Correlation & Covariance of Random Processes

Young W Lim

Sep 11 2021

Based on
Probability, Random Variables and Random Signal Principles,
P.Z. Peebles,Jr. and B. Shi

Outline

Outline

Correlation functions (1)

- A **correlation function** gives
  the statistical correlation between random variables,
  contingent on the spatial or temporal distance between those variables.
- An **auto-correlation function**
  random variables represent the same quantity
  measured at two different points
- A **cross-correlation function**
  random variables are different quantities
  measured at two different points

https://en.wikipedia.org/wiki/Correlation_function

Correlation functions (2)

- a useful indicator of dependencies as a function of distance in time or space
- can be used to assess the distance required between sample points for the values to be effectively uncorrelated.
- can form the basis of rules for interpolating values at points for which there are no observations.

https://en.wikipedia.org/wiki/Correlation_function

Auto-correlation functions (1)

in deterministic processes

- **Autocorrelation** is the correlation of a signal with a delayed copy of itself as a function of delay.
- the similarity between observations as a function of the time lag between them.

https://en.wikipedia.org/wiki/Autocorrelation

Auto-correlation functions (2)

in deterministic processes

- auto-correlation of continuous time signal (deterministic)

$$R_{ff}(\tau) = \int_{-\infty}^{\infty} f(t+\tau)\overline{f(t)}dt = \int_{-\infty}^{\infty} f(t)\overline{f(t-\tau)}dt$$

- auto-correlation of discrete time signal (deterministic)

$$R_{gg}[k] = \sum_n g[n+k]\overline{g[k]}dt = \sum_n g[n]\overline{g[n-\tau]}dt$$

https://en.wikipedia.org/wiki/Autocorrelation

Auto-correlation functions (3)

in deterministic processes

- a mathematical tool for
  - finding repeating patterns,
    such as the presence of a periodic signal obscured by noise,
  - identifying the missing fundamental frequency in a signal implied by its harmonic frequencies.
- often used in signal processing for analyzing functions or series of values, such as time domain signals.

https://en.wikipedia.org/wiki/Autocorrelation

Auto-correlation functions (4)

In random processes

- the **autocorrelation** of a real or complex random process
- the **Pearson correlation**
  between values of the process at different times,
  as a function of the two times (t_1, t_2)
  or a function of the time lag (τ)

https://en.wikipedia.org/wiki/Autocorrelation

Auto-correlation functions (5)

In random processes

- Let $\{X(t)\}$ be a random process, and t be any point in time
  t may be an integer for a discrete-time process
  or a real number for a continuous-time process
- Then $X(t)$ is the value (or **realization**) produced
  by a given run of the process at time t.
- the auto-correlation function between times t_1 and t_2

$$R_{XX}(t_1, t_2) = E\left[X(t_1)\overline{X(t_2)}\right]$$

- the auto-correlation function for WSS randdom process $\{X(t)\}$

$$R_{XX}(\tau) = E\left[X(t)\overline{X(t+\tau)}\right]$$

https://en.wikipedia.org/wiki/Autocorrelation

The properties of autocorrelation functions (1)

$|R_{XX}(\tau)|, R_{XX}(-\tau), R_{XX}(0)$

$$|R_{XX}(\tau)| \le R_{XX}(0)$$

$$R_{XX}(-\tau) = R_{XX}(\tau)$$

$$R_{XX}(0) = E\left[X^2(t)\right]$$

$$P[|X(t+\tau) - X(t)| > \varepsilon] = \frac{2}{\varepsilon^2}(R_{XX}(0) - R_{XX}(\tau))$$

The properties of autocorrelation functions (2)

$R_{NN}(\tau)$,$R_{XX}(\tau)$

if

$$X(t) = \overline{X} + N(t)$$

where $N(t)$ is WSS,
is **zero-mean**,

$$m_N(t) = 0$$

has **autocorrelation** function

$$R_{NN}(\tau) \to 0 \ \ as \ \ |\tau| \to \infty$$

then

$$R_{XX}(\tau) \to \overline{X}^2 \ \ as \ \ |\tau| \to \infty$$

The properties of autocorrelation functions (3)

$R_{NN}(\tau)$,$R_{XX}(\tau)$

If

$$X(t) = \overline{X} + N(t)$$
$$m_N(t) = 0$$
$$\lim_{|\tau|\to\infty} R_{NN}(\tau) = 0$$

then

$$\lim_{|\tau|\to\infty} R_{XX}(\tau) = \overline{X}^2$$

The properties of autocorrelation functions (4)

$R_{NN}(\tau)$,$R_{XX}(\tau)$

if $X(t)$ is **mean square periodic**, i.e,
there exists a $T \neq 0$ such that

$$E\left[\{X(t+T)-X(t)\}^2\right]=0 \qquad \textit{for all } t,$$

then $R_{XX}(t)$ will have a **periodic** component
with the same period

$$R_{XX}(\tau+T)=R_{XX}(\tau)$$

The properties of autocorrelation functions (5)

$R_{NN}(\tau)$,$R_{XX}(\tau)$

- $R_{XX}(\tau + T) = R_{XX}(\tau)$
- $R_{XX}(T) = R_{XX}(0)$
- $E[X(t)X(t+T)] = E[X(t)X(t)]$
- $E[X(t)X(t+T)] = E[X(t+T)X(t+T)]$
- $2E[X(t)X(t+T)] = E[X(t)X(t)] + E[X(t+T)X(t+T)]$
- $E\left[\{X(t) - X(t+T)\}^2\right] = 0$

http://www.math.pitt.edu/~troy/stochastic/meansquareperiodic.pdf

The properties of autocorrelation functions (6)

$R_{NN}(\tau)$,$R_{XX}(\tau)$

$R_{XX}(\tau)$ cannot have an arbitrary shape

any arbitrary function cannot be an autocorrelation funciton.

$R_{XX}(\tau)$ is related to the **power density spectrum**
through the Fourier transform and
the form of the spectrum is not arbitrary

Outline

Crosscorrelation functions (1)

In signal processing

- a measure of similarity of two series as a function of the displacement of one relative to the other.
- commonly used for searching a long signal for a shorter, known feature.
- It has applications in pattern recognition, single particle analysis, electron tomography, averaging, cryptanalysis, and neurophysiology.

https://en.wikipedia.org/wiki/Cross-correlation

Crosscorrelation functions (2)

In signal processing

- similar in nature to the **convolution** of two functions.
- In an **autocorrelation**,
  which is the cross-correlation of a signal with itself,
  there will always be a peak at a lag of zero,
  and its size will be the signal energy.

https://en.wikipedia.org/wiki/Cross-correlation

Crosscorrelation functions (3)

for deterministic signals

- consider two real valued functions f and g differing only by an unknown shift along the x-axis.
- One can use the cross-correlation to find how much g must be shifted along the x-axis to make it identical to f.
- The formula essentially slides the g function along the x-axis, calculating the integral of their product at each position.

https://en.wikipedia.org/wiki/Cross-correlation

Crosscorrelation functions (4)

for deterministic signals

- With complex-valued functions f and g, taking the **conjugate** of f ensures that aligned peaks (or aligned troughs) with imaginary components will contribute positively to the integral.

$$(f \star g)(\tau) \triangleq \int_{-\infty}^{\infty} \overline{f(t)} g(t+\tau)\, dt$$

https://en.wikipedia.org/wiki/Cross-correlation

Crosscorrelation functions (5)

for deterministic signals

- When the functions match, the value of $(f \star g)$ is maximized.
- This is because when peaks (positive areas) are aligned, they make a large contribution to the integral.
- Similarly, when troughs (negative areas) align, they also make a positive contribution to the integral because the product of two negative numbers is positive.

$$(f \star g)(\tau) \triangleq \int_{-\infty}^{\infty} \overline{f(t)} g(t+\tau) \, dt$$

https://en.wikipedia.org/wiki/Cross-correlation

Crosscorrelation functions (6)

for random processes

- the **cross-correlation** of a pair of **random processes** is the correlation between values of the **processes** at different times, as a function of the two times.

https://en.wikipedia.org/wiki/Cross-correlation

Crosscorrelation functions (6)

for random processes

- Let $(X(t), Y(t))$ be a pair of **random processes**, and t be any point in time
- t may be an integer for a discrete-time process or a real number for a continuous-time process
- Then $X(t)$ is the value (or realization) produced by a given run of the process at time t.

$$R_{XY}(t_1, t_2) = E[X_{t_1} \overline{Y_{t_2}}]$$

https://en.wikipedia.org/wiki/Cross-correlation

Crosscorrelation functions (7)

Normalization

- It is common practice in some disciplines
  (e.g. statistics and time series analysis)
  to **normalize** the cross-correlation function
  to get a time-dependent **Pearson correlation coefficient**.
- However, in other disciplines (e.g. engineering)
  the normalization is usually dropped and
  the terms "**cross-correlation**" and "**cross-covariance**"
  are used interchangeably.

https://en.wikipedia.org/wiki/Cross-correlation

Crosscorrelation functions (8)

$R_{XY}(t_1, t_2), R_{XY}(t, t+\tau)$

$$R_{XY}(t_1, t_2) = E[X(t_1)Y(t_2)]$$

$$R_{XY}(t, t+\tau) = E[X(t)Y(t+\tau)] = R_{XY}(\tau)$$

Crosscorrelation functions (9)

$R_{XY}(t_1, t_2), R_{XY}(t, t+\tau)$

if

$$R_{XY}(t, t+\tau) = R_{XY}(\tau) = 0$$

then $X(t)$ and $Y(t)$ are called **orthogonal processes**

Orthogonal Random Variables (1)

$R_{XY}(t_1, t_2), R_{XY}(t, t+\tau)$

vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are **orthogonal**
if their dot product is zero, i.e. $\boldsymbol{x}^T \boldsymbol{y} = 0$.

However for vectors with random components,
the **orthogonality** condition is modified to be
Expected Value $E[\boldsymbol{x}^T \boldsymbol{y}] = 0$.

for orthogonality, each random outcome of $\boldsymbol{x}^T \boldsymbol{y}$ may not be zero,
sometimes positive, sometimes negative, possibly also zero,
but Expected Value $E[\boldsymbol{x}^T \boldsymbol{y}] = 0$.

https://math.stackexchange.com/questions/474840/what-does-orthogonal-random-variables-mean

Orthogonal Random Variables (2)

$R_{XY}(t_1, t_2), R_{XY}(t, t+\tau)$

Two random processes $X(t)$ and $Y(t)$ are called **orthogonal** if their cross-correlation

$$R_{XY}(\tau) = E\left[X(t)Y(t+\tau)\right] = 0$$

https://en.wikipedia.org/wiki/Stochastic_process#Orthogonality

Statistically independent

$R_{XY}(t, t+\tau), R_{XY}(\tau)$

if $X(t)$ and $Y(t)$ are **statistically independent**

$$R_{XY}(t, t+\tau) = E[X(t)Y(t+\tau)] = m_X(t)m_Y(t+\tau)$$

if $X(t)$ and $Y(t)$ are **stistically independent** and are at least **WSS**,

$$R_{XY}(\tau) = \overline{X}\ \overline{Y}$$

which is constant

The properties of crosscorrelation functions (1)

$R_{XY}(\tau), |R_{XY}(\tau)|$

$$R_{XY}(-\tau) = R_{YX}(\tau)$$

$$|R_{XY}(\tau)| \leq \sqrt{R_{XX}(0)R_{YY}(0)}$$

$$|R_{XY}(\tau)| \leq \frac{1}{2}[R_{XX}(0) + R_{YY}(0)]$$

The properties of crosscorrelation functions (2)

$R_{YX}(-\tau)$

$$R_{YX}(-\tau) = E[Y(t)X(t-\tau)] = E[Y(s+\tau)X(s)] = R_{XY}(\tau)$$

$$E\left[\{Y(t+\tau)+\alpha X(t)\}^2\right] \geq 0$$

the **geometric mean** of two positive numbers
cannot exceed their **arithmetic mean**

The properties of crosscorrelation functions (3)

$|R_{XY}(\tau)|$

$$|R_{XY}(\tau)| \leq \frac{1}{2}[R_{XX}(0) + R_{YY}(0)]$$

$$\sqrt{R_{XX}(0)R_{YX}(0)} \leq \frac{1}{2}[R_{XX}(0) + R_{YY}(0)]$$

Outline

Auto-covariance (1)

- With the usual notation E for the expectation operator, if the stochastic process $\{X(t)\}$ has the mean function $\mu(t) = \mathrm{E}[X(t)]$, then the autocovariance is given by

$$\begin{aligned} \mathrm{K}_{XX}(t_1, t_2) &= \operatorname{cov}[X(t_1), X(t_2)] \\ &= \mathrm{E}[(X(t_1) - \mu(t_1))(X(t_2) - \mu(t_2))] \\ &= \mathrm{E}[X(t_1)X(t_2)] - \mu(t_1)\mu(t_2) \end{aligned}$$

- where t_1 and t_2 are two moments in time.

https://en.wikipedia.org/wiki/Autocovariance

Auto-covariance (2)

If $\{X(t)\}$ is a weakly stationary (WSS) process,
then the following are true

- $\mu(t_1) = \mu(t_2) \triangleq \mu$ for all $t_1,\ t_2$
- $\mathrm{E}[|X(t)|^2] < \infty$ for all t
- $\mathrm{K}_{XX}(t_1, t_2) = \mathrm{K}_{XX}(t_2 - t_1, 0) \triangleq \mathrm{K}_{XX}(t_2 - t_1) = \mathrm{K}_{XX}(\tau)$

where $\tau = t_2 - t_1$ is the lag time, or the amount of time
by which the signal has been shifted.

https://en.wikipedia.org/wiki/Autocovariance

Auto-covariance (3)

The autocovariance function of a WSS process is therefore given by

$$\begin{aligned} \mathrm{K}_{XX}(\tau) &= \mathrm{E}[(X(t)-\mu(t))(X(t-\tau)-\mu(t-\tau))] \\ &= \mathrm{E}[X(t)X(t-\tau)]-\mu^2 \end{aligned}$$

which is equivalent to

$$\begin{aligned} \mathrm{K}_{XX}(\tau) &= \mathrm{E}[(X(t+\tau)-\mu(t+\tau))(X(t)-\mu(t))] \\ &= \mathrm{E}[X(t)X(t+\tau)]-\mu^2 \end{aligned}$$

https://en.wikipedia.org/wiki/Autocovariance

Outline

Cross-covariance (1)

- Let $\{X(t)\}$ and $\{Y(t)\}$ denote stochastic processes. Then the cross-covariance function of the processes K_{XY} is defined by

$$\begin{aligned} \mathrm{K}_{XY}(t_1,t_2) &= \operatorname{cov}[X(t_1),Y(t_2)] \\ &= \mathrm{E}[(X(t_1)-\mu_X(t_1))(Y(t_2)-\mu_Y(t_2))] \\ &= \mathrm{E}[X(t_1)X(t_2)] - \mu_X(t_1)\mu_Y(t_2) \end{aligned}$$

- where t_1 and t_2 are two moments in time.

https://en.wikipedia.org/wiki/Autocovariance

Cross-covariance (2)

If $\{X(t)\}$ and $\{Y(t)\}$ are a jointly weakly stationary (WSS) process, then the following are true

- $\mu_X(t_1) = \mu_X(t_2) \triangleq \mu_X$ for all $t_1,\ t_2$
- $\mu_Y(t_1) = \mu_Y(t_2) \triangleq \mu_Y$ for all $t_1,\ t_2$
- $\mathrm{K}_{XY}(t_1, t_2) = \mathrm{K}_{XY}(t_2 - t_1, 0) \triangleq \mathrm{K}_{XY}(t_2 - t_1) = \mathrm{K}_{XY}(\tau)$

where $\tau = t_2 - t_1$ is the lag time, or the amount of time by which the signal has been shifted.

https://en.wikipedia.org/wiki/Autocovariance

Cross-covariance (3)

The autocovariance function of a WSS process is therefore given by

$$\begin{aligned} \mathrm{K}_{XY}(\tau) &= \mathrm{E}[(X(t)-\mu_X(t))(Y(t-\tau)-\mu_Y(t-\tau))] \\ &= \mathrm{E}[X(t)Y(t-\tau)]-\mu_X\mu_Y \end{aligned}$$

which is equivalent to

$$\begin{aligned} \mathrm{K}_{XY}(\tau) &= \mathrm{E}[(X(t+\tau)-\mu_X(t+\tau))(Y(t)-\mu_Y(t))] \\ &= \mathrm{E}[X(t)Y(t+\tau)]-\mu_X\mu_Y \end{aligned}$$

https://en.wikipedia.org/wiki/Autocovariance

Units of measurement

- The units of measurement of the **covariance** $\text{cov}(X, Y)$ are those of X times those of Y.
- By contrast, **correlation coefficients**, which depend on the **covariance**, are a dimensionless measure of linear dependence. (In fact, **correlation coefficients** can simply be understood as a **normalized** version of **covariance**.)

https://en.wikipedia.org/wiki/Covariance

Covariance with itself

- The variance is a special case of the covariance in which the two variables are identical (that is, in which one variable always takes the same value as the other)

$$\operatorname{cov}(X,X) = \operatorname{var}(X) \equiv \sigma^2(X) \equiv \sigma_X^2.$$

https://en.wikipedia.org/wiki/Covariance

Covariance of linear combination

- If X, Y, W, and V are real-valued random variables and a, b, c, d are real-valued constants, then the following facts are a consequence of the definition of covariance:

$$\begin{aligned}
\operatorname{cov}(X, a) &= 0 \\
\operatorname{cov}(X, X) &= \operatorname{var}(X) \\
\operatorname{cov}(X, Y) &= \operatorname{cov}(Y, X) \\
\operatorname{cov}(aX, bY) &= ab \operatorname{cov}(X, Y) \\
\operatorname{cov}(X + a, Y + b) &= \operatorname{cov}(X, Y) \\
\operatorname{cov}(aX + bY, cW + dV) &= ac \operatorname{cov}(X, W) + ad \operatorname{cov}(X, V) \\
&\quad + bc \operatorname{cov}(Y, W) + bd \operatorname{cov}(Y, V)
\end{aligned}$$

https://en.wikipedia.org/wiki/Covariance

Uncorrelated (1)

- Random variables whose covariance is zero are called uncorrelated.
- For example, let X be uniformly distributed in $[-1, 1]$ and let $Y = X^2$. Clearly, X and Y are not independent, but

$$\begin{aligned}
\operatorname{cov}(X, Y) &= \operatorname{cov}\left(X, X^2\right) \\
&= \mathrm{E}\left[X \cdot X^2\right] - \mathrm{E}[X] \cdot \mathrm{E}\left[X^2\right] \\
&= \mathrm{E}\left[X^3\right] - \mathrm{E}[X]\,\mathrm{E}\left[X^2\right] \\
&= 0 - 0 \cdot \mathrm{E}[X^2] \\
&= 0.
\end{aligned}$$

https://en.wikipedia.org/wiki/Covariance

Uncorrelated (2)

- In this case, the relationship between Y and X is non-linear, while correlation and covariance are measures of linear dependence between two random variables.
- This example shows that if two random variables are uncorrelated, that does not in general imply that they are independent.
- However, if two variables are jointly normally distributed (but not if they are merely individually normally distributed), uncorrelatedness does imply independence.

https://en.wikipedia.org/wiki/Covariance

Inner Product

- Many of the properties of covariance can be extracted elegantly by observing that it satisfies similar properties to those of an inner product:
- bilinear: for constants a and b and random variables X, Y, Z, $\text{cov}(aX + bY, Z) = a\,\text{cov}(X, Z) + b\,\text{cov}(Y, Z)$
- symmetric: $\text{cov}(X, Y) = \text{cov}(Y, X)$
- positive semi-definite: $\sigma^2(X) = \text{cov}(X, X) \geq 0$ for all random variables X, and $\text{cov}(X, X) = 0$ implies that X is constant almost surely. $\text{cov}(X, X) = 0\text{cov}(X, X) = 0$ implies that X is constant almost surely.

https://en.wikipedia.org/wiki/Covariance

https://en.wikipedia.org/wiki/Covariance

Covariance Functions

$C_{XX}(t, t+\tau), C_{XY}(t, t+\tau)$

$C_{XX}(t, t+\tau) = E[\{X(t) - m_X(t)\}\{X(t+\tau) - m_X(t+\tau)\}]$
$C_{XY}(t, t+\tau) = E[\{X(t) - m_X(t)\}\{Y(t+\tau) - m_Y(t+\tau)\}]$
$C_{XX}(t, t+\tau) = R_{XX}(t, t+\tau) - m_X(t)m_X(t+\tau)$
$C_{XY}(t, t+\tau) = R_{XY}(t, t+\tau) - m_X(t)m_Y(t+\tau)$

at least jointly WSS

$$C_{XX}(\tau) = R_{XX}(\tau) - \overline{X}^2$$

$$C_{XY}(\tau) = R_{XY}(\tau) - \overline{XY}$$

The properties of covariance functions

$C_{XX}(0)$

For a **WSS process**, variance does not depend on time and if $\tau = 0$

$$C_{XX}(0) = R_{XX}(0) - \overline{X}^2$$

$$\sigma_X^2 = E\left[\{X(t) - E[X(t)]\}^2\right] = C_{XX}(0)$$

it the two random processes **uncorrelated**

$$C_{XY}(t, t+\tau) = R_{XY}(t, t+\tau) - m_X(t)m_Y(t+\tau) = 0$$

$$R_{XY}(t, t+\tau) = m_X(t)m_Y(t+\tau)$$

Discrete-Time Processes and Sequences (1)

$R_{XX}[n, n+k], R_{YY}[n, n+k], C_{XX}[n, n+k], C_{YY}[n, n+k]$

$m_X[n] = \overline{X}$, $m_Y[n] = \overline{Y}$

$$R_{XX}[n, n+k] = R_{XX}[k]$$

$$R_{YY}[n, n+k] = R_{YY}[k]$$

$$C_{XX}[n, n+k] = R_{XX}[k] - \overline{X}^2$$

$$C_{YY}[n, n+k] = R_{YY}[k] - \overline{Y}^2$$

Discrete-Time Processes and Sequences (2)

$R_{XY}[n, n+k], R_{YX}[n, n+k], C_{XY}[n, n+k], C_{YX}[n, n+k]$

$m_X[n] = \overline{X}$, $m_Y[n] = \overline{Y}$

$$R_{XY}[n, n+k] = R_{XY}[k]$$

$$R_{YX}[n, n+k] = R_{YX}[k]$$

$$C_{XY}[n, n+k] = R_{XY}[k] - \overline{XY}$$

$$C_{YX}[n, n+k] = R_{YX}[k] - \overline{YX}$$

Covariance example (1)

- Suppose that X and Y have the following joint probability mass function, in which the six central cells give the discrete joint probabilities $f(x, y)$ of the six hypothetical realizations

		x = 5	x = 6	x = 7	$f_Y(y)$
y	8	0	0.4	0.1	0.5
	9	0.3	0	0.2	0.5
$f_X(x)$		0.3	0.4	0.3	1

- $\mu_X = 5(0.3) + 6(0.4) + 7(0.1 + 0.2) = 6$
- $\mu_Y = 8(0.4 + 0.1) + 9(0.3 + 0.2) = 8.5$

https://en.wikipedia.org/wiki/Covariance

Covariance example (2)

		x = 5	6	7	$f_Y(y)$
y	8	0	0.4	0.1	0.5
	9	0.3	0	0.2	0.5
$f_X(x)$		0.3	0.4	0.3	1

- $\mu_X = 6$ and $\mu_Y = 8.5$
- $cov(X,Y) = \sigma_{XY} = \sum_{(x,y)\in S} f(x,y)(x-\mu_X)(y-\mu_Y)$

$(0)(5-6)(8-8.5)+(0.4)(6-6)(8-8.5)+(0.1)(7-6)(8-8.5)+$
$(0.3)(5-6)(9-8.5)+(0)(6-6)(9-8.5)+(0.2)(7-6)(9-8.5)$
$=-0.1$

https://en.wikipedia.org/wiki/Covariance