MEK4350
Karsten Trulsen
December 2022

4 Stochastic theory

Stochastic description of the sea surface can be found in the books by Kinsman (1984), Ochi (1998), Goda (2000) and Tucker & Pitt (2001). The book by Naess & Moan (2013) is also recommended, although being more oriented toward structural response. For an introduction to probability and stochastic analysis, see the book by Papoulis & Pillai (2002).

4.1 Stochastic description of the sea surface

Suppose we want to describe the vertical displacement $\eta(\boldsymbol{r},t)$ of the water surface as a function of horizontal position $\boldsymbol{r}=x\boldsymbol{i}_x+y\boldsymbol{i}_y$ and time t. A quick look at any water surface under the action of wind suggests that a deterministic description is highly unsatisfactory. Given otherwise identical conditions (wind, etc.) it is our everyday experience that the waves are random and unpredictable. We may therefore think of the surface displacement as a *stochastic process* with several possible *outcomes* or *realizations*. A collection of realizations will be called an *ensemble*. The process of achieving a realization can be thought of as an *experiment*.

In figure 1 we see a few possible time series for the vertical elevation ${}^j\eta(t)$ of ocean surface waves measured at a fixed point. Here the upper left index j denotes the outcome or realization.

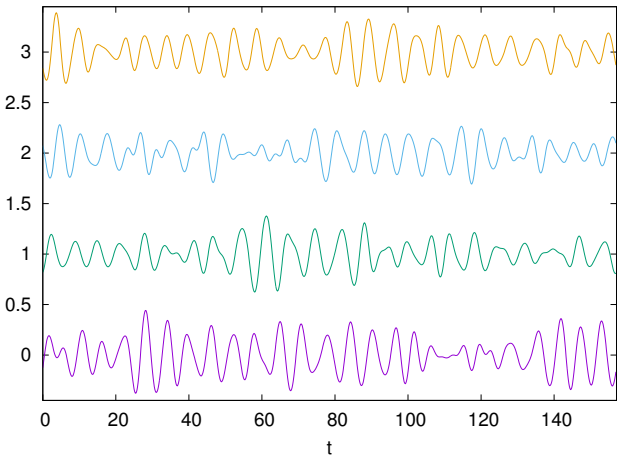

Figure 1: Some realizations for a typical wave time series ${}^j\eta(t)$, for realizations indexed by $j=0,1,2,3$, each plotted offset by the value of j.

It is useful to think about how an ensemble like that shown in figure 1 can be achieved: We could consider the realizations as measurements by distinct buoys at different locations in the ocean, however such a model is not useful if our purpose is to describe how the surface elevation is related between the different locations.

We could consider the realizations as measurements at different hours of the day or different days of the month, however such a model is not useful if our purpose is to describe the long time evolution of the waves. We could consider the realizations as pertaining to different and parallel worlds, this is easier to imagine in our fantasy than to achieve in reality. A different approach is to carry out a set of experiments in the laboratory or a set of simulations on the computer.

A *stochastic variable* Z is a rule that assigns a number ${}^{j}z$ to every outcome j of an experiment.

A *stochastic process* $Z(t)$ is a rule that assigns a function ${}^{j}z(t)$ to every outcome j of an experiment. Of fundamental interest to us, we shall describe the vertical displacement of the sea surface as a stochastic process $Z(\boldsymbol{r}, t)$ being a rule that assigns a spatio-temporal sea surface ${}^{j}z(\boldsymbol{r}, t)$ to every outcome j of an experiment. (We switched notation from η to Z and from ${}^{j}\eta$ to ${}^{j}z$.)

We may now distinguish at least four different interpretations of a stochastic process $Z(t)$ like the one illustrated in figure 1:

1. We consider the process $Z(t)$ for all times t and all outcomes j.
2. Given a particular outcome j_1, we consider a time series ${}^{j_1}z(t)$ for all times t.
3. Given a particular time t_1, we consider a stochastic variable $Z(t_1)$ for all outcomes j.
4. Given a particular outcome j_1 and a particular time t_1, we consider a number ${}^{j_1}z(t_1)$.

In order to characterize a stochastic process we may want to extract averaged parameters. Figure 1 suggests there are at least two ways to extract such averages:

1. Averaging in time (or in space) for a single realization.
2. Averaging over an ensemble.

As an example, we consider the mean level $\bar{\eta}$ of the ocean surface. Averaging in time for a single realization j we have

$$ {}^{j}\bar{\eta} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} {}^{j}\eta(t)\,\mathrm{d}t. $$

Averaging over ensemble for a fixed time t we have

$$ \bar{\eta}(t) = \lim_{N\to\infty} \frac{1}{N} \sum_{j=0}^{N-1} {}^{j}\eta(t). $$

Here we have made some assumptions: The less important assumption is that the realizations are countable, however there is no reason why they cannot be uncountable. The more important assumptions are that each realization is equally likely and that as the size of the ensemble becomes large, the ensemble does indeed represent the true properties of the original process.

If the average over ensemble is independent of time (or space), we say that the process is *stationary* (or *homogeneous*). If the average over ensemble is equal to the average over time (or space), $\bar{\eta}(t) = {}^{j}\bar{\eta}$, we say that the process is *ergodic*. It is clear that in order to be ergodic, the process needs to be stationary (or homogeneous). However, it is possible for a process to be stationary (or homogeneous) without being ergodic as the example in figure 2 suggests.

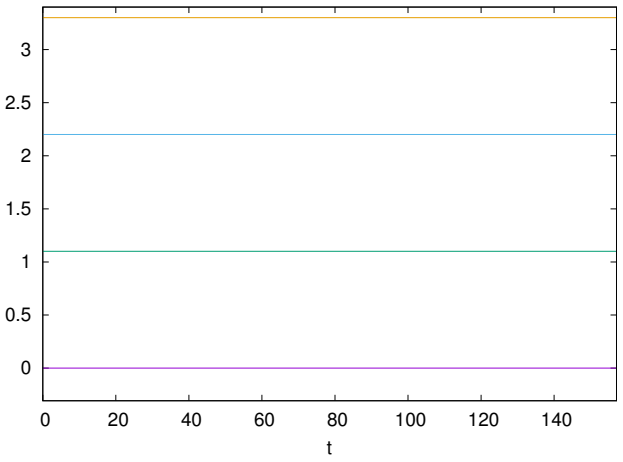

Figure 2: Some realizations for a constant time series ${}^{j}\eta(t)$, for realizations indexed by $j = 0, 1, 2, 3$, each plotted offset by the value of j.

Of particular interest for characterizing a stochastic process $Z(t)$ is the *cumulative distribution function* of the process. The simplest case is to consider one fixed time t, the cumulative distribution function is the proportion of realizations that are bounded from above by a threshold value z

$$F(z;t) = \lim_{N\to\infty} \frac{\#_N\{{}^{j}z(t) \leq z\}}{N}$$

where the notation $\#_N\{\cdot\}$ means the number of realizations within an ensemble of size N that satisfy the condition specified within the curly braces.

We can also consider the *joint* cumulative distribution function for two times t_1 and t_2, which is the proportion of realizations that are bounded from above by the two threshold values z_1 and z_2 at the two times respectively

$$F(z_1, z_2; t_1, t_2) = \lim_{N\to\infty} \frac{\#_N\{{}^{j}z(t_1) \leq z_1, {}^{j}z(t_2) \leq z_2\}}{N}.$$

This can be further generalized to the joint cumulative distribution function for any number of times and threshold values.

If the process is stationary then the cumulative distribution function for a single time is not a function of time, and if the process is in addition ergodic then we can obtain the cumulative distribution function by consideration of a single realization

$$F(z) = \lim_{T\to\infty} \frac{\#_T\{{}^{j}z(t) \leq z\}}{T}$$

where the notation $\#_T\{\cdot\}$ means the amount of time within a total time T that the condition specified within the curly braces is satisfied.

For the above limits we have again assumed that each realization is equally likely and that as the size of the ensemble becomes large, the ensemble does indeed represent the true properties of the original process.

The above considerations are rather empirical. In the following we continue with an elementary introduction from a more non-empirical point of view without evaluation of the limits above.

Example: Staircase time series assumed to be ergodic.

Suppose a stochastic process $Z(t)$ has been measured by the single time series shown in figure 3. The time series is a repeating staircase with three horizontal steps, each horizontal step has equal length and the time series is periodic with period equal to three such lengths.

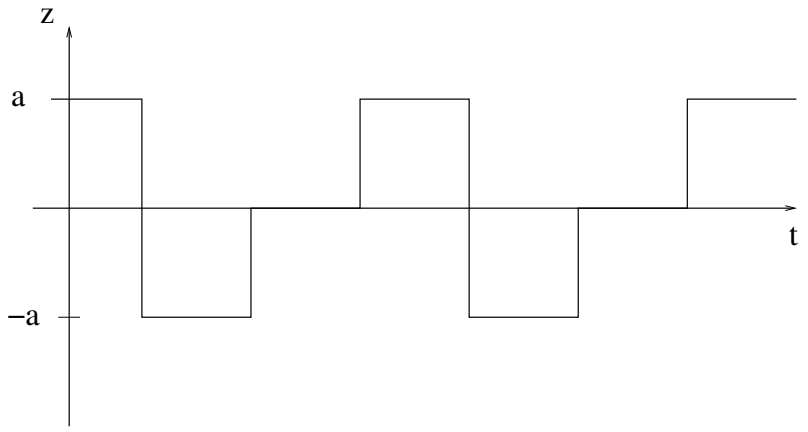

Figure 3: Staircase time series with horizontal parts of equal length.

Provided the process is ergodic, we can compute the statistics from this single realization, and we then find the mean value $\mu = 0$ and the cumulative distribution function

$$F(z) = \begin{cases} 0 & \text{for } z < -a \\ \frac{1}{3} & \text{for } -a \leq z < 0 \\ \frac{2}{3} & \text{for } 0 \leq z < a \\ 1 & \text{for } a \leq z \end{cases}$$

4.2 One real stochastic variable

Suppose X is a real stochastic variable. With the notation $\{X \leq x\}$ we refer to the collection of all outcomes ${}^j x$ of the stochastic variable X such that ${}^j x \leq x$. The probability for this collection of outcomes defines the *cumulative distribution function*

$$F(x) \equiv P\{X \leq x\}$$

where $P\{\cdot\}$ reads the "probability of $\{\cdot\}$". The cumulative probability function has the properties that

1. $F(-\infty) = 0$,
2. $F(\infty) = 1$,
3. $F(x)$ is a non-decreasing function, $F(x_1) \leq F(x_2)$ for $x_1 < x_2$.

The probability that an outcome is between a lower and an upper bound is $P\{a < X \leq b\} = F(b) - F(a)$. Similarly, the probability that an outcome is in an interval of infinitesimal width is $P\{x < X \leq x + \mathrm{d}x\} = F(x + \mathrm{d}x) - F(x) \approx \frac{\mathrm{d}F}{\mathrm{d}x}\mathrm{d}x = f(x)\mathrm{d}x$.

We define the *probability density function* as

$$f(x) \equiv \frac{\mathrm{d}F}{\mathrm{d}x}.$$

The probability density function has the properties that

1. $f(x) \geq 0$,
2. $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$,
3. $F(x) = \int_{-\infty}^{x} f(\xi)\,\mathrm{d}\xi$.

We define the *mode* (Norwegian *typetall*) of a stochastic variable X to be the value of x such that the probability density function $f(x)$ achieves its maximum. If the probability density has a single maximum it is said to be *unimodal*. If it has two maxima it is said to be *bimodal*, etc.

We define the *median* of a stochastic variable X to be the value of x such that the cumulative distribution function $F(x) = 0.5$. It is equally probable that the stochastic variable X gives an outcome smaller than or greater than the median.

The *expected value* μ of a stochastic variable X is defined as the weighted average

$$\mu = \mathrm{E}[X] = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x.$$

The expected value of a function $g(X)$ of the stochastic variable X is the weighted average

$$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x.$$

It is seen that the expected value operator is linear. Suppose we have two functions $g(X)$ and $h(X)$ and two constants a and b, then we have

$$\mathrm{E}[ag(X) + bh(X)] = a\,\mathrm{E}[g(X)] + b\,\mathrm{E}[h(X)].$$

The *variance* $\sigma^2 = \mathrm{Var}[X]$ of a stochastic variable X is defined by

$$\sigma^2 = \mathrm{Var}[X] = \mathrm{E}[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\,\mathrm{d}x.$$

By the linearity of the expected value operator, this can be written

$$\sigma^2 = \mathrm{E}[(X - \mu)^2] = \mathrm{E}[X^2 - 2\mu X + \mu^2] = \mathrm{E}[X^2] - 2\mu\,\mathrm{E}[X] + \mu^2 = \mathrm{E}[X^2] - \mu^2.$$

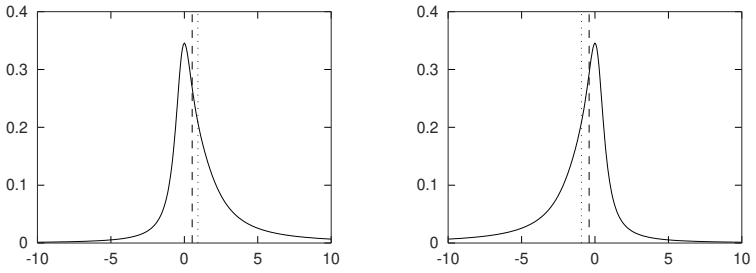

Figure 4: Probability distributions with positive (left) and negative (right) skewness. Solid lines are probability density functions, dashed lines are medians, dottet lines are means. Left: mode 0, median ≈ 0.49, mean $\mu \approx 0.92$, skewness $\gamma \approx 0.56$. Right: Same values with opposite signs.

The *standard deviation* σ of a stochastic variable is defined as the square root of the variance.

The *nth moment* of a stochastic variable X is defined as

$$m_n = \mathrm{E}[X^n] = \int_{-\infty}^{\infty} x^n f(x)\,\mathrm{d}x,$$

while the *nth central moment* is defined as

$$\mu_n = \mathrm{E}[(X-\mu)^n] = \int_{-\infty}^{\infty} (x-\mu)^n f(x)\,\mathrm{d}x.$$

The variance can thus be defined as the second central moment of the stochastic variable. We see that $\mu = m_1$ and $\sigma^2 = m_2 - m_1^2 = \mu_2$ and $\mu_1 = 0$.

The *skewness* γ of a stochastic variable is defined as the third central moment normalized by the cube of the standard deviation

$$\gamma = \frac{\mathrm{E}[(X-\mu)^3]}{\sigma^3} = \frac{\mu_3}{\sigma^3}.$$

The skewness is a measure of the asymmetry of the probability distribution. In figure 4 we see examples of probability distributions with positive and negative skewness.

The *kurtosis* κ of a stochastic variable is defined as the fourth central moment normalized by the square of the variance

$$\kappa = \frac{\mathrm{E}[(X-\mu)^4]}{\sigma^4} = \frac{\mu_4}{\sigma^4}.$$

Sometimes the *excess kurtosis* is introduced as the kurtosis minus 3. For zero excess kurtosis ($\kappa = 3$) the distribution is said to be *mesokurtic*. For positive excess kurtosis ($\kappa > 3$) the distribution is said to be *leptokurtic*. For negative excess kurtosis ($\kappa < 3$) the distribution is said to be *platykurtic*.

The kurtosis is a measure of the weight of the tails of the probability distribution. In figure 5 we see examples of a mesokurtic, a leptokurtic and a platykurtic distribution.

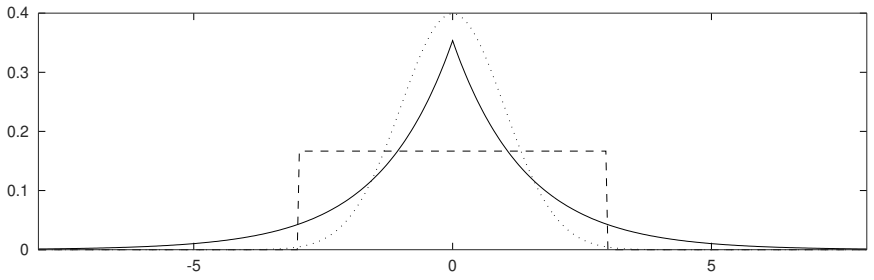

Figure 5: Probability density functions with mean 0 and variance 1. Solid line: Laplace distribution, $\kappa = 6$, positive excess kurtosis, leptokurtic. Dotted line: Gaussian distribution, $\kappa = 3$, zero excess kurtosis, mesokurtic. Dashed line: Uniform distribution, $\kappa = 1.8$, negative excess kurtosis, platykurtic.

Given a stochastic variable X we can transform it into a new stochastic variable Y as a function of X. Particularly useful is the transformation given by

$$Y(X) = \frac{X - \mu}{\sigma}. \tag{1}$$

The cumulative distribution function of Y is

$$F_Y(y) = P\{Y \leq y\} = P\{\frac{X - \mu}{\sigma} \leq y\} = P\{X \leq \mu + \sigma y\} = F_X(\mu + \sigma y).$$

The probability density function of Y is

$$f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} = \frac{\mathrm{d}F_X(\mu + \sigma y)}{\mathrm{d}y} = \sigma f_X(\mu + \sigma y).$$

For the special transformation (1) with μ and σ^2 being the mean and variance of X, we notice that the mean of Y is

$$\mathrm{E}[Y] = \mathrm{E}[\frac{X - \mu}{\sigma}] = \frac{1}{\sigma}\,\mathrm{E}[X] - \frac{\mu}{\sigma} = 0,$$

the variance of Y is

$$\mathrm{E}[Y^2] = \mathrm{E}[\frac{(X - \mu)^2}{\sigma^2}] = \frac{1}{\sigma^2}\,\mathrm{E}[(X - \mu)^2] = 1,$$

while the skewness and kurtosis of Y are the same as the skewness and kurtosis of X, respectively.

The *characteristic function* $\phi(k)$ of a stochastic variable X is defined as the expected value of $\mathrm{e}^{\mathrm{i}kX}$,

$$\phi(k) = \mathrm{E}[\mathrm{e}^{\mathrm{i}kX}] = \int_{-\infty}^{\infty} f(x)\mathrm{e}^{\mathrm{i}kx}\,\mathrm{d}x.$$

We immediately recognize this as the Fourier transform of the probability density function, and we therefore have the inverse transform

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \phi(k)\mathrm{e}^{-\mathrm{i}kx}\,\mathrm{d}k.$$

The variable k used here should not be confused with the wavenumber of some waves. We can formally expand the complex exponential function in a power series

$$\mathrm{e}^{\mathrm{i}kx} = 1 + \mathrm{i}kx - \frac{1}{2}k^2x^2 - \frac{\mathrm{i}}{6}k^3x^3 + \ldots + \frac{\mathrm{i}^n}{n!}k^nx^n + \ldots$$

and thus by the linearity of the expectation operator the characteristic function can be written as a superposition of all the moments

$$\phi(k) = 1 + \mathrm{i}km_1 - \frac{1}{2}k^2m_2 - \frac{\mathrm{i}}{6}k^3m_3 + \ldots + \frac{\mathrm{i}^n}{n!}k^nm_n + \ldots$$

This can be used to derive a convenient formula for the n-th moment simply by differentiating the characteristic function n times and evaluating the result at $k = 0$

$$m_n = (-\mathrm{i})^n \frac{\mathrm{d}^n}{\mathrm{d}k^n}\phi(k)\bigg|_{k=0}.$$

Example: A discrete distribution.

Let us again consider the staircase in figure 3. We find the probability density function by computing the derivative

$$f(z) = \frac{1}{3}\left(\delta(z+a) + \delta(z) + \delta(z-a)\right)$$

where we get a sum of three Dirac deltas. This is, in fact, a discrete probability distribution, but we treat it as a special case of a continuous probability distribution with the help of the Dirac delta generalized function.

The mean is $\mu = \frac{1}{3}(-a + 0 + a) = 0$. The variance is $\sigma^2 = \frac{1}{3}((-a)^2 + 0^2 + a^2) = \frac{2}{3}a^2$. The standard deviation is $\sigma = \sqrt{\frac{2}{3}}a$. The third central moment is $\mu_3 = \frac{1}{3}((-a)^3 + 0^3 + a^3) = 0$, and thus the skewness is also zero. The fourth central moment is $\mu_4 = \frac{1}{3}((-a)^4 + 0^4 + a^4) = \frac{2}{3}a^4$, and thus the kurtosis is $\kappa = \frac{\mu_4}{\sigma^4} = \frac{3}{2}$ so the distribution is platykurtic. The characteristic function is $\phi(k) = \frac{1}{3}(1 + 2\cos(ak))$.

Example: Gaussian or normal distribution.

A stochastic variable X is said to be Gaussian or normally distributed with mean μ and variance σ^2 with the probability density function given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

This distribution is shown in figure 5 above. The cumulative distribution function is

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(\xi-\mu)^2}{2\sigma^2}}\,\mathrm{d}\xi = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}(\frac{x-\mu}{\sqrt{2}\sigma}) = 1 - \frac{1}{2}\,\mathrm{erfc}(\frac{x-\mu}{\sqrt{2}\sigma})$$

where $\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \mathrm{e}^{-t^2}\,\mathrm{d}t$ is the error function and $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$ is the complementary error function. The stochastic variable X is unimodal with mode μ. The median is also μ since $\mathrm{erf}(0) = 0$.

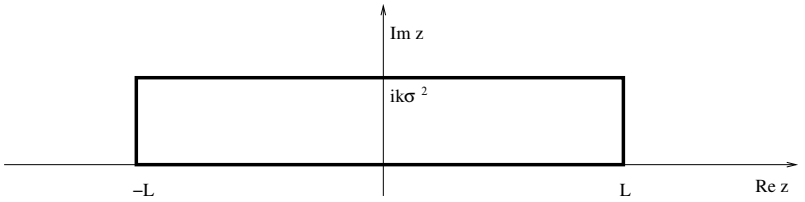

Figure 6: Contour for computation of the characteristic function of the normal distribution.

The characteristic function is[1]

$$\phi(k) = \mathrm{e}^{-\frac{\sigma^2 k^2}{2} + \mathrm{i}\mu k}.$$

The computation of the moments of the normal distribution can be done by direct computation of the expectation integrals, or by means of the characteristic function. The results are the mean $\mathrm{E}[X] = \mu$, the variance $\mathrm{E}[(X-\mu)^2] = \sigma^2$, the skewness $\gamma = 0$ and the kurtosis $\kappa = 3$

Notice that in the limit of zero variance the characteristic function converges to $\phi(k) = \mathrm{e}^{\mathrm{i}\mu k}$ and thus the probability density function converges to a Dirac delta $f(x) = \delta(x-\mu)$.

Example: Uniform distribution.

A stochastic variable X is said to be uniformly distributed on the interval $a \leq X \leq b$ with the probability density function given by

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

This distribution is shown in figure 5 above. The cumulative distribution function is

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x \geq b \end{cases}$$

The mode (maximum) is not well defined since $f(x)$ does not have an isolated extremal point. The median is $x_{\text{median}} = \frac{a+b}{2}$.

The characteristic function is

$$\phi(k) = \frac{\mathrm{e}^{\mathrm{i}kb} - \mathrm{e}^{\mathrm{i}ka}}{\mathrm{i}k(b-a)}.$$

[1] In order to show that this is the characteristic function we extend consideration from the real line to the complex plane. Rewrite the exponent into a square

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(z-\mu)^2}{2\sigma^2} + \mathrm{i}\mu k}\, \mathrm{d}z = \mathrm{e}^{-\frac{\sigma^2 k^2}{2} + \mathrm{i}\mu k} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(z-\mu-\mathrm{i}k\sigma^2)^2}{2\sigma^2}}\, \mathrm{d}z$$

Then we integrate around the contour in the complex plane as suggested in figure 6, along the real axis from $-L$ to L, up to a line parallell to the real axis at imaginary value $\mathrm{i}k\sigma^2$, to the left and back. As $L \to \infty$ the contribution from the vertical segments vanish. There are no singular points so the integral around the closed loop is zero.

The mean is $\mu = \mathrm{E}[X] = \frac{a+b}{2}$. The variance is $\sigma^2 = \mathrm{E}[(X-\mu))^2] = \frac{1}{12}(b-a)^2$. The skewness is $\gamma = 0$. The kurtosis is $\kappa = 9/5$.

4.2.1 Exercises

1. **Central moments computed by the characteristic function**

   Show that the central moments can be obtained from the characteristic function by the formula

   $$\mu_n = (-\mathrm{i})^n \frac{\mathrm{d}^n}{\mathrm{d}k^n} \left(\mathrm{e}^{-\mathrm{i}\mu k}\phi(k)\right)\Big|_{k=0}.$$

2. **Rayleigh distribution**

   A stochastic variable X is said to be Rayleigh distributed with parameter α with the probability density function given by

   $$f(x) = \begin{cases} \frac{x}{\alpha^2}\mathrm{e}^{-\frac{x^2}{2\alpha^2}} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

   Find the mode, median, mean, variance, standard deviation, skewness and kurtosis.

   Hint: Maybe the characteristic function is difficult to compute, so it is easier to find the moments by direct integration?

3. **Exponential distribution**

   A stochastic variable X is said to be exponentially distributed with parameter α with the probability density function given by

   $$f(x) = \begin{cases} \alpha\mathrm{e}^{-\alpha x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

   where $\alpha > 0$.

   Find the mode, median, characteristic function, mean, variance, standard deviation, skewness and kurtosis.

4. **Laplace distribution**

   A stochastic variable X is said to be Laplace distributed with parameter α with the probability density function given by

   $$f(x) = \frac{1}{2\alpha}\mathrm{e}^{-\frac{|x|}{\alpha}}$$

   where $\alpha > 0$. This distribution is shown in figure 5 above.

   Find the mode, median, characteristic function, mean, variance, standard deviation, skewness and kurtosis.

5. **Cauchy distribution**

   A stochastic variable X is said to be Cauchy distributed with parameter α with the probability density function given by

   $$f(x) = \frac{1}{\pi}\frac{\alpha}{x^2+\alpha^2}$$

   Find the mode, median and the characteristic function.

   Notice that the characteristic function is not differentiable at the origin, so the convenient formula for computing the moments by differentiation of the characteristic function will not work. In fact, the Cauchy distribution does not have moments higher than the zeroth moment. The mean, variance, standard deviation, skewness and kurtosis are not defined!

6. **Markov's inequality**

   Let X be a nonnegative random variable with mean μ, and let $a > 0$, show that

   $$P\{X \geq a\} \leq \frac{\mu}{a}.$$

7. **Chebyshev's inequality**

   Let X be a random variable with mean μ and variance σ^2, and let $k > 0$, show that

   $$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}.$$

4.3 Two real stochastic variables

Let X and Y be two real stochastic variables. With the notation $\{X \leq x \text{ and } Y \leq y\}$ we refer to the collection of all outcomes of the two stochastic variables X and Y such that $X \leq x$ and $Y \leq y$ simultaneously. The probability for this collection of outcomes defines the *joint* (Norwegian *simultan*) *cumulative distribution function* $F(x,y) \equiv P\{X \leq x \text{ and } Y \leq y\}$. The joint cumulative probability function has the properties that

1. $F(-\infty,-\infty) = 0$,
2. $F(\infty,\infty) = 1$,
3. $F(x_1,y_1) \leq F(x_2,y_2)$ for $x_1 \leq x_2$ and $y_1 \leq y_2$.

The *marginal cumulative distribution functions* are defined as

$$F_X(x) = F(x,\infty) \qquad \text{and} \qquad F_Y(y) = F(\infty,y).$$

The probability that the joint outcome is within a rectangle is $P\{a_x < X \leq b_x \text{ and } a_y < Y \leq b_y\} = F(b_x,b_y) - F(b_x,a_y) - F(a_x,b_y) + F(a_x,a_y)$. Therefore the probability that an outcome is in a rectangle of infinitesimal widths is $P\{x < X \leq x+\mathrm{d}x \text{ and } y < Y \leq y+\mathrm{d}y\} \approx \frac{\partial^2 F}{\partial x \partial y}\mathrm{d}x\mathrm{d}y = f(x,y)\mathrm{d}x\mathrm{d}y$.

The *joint probability density function* is defined as

$$f(x,y) \equiv \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

and has the properties that

1. $f(x,y) \geq 0$,
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x\,\mathrm{d}y = 1$,
3. $F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(\xi,\eta)\,\mathrm{d}\eta\mathrm{d}\xi$.

The *marginal probability density functions* are defined as

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \frac{\partial F_X(x)}{\partial x} \qquad \text{and} \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x = \frac{\partial F_Y(y)}{\partial y}.$$

Two stochastic variables X and Y are said to be *statistically independent* if the joint distribution function can be factorized $F(x,y) = F_X(x)F_Y(y)$, or equivalently the probability density function can be factorized $f(x,y) = f_X(x)f_Y(y)$.

The mean values of X and Y are

$$\mu_X = \mathrm{E}[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x,$$

$$\mu_Y = \mathrm{E}[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{-\infty}^{\infty} y f_Y(y)\,\mathrm{d}y.$$

The *covariance* of two stochastic variables X and Y is defined as

$$\mathrm{Cov}[X,Y] = \mathrm{E}[(X-\mu_X)(Y-\mu_Y)] = \mathrm{E}[XY] - \mu_X\mu_Y.$$

The *correlation coefficient* of the two stochastic variables is defined as the ratio

$$r = \frac{\mathrm{Cov}[X,Y]}{\sigma_X\sigma_Y}.$$

Example: Show that the magnitude of the correlation coefficient is not greater than one. This can be done by looking at

$$\mathrm{E}[(a(X-\mu_X)+(Y-\mu_Y))^2] = a^2\sigma_X^2 + 2a\,\mathrm{Cov}[X,Y] + \sigma_Y^2 \geq 0$$

where a is a real parameter. Solving for a when the above expression is zero we get

$$a = \frac{-\,\mathrm{Cov}[X,Y] \pm \sqrt{\mathrm{Cov}[X,Y]^2 - \sigma_X^2\sigma_Y^2}}{\sigma_X^2},$$

but observing that there cannot be two distinct real solutions for a, the radicand must be non-positive $\mathrm{Cov}[X,Y]^2 \leq \sigma_X^2\sigma_Y^2$. So we conclude that $|r| \leq 1$.

For another proof of $|r| \leq 1$ see exercise below.

Two stochastic variables X and Y are said to be *uncorrelated* if $\mathrm{Cov}[X,Y] = 0$, which is equivalent to the correlation coefficient $r = 0$, and which is equivalent to $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$.

We observe that if two stochastic variables are statistically independent, then they are uncorrelated. The converse is not always true, see exercise below.

4.3.1 Exercises

1. **Uncorrelated does not imply independent**

   Let X be uniformly distributed on the interval $-1 \leq X \leq 1$ and let $Y = X^2$. Compute the means of X and Y, μ_x and μ_y. Compute the covariance $\mathrm{Cov}[X, Y]$ and show that X and Y are uncorrelated. Compute the two marginal distribution functions $F_X(x)$ and $F_Y(y)$. Show that the joint distribution function is not the product of the two marginal distribution functions. Thus conclude that X and Y are not independent.

   Hint: It is easier to work with the joint cumulative distribution function than the joint probability density.

2. **Geometric interpretation of real stochastic variables**

   First, if we interpret the two real stochastic variables X and Y as vectors, show that $\langle X, Y \rangle = \mathrm{E}[XY]$ satisfies all the requirements for being an inner product. Show that the associated norm is $\|X\| = \sqrt{\mathrm{E}[X^2]}$.

   Show that the cosine of the associated angle between the vectors X and Y is bounded by 1 in magnitude as a consequence of the Cauchy–Schwartz inequality. We say that X and Y are *orthogonal* if $\mathrm{E}[XY] = 0$ and we can write $X \perp Y$.

   Next, if we interpret the two real stochastic variables X and Y as vectors, show that $\langle X, Y \rangle = \mathrm{Cov}[X, Y]$ satisfies all the requirements for being an inner product. Show that the associated norm is the standard deviation of X.

   Show that the bound for the correlation coefficient, $|r| \leq 1$, is a consequence of the Cauchy–Schwartz inequality.

3. Let X be a Rayleigh distributed real stochastic variable with mode α. Find the distribution of $Y = X^2$.

4. Let X and Y be two independent identically distributed real Gaussian stochastic variables with mean μ and variance σ^2. Introduce the new variables R and Θ by $X = R\cos\Theta + \mu$ and $Y = R\sin\Theta + \mu$. Show that R and Θ are independent variables and find their distributions.

5. Let X be a Laplace distributed real stochastic variable with parameter α. Find the distribution of $Y = \sqrt{|X|}$.

6. Let X and Y be two independent identically distributed real Gaussian stochastic variables. Conclude that the sum of squared $S = X^2 + Y^2$ is exponentially distributed with standard deviation equal to its expected value.

4.4 Complex stochastic variable

A *complex stochastic variable* $Z = X + \mathrm{i}Y$ is defined by the joint distribution of its real and imaginary parts X and Y. It does not make sense to compare is one complex number is smaller than or greater than another complex number, but we

can apply such criteria to the real and imaginary parts separately. The probability distribution of a complex variable Z is therefore the joint distribution of its real and imaginary parts X and Y.

For the expected value of Z we have

$$\mu_Z = \mathrm{E}[Z] = \mathrm{E}[X] + \mathrm{i}\,\mathrm{E}[Y] = \mu_X + \mathrm{i}\mu_Y.$$

There are three real second order moments associated with the complex variable $Z = X + \mathrm{i}Y$: The variances $\sigma_X^2 = C_{XX} = \mathrm{Var}[X]$ and $\sigma_Y^2 = C_{YY} = \mathrm{Var}[Y]$ and the covariance $C_{XY} = \mathrm{Cov}[X, Y]$.

We extend the covariance to complex variables by defining the ordinary *complex covariance* by the criterion that the second argument should be conjugated. Suppose U and V are two complex stochastic variables with means μ_U and μ_V, we define their ordinary complex covariance to be

$$C_{UV} = \mathrm{Cov}[U, V] = \mathrm{E}[(U - \mu_U)(V - \mu_V)^*] = \mathrm{E}[UV^*] - \mu_U\mu_V^*.$$

We also have the complex *pseudo-covariance* (also called the complex *complementary variance*) for which no conjugate is taken

$$P_{UV} = \mathrm{E}[(U - \mu_U)(V - \mu_V)] = \mathrm{E}[UV] - \mu_U\mu_V.$$

The variance of a complex variable is the ordinary complex covariance of the variable and itself

$$\mathrm{Var}[Z] = C_{ZZ} = \mathrm{Cov}[Z, Z] = \mathrm{E}[(Z - \mu_Z)(Z - \mu_Z)^*] = \mathrm{E}[|Z|^2] - |\mu_Z|^2 = \sigma_X^2 + \sigma_Y^2$$

while the complex pseudo-covariance of the variable and itself is

$$P_{ZZ} = \mathrm{E}[(Z - \mu_Z)(Z - \mu_Z)] = \sigma_X^2 - \sigma_Y^2 + 2\mathrm{i}C_{XY}.$$

We can now recover the real second-order moments from the complex ones by

$$\sigma_X^2 = \frac{1}{2}\mathrm{Re}(C_{ZZ} + P_{ZZ}), \quad \sigma_Y^2 = \frac{1}{2}\mathrm{Re}(C_{ZZ} - P_{ZZ}) \quad \text{and} \quad C_{XY} = \frac{1}{2}\mathrm{Im}P_{ZZ}.$$

4.4.1 Exercises

1. **Geometric interpretation of complex stochastic variables**

   First, if we interpret the two complex stochastic variables U and V as vectors, show that $\langle U, V\rangle = \mathrm{E}[UV^*]$ satisfies all the requirements for being an inner product. Show that the associated norm is $\|U\| = \sqrt{\mathrm{E}[|U|^2]}$.

   Show that the cosine of the associated angle between the vectors U and V is bounded by 1 in magnitude as a consequence of the Cauchy–Schwartz inequality. We say that U and V are *orthogonal* if $\mathrm{E}[UV^*] = 0$ and we can write $U \perp V$.

   Next, if we interpret the two complex stochastic variables U and V as vectors, show that $\langle U, V\rangle = \mathrm{Cov}[U, V]$ satisfies all the requirements for being an inner product. Show that the associated norm is the standard deviation of U.

   Show that the bound for the correlation coefficient $|r| \leq 1$, is a consequence of the Cauchy–Schwartz inequality.

4.5 Multivariate stochastic variable

Sometimes we are interested in vector-valued stochastic variables. Let $\boldsymbol{X}$ be a real $n \times 1$ vector

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}$$

The mean of $\boldsymbol{X}$ is $\boldsymbol{\mu_X} = \mathrm{E}[\boldsymbol{X}]$ which can be computed componentwise due to the linearity of the expected value.

The joint cumulative distribution function $F(x_1, x_2, \ldots, x_n)$ and the joint probability density function are related by

$$f(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1 \partial x_2 \ldots \partial x_n}.$$

We say that the n variables $X_1, X_2, \ldots, X_n$ are statistically independent if the distribution functions can be factorized

$$f(x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} f_{X_j}(x_j) \qquad \text{and} \qquad F(x_1, x_2, \ldots, x_n) = \prod_{j=1}^{n} F_{X_j}(x_j).$$

The covariance matrix of the components of $\boldsymbol{X}$ is

$$\mathcal{C} = \mathrm{Cov}[\boldsymbol{X}, \boldsymbol{X}] = \mathrm{E}[(\boldsymbol{X} - \boldsymbol{\mu_X})(\boldsymbol{X} - \boldsymbol{\mu_X})^T]$$

where $\mathcal{A}^T$ denotes the transpose of $\mathcal{A}$.

We say that the n variables $X_1, X_2, \ldots, X_n$ are mutually uncorrelated if the covariances are zero $\mathrm{Cov}[X_j, X_l] = 0$ for $j \neq l$.

The characteristic function of $\boldsymbol{X}$ is

$$\phi(\boldsymbol{k}) = \mathrm{E}\left[\exp\left(\mathrm{i}\sum_{j=1}^{n} k_j X_j\right)\right] = \mathrm{E}\left[\exp\left(\mathrm{i}\boldsymbol{k}^T\boldsymbol{X}\right)\right]$$

where $\boldsymbol{k} = (k_1, k_2, \ldots, k_n)^T$.

Sometimes we are also interested in multivariate complex stochastic variables. Let $\boldsymbol{Z}$ be a complex $n \times 1$ vector

$$\boldsymbol{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \end{pmatrix}$$

The mean of $\boldsymbol{Z}$ is $\boldsymbol{\mu_Z} = \mathrm{E}[\boldsymbol{Z}]$.

The covariance matrix of the components of $\boldsymbol{Z}$ is

$$\mathcal{C} = \mathrm{Cov}[\boldsymbol{Z}, \boldsymbol{Z}] = \mathrm{E}[(\boldsymbol{Z} - \boldsymbol{\mu_Z})(\boldsymbol{Z} - \boldsymbol{\mu_Z})^H]$$

where $\mathcal{A}^H$ means the Hermitian or conjugate transpose of $\mathcal{A}$. We let $\overline{\mathcal{A}}$ be the complex conjugate of $\mathcal{A}$, $\mathcal{A}^T$ be the transpose of $\mathcal{A}$, and $\mathcal{A}^* = \mathcal{A}^H = \overline{\mathcal{A}}^T = \overline{\mathcal{A}^T}$ be the Hermitian or conjugate transpose of $\mathcal{A}$.[2]

The covariance matrix is positive semi-definite, i.e. $\boldsymbol{k}^H \mathcal{C} \boldsymbol{k}$ is real and non-negative

$$\boldsymbol{k}^H \mathcal{C} \boldsymbol{k} = \boldsymbol{k}^H \operatorname{E}[(\boldsymbol{Z} - \boldsymbol{\mu}_{\boldsymbol{Z}})(\boldsymbol{Z} - \boldsymbol{\mu}_{\boldsymbol{Z}})^H] \boldsymbol{k} = \operatorname{E}[|\boldsymbol{k}^H (\boldsymbol{Z} - \boldsymbol{\mu}_{\boldsymbol{Z}})|^2] \geq 0$$

For complex variables it may also necessary to compute the pseudo-covariance matrix of the components of $\boldsymbol{Z}$

$$\mathcal{P} = \operatorname{E}[(\boldsymbol{Z} - \boldsymbol{\mu}_{\boldsymbol{Z}})(\boldsymbol{Z} - \boldsymbol{\mu}_{\boldsymbol{Z}})^T]$$

4.5.1 Multivariate real normal distribution

A real multivariate stochastic variable $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^T$ is said to normally distributed, or equivalently the variables X_j are said to be jointly normally distributed, with the probability density function given by

$$f(\boldsymbol{x}) = \frac{\sqrt{|\mathcal{A}|}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \sum_{j,l=1}^{n} (x_j - \mu_j) a_{jl} (x_l - \mu_l) \right)$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$, where the matrix $\mathcal{A} = \{a_{jl}\} = \mathcal{C}^{-1}$, where $|\mathcal{A}|$ denotes the determinant of $\mathcal{A}$, where $\mu_j = \operatorname{E}[X_j]$ are the mean values, and where $\mathcal{C} = \{c_{jl}\}$ is the covariance matrix with components $c_{jl} = \operatorname{Cov}[X_j, X_l] = \operatorname{E}[(X_j - \mu_j)(X_l - \mu_l)]$.

The characteristic function is given by the n-dimensional Fourier transform

$$\phi(\boldsymbol{k}) = \operatorname{E}\left[\exp\left(\mathrm{i}\boldsymbol{k}^T \boldsymbol{X}\right)\right] = \exp\left(-\frac{1}{2} \boldsymbol{k}^T \mathcal{C} \boldsymbol{k} + \mathrm{i} \boldsymbol{k}^T \boldsymbol{\mu} \right) \tag{2}$$

where $\boldsymbol{k} = (k_1, k_2, \ldots, k_n)^T$.

Notice that since the covariance matrix is positive semi-definite, the real part of the exponent is always non-positive, and therefore $|\phi(\boldsymbol{k})| \leq 1$.

Example: Multivariate real normal distribution in the limit of zero variance.

It is useful to consider the limit when one of the variances vanishes $\sigma_j^2 = c_{jj} = 0$ for one of the variables X_j. In this case the covariance matrix becomes singular, it cannot be inverted, and the probability density function becomes a generalized function. However, the characteristic function is still a well defined ordinary function.

Since it is a perfectly acceptable situation that at least one of the variables X_j has zero variance, it is convenient to take the characteristic function in equation (2) as the definition of the multivariate normal distribution.

Example: For multivariate real Gaussian variables, being uncorrelated is equivalent to being statistically independent.

[2] Matlab and Octave have the very strange commands `A'` for Hermitian transpose and `A.'` for just transpose. More logical notations are also available, `conj(A)` for just conjugate, `transpose(A)` for just transpose, and `ctranspose(A)` for Hermitian transpose.

Suppose that the variable X_1 is uncorrelated with all the other variables X_j. Then the covariances $c_{j,1} = c_{1,j} = 0$ for all $j \neq 1$, and thus the n-dimensional characteristic function can be factorized $\phi(k_1, k_2, \ldots, k_n) = \phi_1(k_1)\phi_{2,\ldots,n}(k_2, \ldots, k_n)$. From the multi-dimensional inverse Fourier transform, it follows that the probability density function can be factorized likewise. Hence for joint normally distributed stochastic variables, being statistically independent is equivalent to being uncorrelated.

Example: $\boldsymbol{X}$ being multivariate real Gaussian is equivalent to any linear combination of the components of $\boldsymbol{X}$ being univariate real Gaussian.

We will show that $\boldsymbol{X}$ being multivariate real Gaussian is equivalent to the variable $Y = \sum_{j=1}^{n} \lambda_j X_j$ being real Gaussian for any choice of λ_j.

First, suppose that $\boldsymbol{X}$ is multivariate real Gaussian. In that case the characteristic function for Y is

$$\phi_Y(\kappa) = \mathrm{E}[\mathrm{e}^{\mathrm{i}\kappa Y}] = \mathrm{E}[\mathrm{e}^{\mathrm{i}\kappa(\lambda_1 X_1 + \ldots + \lambda_n X_n)}] = \exp\left(-\frac{1}{2}\sum_{j,l} \kappa\lambda_j c_{jl} \kappa\lambda_l + \mathrm{i}\sum_j \kappa\lambda_j \mu_j\right)$$
$$= \exp\left(-\frac{1}{2}\kappa^2 \sum_{j,l} \lambda_j c_{jl} \lambda_l + \mathrm{i}\kappa \sum_j \lambda_j \mu_j\right).$$

so we see that Y is Gaussian with mean $\mu_Y = \sum_j \lambda_j \mu_j$ and variance $\sigma_Y^2 = \sum_{j,l} \lambda_j c_{jl} \lambda_l$.

Second, if Y is Gaussian with mean μ_Y and variance σ_Y^2 then from the sum defining Y we have

$$\mu_Y = \mathrm{E}\left[\sum_j \lambda_j X_j\right] = \sum_j \lambda_j \mu_j$$

and

$$\sigma_Y^2 = \mathrm{E}\left[\left(\sum_j \lambda_j X_j - \sum_j \lambda_j \mu_j\right)^2\right] = \mathrm{E}\left[\sum_{j,l} \lambda_j \lambda_l (X_j - \mu_j)(X_l - \mu_l)\right] = \sum_{j,l} \lambda_j \lambda_l c_{j,l}$$

so the characteristic function of Y can be written as

$$\phi_Y(\kappa) = \mathrm{E}[\mathrm{e}^{\mathrm{i}\kappa Y}] = \mathrm{e}^{-\frac{1}{2}\kappa^2\sigma_Y^2 + \mathrm{i}\kappa\mu_Y} = \exp\left(-\frac{1}{2}\sum_{j,l} \kappa\lambda_j c_{jl} \kappa\lambda_l + \mathrm{i}\sum_j \kappa\lambda_j \mu_j\right)$$

which is the characteristic function of a multivariate Gaussian variable by the association $k_j = \kappa\lambda_j$, so $\boldsymbol{X}$ is multivariate Gaussian.

There also exist a multivatiate complex Gaussian distribution, which is not presented here. More properties of joint multivariate normal variables can be found in Appendix A of Øksendal (2003) and in Papoulis & Pillai (2002).

4.5.2 Exercises

1. Let A and B be bivariate Gaussian stochastic variables, and let $X = \alpha A + \beta B$. Show that X is Gaussian and find its mean and variance.

4.6 The Central Limit Theorem

Let us form a new stochastic variable Y taking the superposition of n statistically independent variables X_j with mean values μ_j and variances σ_j^2, respectively,

$$Y = X_1 + X_2 + \ldots + X_n.$$

The assumption of statistical independence means that the joint probability density function can be factorized

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \ldots f_n(x_n).$$

The expected value of Y is

$$\mathrm{E}[Y] = \mathrm{E}\left[\sum_j X_j\right] = \sum_j \mathrm{E}[X_j] = \sum_j \mu_j = \mu.$$

The variance of Y is

$$\begin{aligned}\mathrm{Var}[Y] = \mathrm{E}[(Y - \mathrm{E}[Y])^2] &= \mathrm{E}\left[\left(\sum_j (X_j - \mu_j)\right)^2\right] \\ &= \sum_j \mathrm{E}[(X_j - \mu_j)^2] + \sum_{j \neq l} \mathrm{E}[(X_j - \mu_j)(X_l - \mu_l)] = \sum_j \sigma_j^2 = \sigma^2\end{aligned}$$

where the sum over all $j \neq l$ vanishes because the X_j are statistical independent (here it is enough that they are mutually uncorrelated). We have defined μ as the sum of the means and σ^2 as the sum of the variances.

We introduce the transformed variable

$$Z = \frac{Y - \mu}{\sigma} = \sum_{j=1}^{n} \frac{X_j - \mu_j}{\sigma} = \sum_{j=1}^{n} Z_j$$

where $Z_j = \frac{X_j - \mu_j}{\sigma}$. It can be shown that $\mathrm{E}[Z] = 0$ and $\mathrm{E}[Z^2] = 1$. We do not assume that the distribution of X_j is known, however we can write down the first few terms of the power series expansion of the characteristic function of Z_j

$$\phi_{Z_j}(k) = \mathrm{E}[\mathrm{e}^{\mathrm{i}k\frac{X_j - \mu_j}{\sigma}}] = \sum_{l=0}^{\infty} \mathrm{i}^l \frac{k^l}{l!\sigma^l} \mathrm{E}[(X_j - \mu_j)^l] = 1 - \frac{k^2\sigma_j^2}{2\sigma^2} - \mathrm{i}\frac{k^3\sigma_j^3}{3!\sigma^3}\gamma_j + \frac{k^4\sigma_j^4}{4!\sigma^4}\kappa_j + \ldots$$

where γ_j and κ_j are the skewness and kurtosis. Notice that if the distributions of X_j are somewhat similar, then we can estimate the asymptotic behavior $\frac{\sigma_j}{\sigma} \propto \frac{1}{\sqrt{n}}$ as $n \to \infty$.

The characteristic function for Z has a power series expansion

$$\begin{aligned}\phi_Z(k) = \mathrm{E}[\mathrm{e}^{\mathrm{i}kZ}] &= \mathrm{E}\left[\mathrm{e}^{\mathrm{i}k\sum_{j=1}^{n}\frac{X_j - \mu_j}{\sigma}}\right] = \mathrm{E}\left[\prod_{j=1}^{n} \mathrm{e}^{\mathrm{i}k\frac{X_j - \mu_j}{\sigma}}\right] \\ &= \prod_{j=1}^{n} \mathrm{E}\left[\mathrm{e}^{\mathrm{i}k\frac{X_j - \mu_j}{\sigma}}\right] = \prod_{j=1}^{n}\left(1 - \frac{k^2\sigma_j^2}{2\sigma^2} + \ldots\right) = (1 - \frac{k^2}{2n})^n + R\end{aligned}$$

where the interchange of the product and the expected value operator depends on statistical independence. Here R is a remainder which we anticipate is proportional to $\frac{1}{\sqrt{n}}$ for the case that the distributions of X_j are somewhat similar.

Now we let $n \to \infty$, we disregard the remainder term R because it vanishes, and we recall the limit

$$\left(1 - \frac{x}{n}\right)^n \to \mathrm{e}^{-x} \qquad \text{as } n \to \infty.$$

Therefore it follows that

$$\phi_Z(k) \to \mathrm{e}^{-\frac{k^2}{2}}$$

and thus the asymptotic probability density of the transformed variable Z is

$$f(z) \to \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{z^2}{2}},$$

which is the Gaussian distribution with mean 0 and variance 1.

If all the variables X_j are equally distributed it becomes particularly simple to demonstrate that $R = O(n^{-1/2})$. If the X_j are not equally distributed, sufficient conditions for the vanishing of R could depend on certain conditions on the variances σ_j^2 and higher moments of X_j, see e.g. Papoulis & Pillai (2002).

4.6.1 Exercises

1. Let X_1 and X_2 be two identically distributed statistically independent variables with mean $\mu_X = 0$, variance σ_X^2, skewness $\gamma_X \neq 0$ and kurtosis $\kappa_X \neq 3$. Let $Y = X_1 + X_2$. Compute the mean μ_Y, variance σ_Y^2, skewness γ_Y and kurtosis κ_Y of Y. Can we say that Y is closer to Gaussian than X_1?

4.7 Stochastic processes

A *stochastic process* $X(t)$ along the time axis t is a collection of stochastic variables $X(t_1), X(t_2), \ldots$, for any selection of times $t_1, t_2, \ldots$. It is also possible to talk about a *stochastic field* $X(\boldsymbol{r})$ where $\boldsymbol{r}$ is a spatial position vector. In the most general case we will allow both spatial and temporal dependence.

The set of possible values of t or $\boldsymbol{r}$ is the so-called *index set* of the process. Given that time t and space $\boldsymbol{r}$ are continuous, there is not only an infinite number of index values, they cannot even be counted. We may fear that a description of the stochastic process $X(t)$ requires consideration of the joint statistics of an uncountable infinity of distinct times. Fortunately, it turns out that it is enough, at least for the processes of interest to us, only to consider the joint statistics for a finite number of times. Thus the stochastic process will be completely described by the multivariate distribution of $\boldsymbol{X} = (X_1, \ldots, X_n)^T$, with $X_j = X(t_j)$, for finite n. The joint distribution of a process at n fixed times is called the *n-th order distribution*.

The *first order distribution* for a real process $X(t)$ is

$$F(x_1; t_1) = P\{X(t_1) \leq x_1\} \qquad \text{and} \qquad f(x_1; t_1) = \frac{\partial F(x_1; t_1)}{\partial x_1}.$$

The *second order distribution* for a real process $X(t)$ is

$$F(x_1, x_2; t_1, t_2) = P\{X(t_1) \leq x_1 \text{ and } X(t_2) \leq x_2\}$$

and

$$f(x_1, x_2; t_1, t_2) = \frac{\partial^2 F(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2}.$$

Several compatibility relations follow, e.g. $F(x_1; t_1) = F(x_1, \infty; t_1, t_2)$, etc.

In principle we can proceed to derive the *n-th order distribution* for the joint behavior at n fixed times $F(x_1, x_2, \ldots, x_n; t_1, t_2, \ldots, t_n)$, however, the first and second order distributions will suffice in the following.

For a complex process $Z(t) = X(t) + \mathrm{i}Y(t)$ we need to consider the joint distribution of the real and imaginary parts, thus the first order distribution is

$$F(x_1, y_1; t_1) = P\{X(t_1) \leq x_1 \text{ and } Y(t_1) \leq y_1\} \quad \text{and} \quad f(x_1, y_1; t_1) = \frac{\partial^2 F(x_1, y_1; t_1)}{\partial x_1 \partial y_1}$$

and similarly for the second order distribution $F(x_1, y_1, x_2, y_2; t_1, t_2)$.

Care should be taken not to be confused by our double use of the word "order": The order of a distribution of a stochastic process refers to the number of simultaneous values from the index set employed for joint statistics. The order of nonlinearity refers to the power of the steepness that measures the importance of wave-wave interactions.

The expected value of a real stochastic process is

$$\mu(t) = \mathrm{E}[X(t)] = \int_{-\infty}^{\infty} x f(x; t)\, \mathrm{d}x.$$

The *autocorrelation function* of a real process is

$$R(t_1, t_2) = \mathrm{E}[X(t_1)X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; t_1, t_2)\, \mathrm{d}x_1\, \mathrm{d}x_2.$$

For the complex stochastic process $Z(t) = X(t) + \mathrm{i}Y(t)$ the expected value is

$$\mu(t) = \mathrm{E}[Z(t)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + \mathrm{i}y) f(x, y; t)\, \mathrm{d}x\mathrm{d}y$$

and the *autocorrelation function* is

$$R(t_1, t_2) = \mathrm{E}[Z(t_1)Z^*(t_2)] = \int_{-\infty}^{\infty} (x_1 + \mathrm{i}y_1)(x_2 - \mathrm{i}y_2) f(x_1, y_1, x_2, y_2; t_1, t_2)\, \mathrm{d}x_1\, \mathrm{d}y_1 \mathrm{d}x_2\, \mathrm{d}y_2$$

(where there should have been four integrals).

The *mean power* of the process is defined as the second moment $R(t, t) = \mathrm{E}[|Z(t)|^2]$.

The *autocovariance function* is defined as

$$C(t_1, t_2) = \mathrm{E}[(Z(t_1) - \mu(t_1))(Z(t_2) - \mu(t_2))^*] = R(t_1, t_2) - \mu(t_1)\mu^*(t_2).$$

A process is said to be *steady state* or *stationary* if the statistical properties are independent of translation of the origin, i.e. $Z(t)$ and $Z(t+\tau)$ have the same distributions. For the first order distribution we need $f(z_1;t_1) = f(z_1)$. For the second order distribution we need $f(z_1,z_2;t_1,t_2) = f(z_1,z_2;\tau)$ where $\tau = t_1 - t_2$. Similarly, the distributions at any higher order should only depend on the time intervals and not the absolute times.

A process is said to be *weakly stationary* if the expected value is constant with respect to time $\mathrm{E}[Z(t)] = \mu$ and the autocorrelation function is independent of translation of the origin $R(t_1,t_2) = \mathrm{E}[Z(t_1)Z^*(t_2)] = R(\tau)$ where $\tau = t_1 - t_2$.

A process $Z(t)$ is said to be *ergodic* (with respect to time-averaging) for the computation of some function $g(Z(t))$ if ensemble-averaging gives the same result as time-averaging, e.g.

$$\mathrm{E}[g(Z(t))] = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g(z(t))\,\mathrm{d}t$$

where $z(t)$ is any realization of $Z(t)$.

It is obvious that ergodicity is only meaningful provided the process has some kind of stationarity. Ergodicity with respect to first and second order statistics, such as the mean and the autocorrelation, is meaningful for a weakly stationary process.

Example: A stationary and non-ergodic process

Consider the process $X(t) = A$ where A is uniformly distributed on the interval $[-1,1]$. We know that the probability density of A is $f(a) = \frac{1}{2}$ for $-1 \le a \le 1$ and zero elsewhere. It is obvious that the process is stationary since it does not depend on time. The mean of the process is

$$\mu(t) = \mathrm{E}[A] = \int_{-1}^{1} a\frac{1}{2}\,\mathrm{d}a = 0.$$

Time averaging over a realization gives

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} a\,\mathrm{d}t = a$$

for some a in the interval $[-1,1]$, which is almost certainly not the mean of the process. Therefore the process is not ergodic.

Example: A weakly stationary and ergodic process

Consider the process $X(t) = \cos(\omega t + \Theta)$ where ω is any real number and where Θ is uniformly distributed on the interval $[0, 2\pi]$. We know that the probability density of Θ is $f(\theta) = \frac{1}{2\pi}$ for $0 \le \theta \le 2\pi$ and zero elsewhere. The mean of the process is

$$\mu(t) = \mathrm{E}[\cos(\omega t + \Theta)] = \int_0^{2\pi} \cos(\omega t + \theta)\frac{1}{2\pi}\,\mathrm{d}\theta = 0$$

and the autocorrelation of the process is

$$\begin{aligned} R(t_1,t_2) &= \mathrm{E}[\cos(\omega t_1 + \Theta)\cos((\omega t_2 + \Theta)] \\ &= \frac{1}{2}\,\mathrm{E}[\cos(\omega(t_1+t_2) + 2\Theta)] + \frac{1}{2}\cos(\omega(t_1 - t_2)) = \frac{1}{2}\cos(\omega\tau) \end{aligned}$$

where we have used the identity $\cos u \cos v = \frac{1}{2}(\cos(u+v) + \cos(u-v))$ and we have set $\tau = t_1 - t_2$. Thus we conclude that the process is weakly stationary.

Time averaging over a realization for $\omega \neq 0$ also gives the mean and the autocorrelation

$$\mu = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \cos(\omega t + \theta)\, \mathrm{d}t = 0$$

and

$$\begin{aligned} R(\tau) &= \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \cos(\omega(t+\tau) + \theta) \cos(\omega t + \theta)\, \mathrm{d}t \\ &= \lim_{T\to\infty} \frac{1}{4T} \int_{-T}^{T} \cos(\omega(2t+\tau) + 2\theta) + \cos(\omega\tau)\, \mathrm{d}t = \frac{1}{2}\cos(\omega\tau) \end{aligned}$$

for some θ in the interval $[0, 2\pi]$. Thus we conclude that the process is ergodic for $\omega \neq 0$ for the computation of mean and autocorrelation.

For $\omega = 0$ the process is not ergodic since we cannot achieve the mean or the autocorrelation by time averaging of a single realization. In this case the process in this example is not the same as in the previous example since $\cos\Theta$ is not uniformly distributed on $[-1, 1]$.

Example: Gaussian process

A process $X(t)$ is said to be Gaussian if and only if for every finite set of times $t_1, \ldots, t_n$ the stochastic variables $X(t_1), \ldots, X(t_n)$ are multivariate Gaussian. We have seen that this is equivalent to saying that every linear combination $Y = \sum_{j=1}^{n} \lambda_j X(t_j)$ is univariate Gaussian for any choice of λ_j. We showed that these two statements are equivalent in the last example of the previous section.

4.7.1 Exercises

1. Consider the real stochastic process $X(t) = a\cos(\omega t + \Theta)$ where Θ is uniformly distributed on $[0, 2\pi]$ and a is a constant. Compute the skewness and the kurtosis of the process. Determine if this is a Gaussian process.

2. Consider the real stochastic process $X(t) = A\cos(\omega t) + B\sin(\omega t)$ where A and B are bivariate stochastic variables (A and B being uncorrelated, statistically independent or equally distributed are special cases). Determine if this is a weakly stationary process, or find some criterion for the bivariate distribution of A and B for it to be weakly stationary.

3. Consider the real stochastic process $X(t) = A\cos(\omega t) + B\sin(\omega t)$ where A and B are bivariate Gaussian stochastic variables (A and B being statistically independent or equally distributed are special cases). Determine if this is a Gaussian process.

4. Consider the complex stochastic process $X(t) = a\mathrm{e}^{-\mathrm{i}(\omega t + \Theta)}$ where Θ is uniformly distributed on $[0, 2\pi]$. Determine if this process is weakly stationary, and if it is ergodic for computation of the mean and the autocorrelation function.

4.8 Properties of the autocorrelation function for a weakly stationary processes

In order to proceed we need to impose some kind of stationarity. We limit to weakly stationary processes, and thus we assume that the mean is a constant

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

and the autocorrelation is only a function of the time difference

$$R(\tau) = \mathrm{E}[X(t+\tau)X^*(t)]. \tag{3}$$

We shall use the convention that the time lag τ should be added to the first factor while the second factor should be complex conjugated, always keeping in mind that process can be complex.

In the following we describe some properties of the autocorrelation function for weakly stationary processes:

- For a complex process $R(-\tau) = R^*(\tau)$, and for a real process $R(-\tau) = R(\tau)$,

  $$R(-\tau) = \mathrm{E}[Z(t-\tau)Z^*(t)] = \mathrm{E}[Z(t)Z^*(t+\tau)] = R^*(\tau).$$

- $R(0)$ is real and non-negative,

  $$R(0) = \mathrm{E}[Z(t)Z^*(t)] = \mathrm{E}[|Z(t)|^2] \geq 0.$$

  The quantity $R(0)$ is often called the *mean power* of the process, and we recognize it as the variance of the process in the case of zero expected value $\mu = 0$.

- $R(0) \geq |R(\tau)|$

  This is most easily seen by considering $U = X(t+\tau)$ and $V = X(t)$ as two vectors, and introducing the inner product $\langle U, V\rangle = \mathrm{E}[UV^*]$. Then the above statement is simply the Cauchy–Schwartz inequality.

- If $X(t)$ and $Y(t)$ are uncorrelated processes with zero mean, and $Z(t) = X(t) + Y(t)$, then

  $$R_{ZZ}(\tau) = \mathrm{E}[(X(t+\tau) + Y(t+\tau))(X^*(t) + Y^*(t))] = R_{XX}(\tau) + R_{YY}(\tau).$$

  Also recall that statistically independent processes are uncorrelated.

4.9 The power spectrum or variance spectrum

The *power spectrum* or *variance spectrum* $S(\omega)$ of a weakly stationary process is defined as the Fourier transform of the autocorrelation function $R(\tau)$.[3]

[3] In mathematics the word "spectrum" means the set of all eigenvalues of an operator. Sometimes the word "spectrum" is used as a synonym for Fourier transform, being interpreted as eigenvalues with corresponding complex exponential functions as eigenvectors. Sometimes we are sloppy and say "spectrum" when we should have said "power spectrum" or "variance spectrum".

Whereas the Fourier transform in general is undetermined by a multiplicative constant, we shall require the spectrum to be uniquely defined by the constraint that the integral of the spectrum over the domain of the frequency axis should be equal to the mean power of the process.

$$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(\tau) \mathrm{e}^{\mathrm{i}\omega\tau} \, \mathrm{d}\tau, \tag{4}$$

$$R(\tau) = \int_{-\infty}^{\infty} S(\omega) \mathrm{e}^{-\mathrm{i}\omega\tau} \, \mathrm{d}\omega, \tag{5}$$

and we notice that the normalization criterion is satisfied

$$R(0) = \int_{-\infty}^{\infty} S(\omega) \, \mathrm{d}\omega.$$

The Fourier transform pair (4)–(5) is known as the Wiener–Khintchine relations. The power spectrum $S(\omega)$ has the following three important properties:

- $S(\omega)$ is real.

  In order to show this recall that $R(-\tau) = R^*(\tau)$, and let us split the integral

  $$S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(\tau) \mathrm{e}^{\mathrm{i}\omega\tau} \, \mathrm{d}\tau = \frac{1}{2\pi} \int_{-\infty}^{0} R(\tau) \mathrm{e}^{\mathrm{i}\omega\tau} + \frac{1}{2\pi} \int_{0}^{\infty} R(\tau) \mathrm{e}^{\mathrm{i}\omega\tau} \, \mathrm{d}\tau.$$

  If we apply the substitution $\tau \to -\tau$ and apply the complex symmetry of $R(\tau)$ in the first integral, we get

  $$S(\omega) = \frac{1}{2\pi} \int_{0}^{\infty} R^*(\tau) \mathrm{e}^{-\mathrm{i}\omega\tau} + \frac{1}{2\pi} \int_{0}^{\infty} R(\tau) \mathrm{e}^{\mathrm{i}\omega\tau} \, \mathrm{d}\tau$$

  so clearly $S(\omega)$ is real.

- For a real process $S(\omega)$ is even.

  In order to show this recall that in this case the autocorrelation function is real and even $R(-\tau) = R(\tau)$ and thus we may write

  $$S(-\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(\tau) \mathrm{e}^{-\mathrm{i}\omega\tau} \, \mathrm{d}\tau = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(-\tau) \mathrm{e}^{-\mathrm{i}\omega\tau} \, \mathrm{d}\tau$$

  and with the substitution $\tau \to -\tau$ we get

  $$S(-\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(\tau) \mathrm{e}^{\mathrm{i}\omega\tau} \, \mathrm{d}\tau = S(\omega).$$

- $S(\omega)$ is nonnegative.

  In order to show this we can introduce

  $$\hat{X}_T(\omega) = \frac{1}{2\pi} \int_{-T}^{T} X(t) \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}t$$

and

$$S_T(\omega) = \mathrm{E}\left[\frac{\pi}{T}|\hat{X}_T(\omega)|^2\right].$$

Clearly, $S_T(\omega) \geq 0$.

We can now do the following manipulations

$$S_T(\omega) = \mathrm{E}\left[\frac{\pi}{T}\frac{1}{2\pi}\int_{-T}^{T} X(t)\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}t\frac{1}{2\pi}\int_{-T}^{T} X^*(s)\mathrm{e}^{-\mathrm{i}\omega s}\,\mathrm{d}s\right]$$

Writing this as a double integral, we let the expected value operator act on the integrand and make use of weak stationarity, $\mathrm{E}[X(t)X^*(s)] = R(t-s)$, we may introduce the new integration variables

$$\left\{\begin{array}{l}\tau = t-s\\ \xi = t+s\end{array}\right\} \quad \text{or} \quad \left\{\begin{array}{l}t = \frac{1}{2}(\xi+\tau)\\ s = \frac{1}{2}(\xi-\tau)\end{array}\right\} \quad \text{with Jacobian} \quad \frac{\partial(t,s)}{\partial(\tau,\xi)} = \begin{vmatrix}\frac{1}{2} & \frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{vmatrix} = \frac{1}{2}$$

such that

$$\begin{aligned}S_T(\omega) &= \frac{\pi}{(2\pi)^2 T}\int_{-T}^{T}\int_{-T}^{T}\mathrm{E}[X(t)X^*(s)]\mathrm{e}^{\mathrm{i}\omega(t-s)}\,\mathrm{d}t\mathrm{d}s\\ &= \frac{1}{8\pi T}\int_{-2T}^{2T}\int_{-2T+|\tau|}^{2T-|\tau|} R(\tau)\mathrm{e}^{\mathrm{i}\omega\tau}\,\mathrm{d}\xi\mathrm{d}\tau = \frac{1}{2\pi}\int_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)R(\tau)\mathrm{e}^{\mathrm{i}\omega\tau}\,\mathrm{d}\tau.\end{aligned}$$

Finally taking the limit we get

$$\lim_{T\to\infty} S_T(\omega) = S(\omega) \geq 0.$$

There is another more heuristic approach to arrive at the same result. Starting with the right-hand side of (4), using (3) and substituting $X(t+\tau)$ and $X^*(t)$ by their Fourier transforms, we have the expression

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{E}\left[\int_{-\infty}^{\infty}\hat{X}(\omega_1)\mathrm{e}^{-\mathrm{i}\omega_1(t+\tau)}\,\mathrm{d}\omega_1\int_{-\infty}^{\infty}\hat{X}^*(\omega_2)\mathrm{e}^{\mathrm{i}\omega_2 t}\,\mathrm{d}\omega_2\right]\mathrm{e}^{\mathrm{i}\omega\tau}\,\mathrm{d}\tau$$

which we bravely rewrite into the form

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{E}[\hat{X}(\omega_1)\hat{X}^*(\omega_2)]\mathrm{e}^{-\mathrm{i}(\omega_1-\omega_2)t}\int_{-\infty}^{\infty}\frac{1}{2\pi}\mathrm{e}^{\mathrm{i}(\omega-\omega_1)\tau}\,\mathrm{d}\tau\,\mathrm{d}\omega_1\,\mathrm{d}\omega_2.$$

Here we notice that the innermost τ-integral is $\delta(\omega-\omega_1)$, and we notice that for the process to be weakly stationary it is necessary that $\mathrm{E}[\hat{X}(\omega_1)\hat{X}^*(\omega_2)] = 0$ for $\omega_1 \neq \omega_2$. We arrive at the desired result making the identification

$$\mathrm{E}[\hat{X}(\omega_1)\hat{X}^*(\omega_2)] = S(\omega_1)\delta(\omega_1-\omega_2)$$

from which it follows that $S(\omega) \geq 0$. Two important remarks must be made regarding this equation. First, the Dirac delta has physical dimension inverse of its argument. Second, even if the Fourier transform of $X(t)$ may contain a Dirac delta, the above equation does not imply the product of two Dirac deltas of the same argument.

4.9.1 One-sided spectrum

The fact that a real process has an even spectrum is often used to introduce a *one-sided spectrum* for non-negative frequencies. The desired Fourier transform pair is then

$$S_{\text{one-sided}}(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} R(\tau)\mathrm{e}^{\mathrm{i}\omega\tau}\,\mathrm{d}\tau = \frac{2}{\pi}\int_{0}^{\infty} R(\tau)\cos(\omega\tau)\,\mathrm{d}\tau, \tag{6}$$

$$R(\tau) = \int_{0}^{\infty} S_{\text{one-sided}}(\omega)\cos(\omega\tau)\,\mathrm{d}\omega. \tag{7}$$

Here we have

$$S_{\text{one-sided}}(\omega) = \begin{cases} 2S(\omega) & \text{for} \quad \omega > 0 \\ S(0) & \text{for} \quad \omega = 0 \end{cases}$$

We notice that the normalization criterion is satisfied

$$\int_{0}^{\infty} S_{\text{one-sided}}(\omega)\,\mathrm{d}\omega = R(0).$$

4.9.2 Power spectrum of continuous periodic process

Let us consider how this becomes for a process $X(t)$ periodic on the interval $0 \leq t < T$. Suppose we have the Fourier transform pair

$$\hat{X}_n = \frac{1}{T}\int_{0}^{T} X(t)\mathrm{e}^{\mathrm{i}\omega_n t}\,\mathrm{d}t$$

$$X(t) = \sum_{n=-\infty}^{\infty} \hat{X}_n \mathrm{e}^{-\mathrm{i}\omega_n t}$$

where $\omega_n = n\Delta\omega$ where $\Delta\omega = \frac{2\pi}{T}$.

We can let the spectrum and the autocorrelation function be given by the Fourier transform pair

$$S_n = \frac{1}{2\pi}\int_{0}^{T} R(\tau)\mathrm{e}^{\mathrm{i}\omega_n\tau}\,\mathrm{d}\tau$$

$$R(\tau) = \Delta\omega \sum_{n=-\infty}^{\infty} S_n \mathrm{e}^{-\mathrm{i}\omega_n\tau}$$

Notice that the normalization criterion is satisfied

$$R(0) = \sum_{n=-\infty}^{\infty} S_n\,\Delta\omega$$

where this sum can be considered as an approximation to the area under the spectral curve. Recall that the criterion for S_n and $R(\tau)$ to be a Fourier transform pair is that $\frac{\Delta\omega}{2\pi} = \frac{1}{T}$, which is obviously satisfied. However, also recall that while the Fourier transform in general is undetermined by a multiplicative constant, the spectrum becomes uniquely defined by the constraint that the integral of the spectrum over the domain of the frequency axis should be equal to the mean power of the process. Hence our particular choice of using $\Delta\omega$ for the normalization.

Let us investigate further the consequences of $X(t)$ being weakly stationary. The mean should be constant

$$\mu = \mathrm{E}[X(t)] = \mathrm{E}\left[\sum_n \hat{X}_n \mathrm{e}^{-\mathrm{i}\omega_n t}\right] = \sum_n \mathrm{E}[\hat{X}_n]\mathrm{e}^{-\mathrm{i}\omega_n t}$$

so weak stationarity requires $\mathrm{E}[\hat{X}_n] = 0$ for $n \neq 0$, therefore we have $\mathrm{E}[\hat{X}_n] = \mu\delta_{n,0}$. Similarly for the autocorrelation function

$$\begin{aligned} R(\tau) = \mathrm{E}[X(t+\tau)X^*(t)] = \mathrm{E}\left[\sum_n \hat{X}_n \mathrm{e}^{-\mathrm{i}\omega_n(t+\tau)} \sum_m \hat{X}_m^* \mathrm{e}^{\mathrm{i}\omega_m t}\right] \\ = \sum_{n,m} \mathrm{E}[\hat{X}_n \hat{X}_m^*]\mathrm{e}^{\mathrm{i}(\omega_m-\omega_n)t - \mathrm{i}\omega_n\tau} \end{aligned}$$

so weak stationarity requires $\mathrm{E}[\hat{X}_n\hat{X}_m^*] = 0$ for $n \neq m$. Comparison with the above reveals that $\mathrm{E}[\hat{X}_n\hat{X}_m^*] = S_n\delta_{n,m}\Delta\omega$, or alternatively

$$S_n = \frac{1}{\Delta\omega}\,\mathrm{E}[|\hat{X}_n|^2]. \tag{8}$$

If $X(t)$ is real we may employ the one-sided spectrum

$$S_{\text{one-sided},n} = \begin{cases} \frac{2}{\Delta\omega}\,\mathrm{E}[|\hat{X}_n|^2] & \text{for} \quad n > 0 \\ \frac{1}{\Delta\omega}\,\mathrm{E}[|\hat{X}_0|^2] & \text{for} \quad n = 0 \end{cases}$$

for which the normalization criterion is satisfied

$$R(0) = \sum_{n=0}^{\infty} S_{\text{one-sided},n}\,\Delta\omega$$

4.9.3 Power spectrum of discrete periodic process

Let us consider how this becomes for a finite length discrete process $X_j = X(t_j)$ periodic on the interval $0 \leq j < N$. Suppose we have the Fourier transform pair

$$\hat{X}_n = \frac{1}{N}\sum_{j=0}^{N-1} X_j \mathrm{e}^{\frac{2\pi\mathrm{i}nj}{N}}$$

$$X_j = \sum_{n=0}^{N-1} \hat{X}_n \mathrm{e}^{-\frac{2\pi\mathrm{i}nj}{N}}$$

where we may think that the exponent is $\omega_n t_j$ where $\omega_n = n\Delta\omega$ where $\Delta\omega = \frac{2\pi}{T}$ and $t_j = j\Delta T$ where $\Delta T = \frac{T}{N}$.

We can let the spectrum and the autocorrelation function be given by the Fourier transform pair

$$S_n = \frac{1}{N}\sum_{j=0}^{N-1} R_j \mathrm{e}^{\frac{2\pi\mathrm{i}nj}{N}}$$

$$R_j = \sum_{n=0}^{N-1} S_n \mathrm{e}^{-\frac{2\pi i n j}{N}}$$

Notice that the normalization criterion is satisfied

$$R_0 = \sum_{n=0}^{N-1} S_n$$

where this sum can be considered as an approximation to the area under the spectral curve.

Let us investigate further the consequences of X_j being weakly stationary. The mean should be constant

$$\mu = \mathrm{E}[X_j] = \mathrm{E}\left[\sum_{n=0}^{N-1} \hat{X}_n \mathrm{e}^{-\frac{2\pi i n j}{N}}\right] = \sum_n \mathrm{E}[\hat{X}_n] \mathrm{e}^{-\frac{2\pi i n j}{N}}$$

so weak stationarity requires $\mathrm{E}[\hat{X}_n] = 0$ for $n \neq 0$, therefore we have $\mathrm{E}[\hat{X}_n] = \mu\delta_{n,0}$. Similarly for the autocorrelation function

$$R_j = \mathrm{E}[X_{l+j} X_l^*] = \mathrm{E}\left[\sum_n \hat{X}_n \mathrm{e}^{-\frac{2\pi i n (l+j)}{N}} \sum_m \hat{X}_m^* \mathrm{e}^{\frac{2\pi i m l}{N}}\right]$$
$$= \sum_{n,m} \mathrm{E}[\hat{X}_n \hat{X}_m^*] \mathrm{e}^{\frac{2\pi \mathrm{i}(m-n)l}{N} - \frac{2\pi i n j}{N}}$$

so weak stationarity requires $\mathrm{E}[\hat{X}_n \hat{X}_m^*] = 0$ for $n \neq m$. Comparison with the above reveals that $\mathrm{E}[\hat{X}_n \hat{X}_m^*] = S_n \delta_{n,m}$, or alternatively

$$S_n = \mathrm{E}[|\hat{X}_n|^2]. \tag{9}$$

If X_j is real we may employ the one-sided spectrum (for $0 \leq n \leq N/2$)

$$S_{\text{one-sided},n} = \begin{cases} \mathrm{E}[|\hat{X}_0|^2] & \text{for} \quad n = 0 \\ 2\,\mathrm{E}[|\hat{X}_n|^2] & \text{for} \quad 0 < n < N/2 \\ \mathrm{E}[|\hat{X}_n|^2] & \text{for} \quad n = N/2 \quad \text{for } n \text{ even} \end{cases}$$

for which the normalization criterion is satisfied

$$R_0 = \sum_{n=0}^{\lfloor N/2 \rfloor} S_{\text{one-sided},n}$$

4.9.4 Exercises

1. Consider the complex stochastic process $X(t) = a\mathrm{e}^{-\mathrm{i}(\omega_0 t + \Theta)}$ where Θ is uniformly distributed on $[0, 2\pi]$. Compute the Fourier transform $\hat{X}(\omega)$, the mean $\mu(t)$, the autocorrelation function $R(t_1, t_2)$, and determine if the process is weakly stationary. If the process is weakly stationary compute the power spectrum $S(\omega)$.

2. Consider the real stochastic process $X(t) = a\cos(\omega_0 t + \Theta)$ where Θ is uniformly distributed on $[0, 2\pi]$. Compute the Fourier transform $\hat{X}(\omega)$, the mean $\mu(t)$, the autocorrelation function $R(t_1, t_2)$, and determine if the process is weakly stationary. If the process is weakly stationary compute the two-sided power spectrum $S(\omega)$ and the one-sided power spectrum.

3. Consider the real stochastic process $X(t) = A\cos(\omega_0 t) + B\sin(\omega_0 t)$ where A and B are independent identically distributed Gaussian stochastic variables with mean 0 and variance σ_0^2. Compute the Fourier transform $\hat{X}(\omega)$, the mean $\mu(t)$, the autocorrelation function $R(t_1, t_2)$, and determine if the process is weakly stationary. If the process is weakly stationary compute the two-sided power spectrum $S(\omega)$ and the one-sided power spectrum.

4.10 Directional wave spectrum

Assume that $\boldsymbol{r}$ is an n-dimensional spatial position vector, and the t is time, and that the process $\eta(\boldsymbol{r}, t)$ is weakly stationary in time and weakly homogeneous in space, we simply say it is weakly stationary.

Example: We can assume a Gaussian wave field of the form

$$\eta(\boldsymbol{r}, t) = \sum_j A_j \cos(\boldsymbol{k}_j \cdot \boldsymbol{r} - \omega_j t) + B_j \sin(\boldsymbol{k}_j \cdot \boldsymbol{r} - \omega_j t)$$

where A_j and B_j are statistically independent Gaussian distributed stochastic variables with mean zero and variance σ_j^2.

The assumption of weak stationarity means that the mean is constant

$$\mu = \mathrm{E}[\eta(\boldsymbol{r}, t)]$$

and the autocorrelation function only depends on the relative position and time

$$R(\boldsymbol{\xi}, \tau) = \mathrm{E}[\eta(\boldsymbol{r} + \boldsymbol{\xi}, t + \tau)\eta^*(\boldsymbol{r}, t)].$$

The $(n+1)$-dimensional wave spectrum is

$$F^{(n+1)}(\boldsymbol{k}, \omega) = \frac{1}{(2\pi)^{n+1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R(\boldsymbol{\xi}, \tau) \mathrm{e}^{-\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{\xi} - \omega\tau)} \, \mathrm{d}^n\boldsymbol{\xi}\mathrm{d}\tau.$$

If the original process is real we have the symmetry $F^{(n+1)}(\boldsymbol{k}, \omega) = F^{(n+1)}(-\boldsymbol{k}, -\omega)$ and we may limit attention to half of the $(\boldsymbol{k}, \omega)$-space, e.g. to only positive frequencies, without sacrificing the resolution of the propagation direction of the waves.

It should be noticed that our convention of considering the most basic wave as $\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)}$ ensures that waves going in "positive" direction, i.e. positive direction of time and all the spatial components of $\boldsymbol{r}$, will be located in the first quadrant/octant/etc. of the $(\boldsymbol{k}, \omega)$ space.

The n-dimensional wavenumber vector spectrum is

$$F^{(n)}(\boldsymbol{k}) = \int_{-\infty}^{\infty} F^{(n+1)}(\boldsymbol{k}, \omega) \, \mathrm{d}\omega.$$

For a real process we have the symmetry $F^{(n)}(\boldsymbol{k}) = F^{(n)}(-\boldsymbol{k})$, so clearly we have lost the resolution of the propagation direction of the waves. However, in the case of a real process we can keep the resolution of propagation direction by only integrating over positive frequencies

$$\Psi(\boldsymbol{k}) = F_+^{(n)}(\boldsymbol{k}) = 2\int_0^\infty F^{(n+1)}(\boldsymbol{k},\omega)\,\mathrm{d}\omega$$

where the quantity $\Psi(\boldsymbol{k}) = F_+^{(n)}(\boldsymbol{k})$ may be called the unambiguous directional wavenumber vector "spectrum" although it is not really a spectrum since it lacks the necessary symmetry properties. The factor 2 has been inserted so that the normalization criterion is still satisfied. The two-dimensional wavenumber vector spectrum can then be recovered by the relation $F^{(n)}(\boldsymbol{k}) = \frac{1}{2}\left(F_+^{(n)}(\boldsymbol{k}) + F_+^{(n)}(-\boldsymbol{k})\right)$.

Likewise, the frequency spectrum can be obtained by integrating the $(n+1)$-dimensional wave spectrum over the wavenumber vector

$$S(\omega) = \int_{-\infty}^{\infty} F^{(n+1)}(\boldsymbol{k},\omega)\,\mathrm{d}^n\boldsymbol{k}.$$

Given the above spectra, we may consider changing variables. The most elementary case is to use frequency instead of angular frequency. With the substitution $\omega = 2\pi f$, and insisting that the amount of power in some domain $\mathcal{D}$ remains the same, we have

$$\int_{\mathcal{D}} S(\omega)\,\mathrm{d}\omega = \int_{\mathcal{D}} S(\omega(f))\,2\pi\,\mathrm{d}f = \int_{\mathcal{D}} \tilde{S}(f)\,\mathrm{d}f$$

and we see that $\tilde{S}(f) = 2\pi S(\omega)$.

Limiting to two spatial dimensions, $n = 2$, we may also want to change from Cartesian wavenumber vector components to polar representation with scalar wavenumber k and directional angle θ. Suppose we want to transform the unambiguous wavenumber vector spectrum $\Psi(\boldsymbol{k})$ for a two-dimensional wavenumber vector (k_x, k_y) to polar representation with scalar wavenumber k and direction θ, $k_x = k\cos\theta$ and $k_y = k\sin\theta$. This transformation has Jacobian k. For the unambiguous wavenumber vector spectrum we have

$$\int_{\mathcal{D}} \Psi(\boldsymbol{k})\,\mathrm{d}^2\boldsymbol{k} = \int_{\mathcal{D}} \Psi(\boldsymbol{k}(k,\theta))k\,\mathrm{d}k\mathrm{d}\theta$$

In none of the above manipulations we assumed any relationship between the wave vector and the frequency.

If we now assume the existence of an isotropic dispersion relation $\omega = \omega(k)$, with $k = |\boldsymbol{k}|$, and we furthermore assume that for any given wavenumber k there is only one frequency ω, then we may make the substitution $\omega = \omega(k)$ with $\mathrm{d}\omega = c_g\mathrm{d}k$, where $\boldsymbol{c}_g = \partial\omega/\partial\boldsymbol{k}$ is the group velocity and $c_g = |\boldsymbol{c}_g| = \mathrm{d}\omega/\mathrm{d}k$,

$$\int_{\mathcal{D}} \Psi(\boldsymbol{k})\,\mathrm{d}^2\boldsymbol{k} = \int_{\mathcal{D}} \Psi(\boldsymbol{k}(k(\omega),\theta))\frac{k(\omega)}{c_g}\,\mathrm{d}\omega\mathrm{d}\theta = \int_{\mathcal{D}} E(\omega,\theta)\mathrm{d}\omega\mathrm{d}\theta$$

We thus have derived the directional frequency "spectrum", which is not really a spectrum because it lacks the necessary symmetry,

$$E(\omega,\theta) = \Psi(\boldsymbol{k}(k(\omega),\theta))\frac{k(\omega)}{c_g(\omega)}.$$

It is common to write this in terms of the one-sided frequency spectrum and a directional distribution

$$E(\omega,\theta) = S_{\text{one-sided}}(\omega)D(\theta,\omega)$$

where the directional distribution satisfies the requirements for being a probability distribution $D(\theta,\omega) \geq 0$ and $\int_0^{2\pi} D(\theta,\omega)\,\mathrm{d}\theta = 1$.

4.10.1 Exercises

1. Let $\boldsymbol{r}$ be an n-dimensional position vector, let t be time, and let $\eta(\boldsymbol{r},t)$ be a real and weakly stationary (and weakly homogeneous) stochastic process. Let the autocorrelation function be

   $$R(\boldsymbol{\xi},\tau) = \mathrm{E}[\eta(\boldsymbol{r}+\boldsymbol{\xi},t+\tau)\eta(\boldsymbol{r},t)].$$

   Show that the power spectrum

   $$F^{(n+1)}(\boldsymbol{k},\omega) = \frac{1}{(2\pi)^{n+1}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(\boldsymbol{\xi},\tau)\mathrm{e}^{-\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{\xi}-\omega\tau)}\,\mathrm{d}^n\boldsymbol{\xi}\,\mathrm{d}\tau$$

   has the symmetry

   $$F^{(n+1)}(\boldsymbol{k},\omega) = F^{(n+1)}(-\boldsymbol{k},-\omega).$$

2. Suppose that the unambiguous directional wave spectrum for wind-generated waves is

   $$\Psi(\boldsymbol{k}) = F_+^{(2)}(\boldsymbol{k}) = 2\int_0^{\infty} F^{(2+1)}(\boldsymbol{k},\omega)\,\mathrm{d}\omega = \begin{cases} 0 & \text{for} \quad |\boldsymbol{k}| \leq k_0 \\ A|\boldsymbol{k}|^{-4}\cos^2\frac{\theta}{2} & \text{for} \quad |\boldsymbol{k}| > k_0 \end{cases}$$

   where $\boldsymbol{k} = k_x\boldsymbol{i} + k_y\boldsymbol{j} = k(\boldsymbol{i}\cos\theta + \boldsymbol{j}\sin\theta)$.

   (a) Explain why $\Psi(\boldsymbol{k})$ is not really a spectrum. Hint: Which symmetry property does it not have?

   (b) Compute the mean power of the process from $\Psi(\boldsymbol{k})$.

   (c) Determine the corresponding directional frequency spectrum $E(\omega,\theta)$, the one-sided frequency spectrum $S(\omega)$ and the directional distribution $D(\theta,\omega)$, where we have $E(\omega,\theta) = S(\omega)D(\theta,\omega)$. Do this for deep water with dispersion relation $\omega^2 = gk$ and for shallow water with dispersion relation $\omega^2 = ghk^2$ where g is the acceleration of gravity and h is the depth.

   (d) Show that the mean power of the process can also be computed from $S(\omega)$.

4.11 Distributions of surface elevation

4.11.1 Exercises

1. Let the surface elevation at a given point be given by the real stochastic process $\eta(t) = A\cos(\omega_p t) + B\sin(\omega_p t)$ where A and B are independent Gaussian variables with mean 0 and variance σ^2.

   (a) Show that the surface elevation is a Gaussian process.

   (b) Show that the surface elevation of one particular realization, *i.e.* a time series for one particular choice of A and B, is not Gaussian with respect to time.

   (c) Derive the distribution of the surface elevation for one particular realization (alternatively, you may compute the mean, variance, skewness and kurtosis of the surface elevation for one particular realization with respect to time).

4.12 Distributions of crests and wave heights for narrowband processes

The (relative) bandwidth of the spectrum $S(\omega)$ is a ratio between of the absolute width and the centerpoint. Here we will limit attention to the case that we can approximate the process by only a single wavenumber vector $\boldsymbol{k}_p$ and frequency ω_p.

In the following we will suppose that the process is Gaussian and that it can be approximated by

$$\eta(\boldsymbol{r}, t) = A\cos(\boldsymbol{k}_p \cdot \boldsymbol{r} - \omega_p t) + B\sin(\boldsymbol{k}_p \cdot \boldsymbol{r} - \omega_p t) = R\cos(\boldsymbol{k}_p \cdot \boldsymbol{r} - \omega_p t + \Theta)$$

where A and B are independent Gaussian variables with mean 0 and variance σ^2. The phase Θ is uniformly distributed and the crest height $\eta_c = R = \sqrt{A^2 + B^2}$ is Rayleigh distributed

$$f_R(r) = \frac{r}{\sigma^2}\mathrm{e}^{-\frac{r^2}{2\sigma^2}} \quad \text{for} \quad r \geq 0. \tag{10}$$

Of particular interest is the exceedance probability $P_e(r) = P\{R > r\}$ that the crest height is greater than some threshold

$$P_e(r) = \mathrm{e}^{-\frac{r^2}{2\sigma^2}} \quad \text{for} \quad r \geq 0$$

which is a parabola when plotted with linear first axis and logarithmic second axis.

In the limit that the bandwidth goes to zero, the wave height is twice the crest height, $H = 2R$, and is also Rayleigh distributed

$$f_H(h) = \frac{h}{4\sigma^2}\mathrm{e}^{-\frac{h^2}{8\sigma^2}} \quad \text{for} \quad h \geq 0. \tag{11}$$

Of particular interest is the exceedance probability $P_e(h) = P\{H > h\}$ that the wave height is greater than some threshold

$$P_e(h) = \mathrm{e}^{-\frac{h^2}{8\sigma^2}} \quad \text{for} \quad h \geq 0$$

which is a parabola when plotted with linear first axis and logarithmic second axis.

Let us consider the distribution of the $1/N$ highest waves. The probability density for wave height (11) is shown in figure 7. The threshold height H_* that divides the $1/N$ highest waves from the smaller waves is the solution of

$$\int_{H_*}^{\infty} \frac{z}{4\sigma^2} \mathrm{e}^{-\frac{z^2}{8\sigma^2}} \,\mathrm{d}z = \frac{1}{N}$$

which is $H_* = \sqrt{8 \ln N}\sigma$. The probability distribution for the $1/N$ highest waves is

$$f_{H \geq H_*}(z) = N \frac{z}{4\sigma^2} \mathrm{e}^{-\frac{z^2}{8\sigma^2}} \quad \text{for} \quad z \geq H_*$$

and the mean height of the $1/N$ highest waves is

$$H_{1/N} = N \int_{H_*}^{\infty} \frac{z^2}{4\sigma^2} \mathrm{e}^{-\frac{z^2}{8\sigma^2}} \,\mathrm{d}z = \left[\sqrt{8 \ln N} + \sqrt{2\pi} N \operatorname{erfc} \sqrt{\ln N} \right] \sigma$$

where $\operatorname{erfc} z = \frac{2}{\sqrt{\pi}} \int_z^{\infty} \mathrm{e}^{-t^2} \,\mathrm{d}t$ is the complementary error function. Traditionally the significant wave height was defined as the mean height of the $1/3$ highest waves. If we set $N = 3$ then we get $H_{1/3} = 4.0043\sigma$ which should be compared with the modern definition $H_s = 4\sigma$.

It should be stressed that the relationship between $H_{1/N}$ and σ depends on the distribution of the wave heights. The above derivation assumes Rayleigh distribution with certain parameters appropriate for the limit of zero bandwidth, these assumptions are usually not satisfied, and the actual value of $H_{1/3}$ is typically smaller than the estimate above.

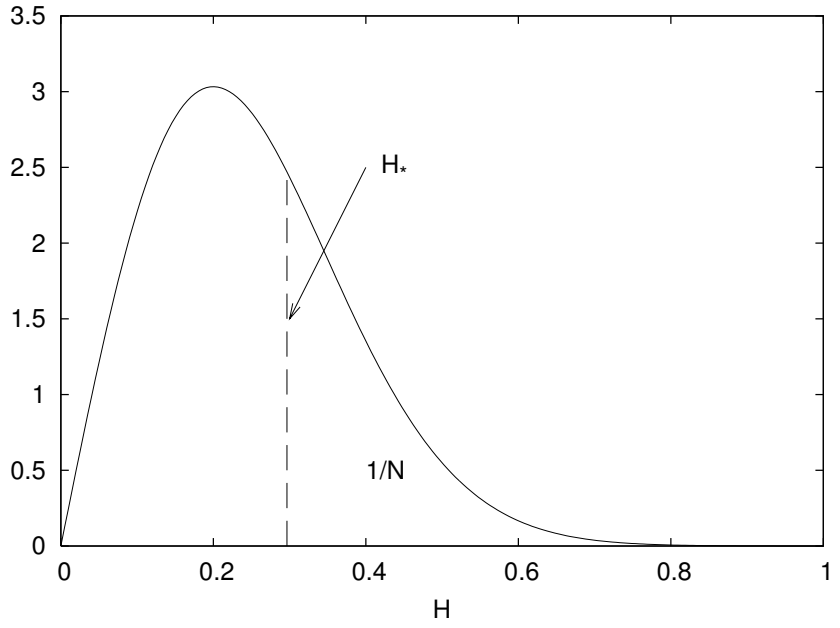

Figure 7: Highest $1/N$ waves, threshold height H_*, definition sketch using Rayleigh distribution ($\sigma = 0.1$, $N = 3$).

4.12.1 Second-order nonlinear narrowbanded waves with Gaussian first harmonic — Tayfun distributions

In order to capture the statistical distribution of nonlinear waves we may distinguish two different types of nonlinear contributions. One is the higher harmonic contribution to the reconstruction of the wave profile. Another is the effect of the nonlinear evolution equations that show up as solvability conditions, such as the nonlinear Schrödinger (NLS) equation. If the second type of contribution can be neglected (corresponding to the Stokes wave), then it is reasonable to assume that the first-order contribution to the harmonic expansion has a Gaussian distribution. Let us make this assumption, and consider the second-order expansion

$$\eta(\boldsymbol{x},t) = \frac{1}{2}\left(B\mathrm{e}^{\mathrm{i}(k_c x-\omega_c t)} + \beta B^2 \mathrm{e}^{2\mathrm{i}(k_c x-\omega_c t)} + c.c.\right)$$

where β is a constant that depends on the depth. For deep water we have $\beta = k_c/2$.

We now recognize that the nonlinear crest heights are

$$\eta_c = |B| + \beta|B|^2$$

and the nonlinear trough depths are

$$\eta_t = -|B| + \beta|B|^2.$$

These formulas are valid only in the case of zero bandwidth, i.e. the case that $|B|$ does not depend on x or t, corresponding to the Stokes wave. The distribution of wave height is the distribution of the distance between upper and lower envelope, $H = \eta_c - \eta_t = 2|B|$, thus the wave height is Rayleigh distributed to second nonlinear order in the limit of vanishing bandwidth.

Given that $|B|$ is Rayleigh distributed, we get the distribution for the second order nonlinear crest height

$$f_{\eta_c}(z) = \frac{1}{2\gamma\sigma^2}\left(1 - \frac{1}{\sqrt{1+4\beta z}}\right)\exp\{\frac{\sqrt{1+4\beta z}-1-2\beta z}{(2\beta\sigma)^2}\} \quad \text{for} \quad z>0. \tag{12}$$

This distribution was first derived by Tayfun (1980).

Using the second-order nonlinear Tayfun distribution (12) the exceedance probability for the crest height is

$$P_e(z) = P\{\eta_c > z\} = \exp\left(\frac{\sqrt{1+4\beta z}-1-2\beta z}{(2\beta\sigma)^2}\right)$$

Probability densities corresponding to the Rayleigh and Tayfun distributions for infinite depth and steepness 0.1 are seen in figure 8.

4.12.2 Exercises

1. (not so easy) Find the Tayfun distribution for wave trough depths η_t in the case of zero bandwidth.

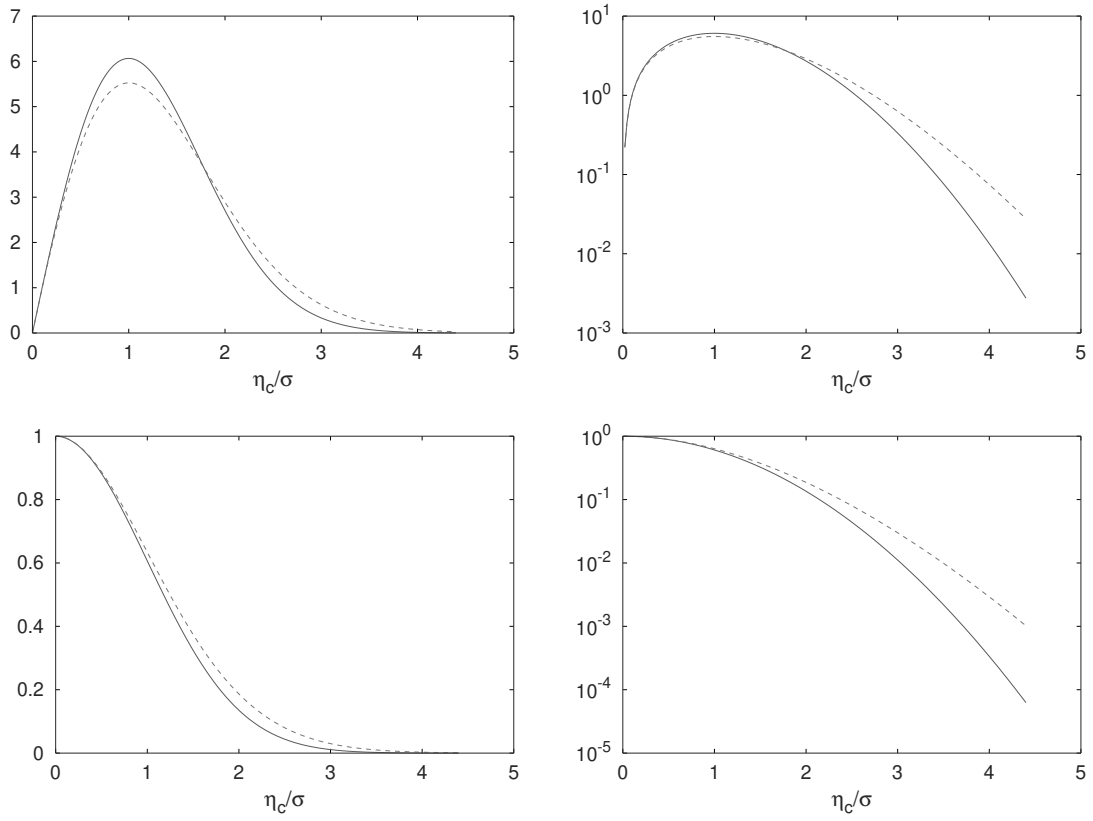

Figure 8: Probability density functions (top) and exceedance probability (bottom) for linear axes (left) and logarithmic second axis (right) for crest height corresponding to the Rayleigh distribution (—) and the second-order nonlinear Tayfun distribution (– –). These plots correspond to infinite depth and steepness 0.1.

4.13 The envelope, Hilbert transform, distribution of the envelope

There are several alternative definitions of the envelope of a process. The discussion below considers the envelope associated with the Hilbert transform. The present discussion is based on information found in Ochi (1998), Papoulis & Pillai (2002) and Lindgren (2013).

First consider an elementary case, the monochromatic oscillation $x(t) = \cos \omega_p t$. The function $x(t)$ is the real part of the complex function $z(t) = \mathrm{e}^{\mathrm{i}\omega_p t} = \cos \omega_p t + \mathrm{i} \sin \omega_p t$. A useful envelope is achieved by computing $|z(t)| = 1$.

Then consider a slightly more complicated case, the bichromatic oscillation $x(t) = \cos \omega_1 t + \cos \omega_2 t$. We have $x(t) = 2 \cos \frac{(\omega_2 - \omega_1)t}{2} \cos \frac{(\omega_2 + \omega_1)t}{2}$, for $\omega_1 \approx \omega_2$ this is the slow beat of a rapid oscillation. The function $x(t)$ is the real part of a complex function $z(t) = \mathrm{e}^{\mathrm{i}\omega_1 t} + \mathrm{e}^{\mathrm{i}\omega_2 t} = \cos \omega_1 t + \cos \omega_2 t + \mathrm{i} \sin \omega_1 t + \mathrm{i} \sin \omega_2 t$. A useful envelope is achieved by computing the absolute value with turns out to be the slow beat $|z(t)| = 2|\cos \frac{(\omega_2 - \omega_1)t}{2}|$.

Closer inspection reveals the reason why the above examples work: The imaginary part of $z(t)$ is a phase-shifted version of the real part of $z(t)$, for positive frequencies the phase-shift is $\pi/2$ in the positive time direction, for negative frequencies the phase-shift is $\pi/2$ in the negative time direction. Thus if $\hat{x}(\omega)$ is the Fourier transform of $x(t)$ then the Fourier transform of the imaginary part of $z(t)$ is $\hat{g}(\omega)\hat{x}(\omega)$ where $\hat{g}(\omega) = -\mathrm{i} \operatorname{sign} \omega$. The imaginary part of $z(t)$ is the Hilbert transform of $x(t)$, which we will denote $\tilde{x}(t)$.

4.13.1 Hilbert transform defined by Fourier transform

Let the Fourier transform pair of the process $X(t)$ be

$$\hat{X}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(t) \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}t$$

$$X(t) = \int_{-\infty}^{\infty} \hat{X}(\omega) \mathrm{e}^{-\mathrm{i}\omega t} \, \mathrm{d}\omega$$

If we assume $X(t)$ is a real process then we know that the Fourier transform is complex conjugate symmetric $\hat{X}(-\omega) = \hat{X}^*(\omega)$, thus it is natural to look for a representation of $X(t)$ using only half of the Fourier integral. The Hilbert transform will be defined as the imaginary part of this half of the Fourier integral.

We write

$$X(t) = \int_{-\infty}^{\infty} \hat{X}(\omega) \mathrm{e}^{-\mathrm{i}\omega t} \, \mathrm{d}\omega = \int_{-\infty}^{0} \hat{X}(\omega) \mathrm{e}^{-\mathrm{i}\omega t} \, \mathrm{d}\omega + \int_{0}^{\infty} \hat{X}(\omega) \mathrm{e}^{-\mathrm{i}\omega t} \, \mathrm{d}\omega$$

For simplicity we have assumed that $\hat{X}(0)$ is finite at the origin.

With the transformation $\omega' = -\omega$ in the first integral, and using the complex conjugate symmetry, we get upon dropping the primes

$$X(t) = \int_{0}^{\infty} \hat{X}^*(\omega) \mathrm{e}^{\mathrm{i}\omega t} \, \mathrm{d}\omega + \int_{0}^{\infty} \hat{X}(\omega) \mathrm{e}^{-\mathrm{i}\omega t} \, \mathrm{d}\omega$$

Defining the complex process

$$Z(t) = X(t) + \mathrm{i}\tilde{X}(t) = 2\int_0^\infty \hat{X}(\omega)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega$$

we define the Hilbert transform of $X(t)$ to be

$$\tilde{X}(t) = \frac{\mathrm{i}}{2}\left(Z(t) - Z^*(t)\right) = -\mathrm{i}\int_0^\infty \hat{X}^*(\omega)\mathrm{e}^{\mathrm{i}\omega t}\,\mathrm{d}\omega + \mathrm{i}\int_0^\infty \hat{X}(\omega)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega = \int_{-\infty}^\infty \hat{g}(\omega)\hat{X}(\omega)\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{d}\omega$$

where we have introduced

$$\hat{g}(\omega) = -\mathrm{i}\,\mathrm{sign}\,\omega \equiv \begin{cases} \mathrm{i} & \text{for} \quad \omega < 0 \\ 0 & \text{for} \quad \omega = 0 \\ -\mathrm{i} & \text{for} \quad \omega > 0 \end{cases}$$

Matlab: The function `hilbert(x)` is equivalent to:

```
n = length(x);
ft = ifft(x);
hilt = ft.*[0; i*ones(n/2-1,1); 0; -i*ones(n/2-1,1)];
hil = x + i*fft(hilt);
```

or

```
n = length(x);
ft = ifft(x);
hilt = ft.*[0; i*ones(ceil(n/2)-1,1); -i*ones(ceil(n/2)-1,1)];
hil = x + i*fft(hilt);
```

for even and odd n, respectively. It should therefore be clear that this function returns our $Z(t)$ and that the Hilbert transform of $X(t)$ is achieved by taking the imaginary part `imag(hilbert(x))`.

4.13.2 Convolution integral

It can be seen that $\tilde{X}(t)$ is given as an inverse Fourier transform of a product of two Fourier transforms. By the convolution theorem

$$\widehat{g(t) * X}(t)(\omega) = 2\pi\hat{g}(\omega)\hat{X}(\omega)$$

we then have

$$\tilde{X}(t) = \frac{1}{2\pi} g(t) * X(t)$$

It can be shown that

$$g(t) = \frac{2}{t}.$$

We then finally have the desired result

$$\tilde{X}(t) = \frac{1}{\pi}\mathcal{P}\int_{-\infty}^\infty \frac{X(\xi)}{t-\xi}\,\mathrm{d}\xi \tag{13}$$

where the symbol $\mathcal{P}$ indicates the Cauchy Principal Value, i.e. the Fourier integral must be evaluated upon approaching the singular point at the origin equally fast from both sides.

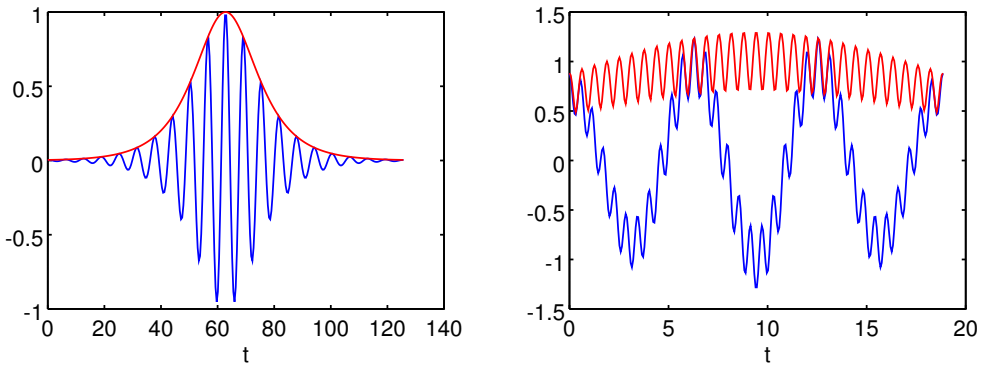

Figure 9: Original time series $X(t)$ (blue) and Hilbert envelope $R(t)$ (red) for a narrowbanded process (left) and a broadbanded process (right).

4.13.3 Analytical signal, Hilbert envelope, total phase and instantaneous frequency

The complex process $Z(t)$ defined above, which is what the Matlab function `hilbert(x)` returns, is known as the *analytical signal* of $X(t)$.

Writing

$$X(t) = \mathcal{E}(t)\cos\psi(t)$$
$$\tilde{X}(t) = \mathcal{E}(t)\sin\psi(t)$$

we say that $\mathcal{E}(t) = \sqrt{X^2(t) + \tilde{X}^2(t)}$ is the *Hilbert envelope* of $X(t)$ and $\psi(t)$ is the *total phase* of $X(t)$. Sometimes we also say that $\frac{\mathrm{d}\psi}{\mathrm{d}t}$ is the *instantaneous angular frequency* of $X(t)$.

Figure 9 should make it clear that the Hilbert envelope touches the crests of a narrowbanded series, but does not touch the crests of a broadbanded series.

4.13.4 Optimum envelope

It can be anticipated from figure 9 that we could have imagined several different "envelopes". For the broadbanded process it appears to be an issue that the Hilbert envelope is necessarily non-negative while the process maybe could have been better characterized by an envelope that could have been negative. It appears to be an issue if the envelope should touch every local crest or if it should only touch some extreme crests.

In Papoulis & Pillai (2002) chapter 10 it is shown that the Hilbert envelope is optimum subject to the criterion that $\mathrm{E}[|Z'(t)|^2]$ should be minimized.

4.13.5 Correlation between $X(t)$ and $\tilde{X}(t)$

Let $X(t)$ be weakly stationary, with zero mean $\mathrm{E}[X(t)] = 0$, and with spectrum $S(\omega)$. We have $\mathrm{E}[\tilde{X}(t)] = 0$. The covariance between $X(t)$ and $\tilde{X}(t)$ is (recall that they are both real)

$$\mathrm{Cov}[X, \tilde{X}] = \mathrm{E}[X(t+\tau)\tilde{X}^*(t)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \hat{g}^*(\omega_2)\,\mathrm{E}[\hat{X}(\omega_1)\hat{X}^*(\omega_2)]\mathrm{e}^{-\mathrm{i}\omega_1(t+\tau)+\mathrm{i}\omega_2 t}\,\mathrm{d}\omega_1\mathrm{d}\omega_2$$

Using the fact that for weakly stationary processes we have $\mathrm{E}[\hat{X}(\omega_1)\hat{X}^*(\omega_2)] = \delta(\omega_1 - \omega_2)S(\omega_1)$ we arrive at

$$\mathrm{Cov}[X, \tilde{X}] = -\int_0^\infty S_{\text{one-sided}}(\omega)\sin(\omega\tau)\,\mathrm{d}\omega$$

This expression evaluates to zero for $\tau = 0$, therefore the process $X(t)$ is uncorrelated with its Hilbert transform $\tilde{X}(t)$ *at the same time.*

4.13.6 Distribution of $\tilde{X}(t)$

We first show that the variance of $\tilde{X}(t)$ is equal to the variance of $X(t)$,

$$\mathrm{Var}[\tilde{X}] = \mathrm{E}[\tilde{X}(t)\tilde{X}^*(t)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{g}(\omega_1)\hat{g}^*(\omega_2)\,\mathrm{E}[\hat{X}(\omega_1)\hat{X}^*(\omega_2)]\mathrm{e}^{\mathrm{i}(\omega_2-\omega_1)t}\,\mathrm{d}\omega_1\mathrm{d}\omega_2$$

Again using the fact that for weakly stationary processes we have $\mathrm{E}[\hat{X}(\omega_1)\hat{X}^*(\omega_2)] = \delta(\omega_1 - \omega_2)S(\omega_1)$ we see that $\tilde{X}(t)$ indeed has the same variance as $X(t)$.

Second we note that the convolution integral (13) gives $\tilde{X}(t)$ as a linear superposition of different time instances of $X(t)$ and therefore since $X(t)$ is a Gaussian process we have that $\tilde{X}(t)$ is Gaussian.

4.13.7 Distribution of the Hilbert envelope $\mathcal{E}(t)$

We have shown that if $X(t)$ is a weakly stationary Gaussian process then $\tilde{X}(t)$ is an identically distributed Gaussian process. We have also shown that these two processes are uncorrelated at the same time. It then follows that they are independent at the same time. In this case we know that the envelope $\mathcal{E}(t)$ is Rayleigh distributed and the total phase $\psi(t)$ is uniformly distributed.

In practical applications, nonlinearity will introduce dependencies between different frequencies/wavenumbers. Then the Central Limit Theorem can be broken, and the non-Gaussian behavior of $X(t)$ can imply deviation from the Rayleigh distribution for $\mathcal{E}(t)$.

4.13.8 Exercises

1. Show that we arrive at the same expression for the Hilbert transform also with the opposite sign convention for the exponents of the Fourier transform pair.

2. With $X(t) = \cos t$, show that the Hilbert transform is $\tilde{X}(t) = \sin t$ and that both the envelope and the instantaneous angular frequency are 1.

3. With $X(t) = \sin t$, show that the Hilbert transform is $\tilde{X} = -\cos t$ and that both the envelope and the instantaneous angular frequency are 1.

4. (trivial) Show that $|X(t)| \leq \mathcal{E}(t)$ and that $\mathcal{E}(t) \geq 0$.

5. For the narrowbanded model $\eta(x,t) = a\cos(k_c x - \omega_c t + \theta)$ where a and θ are slowly varying with respect to x and t, show that the Hilbert envelope is $\mathcal{E} = a$.

4.14 Estimation

We will be concerned about the estimation of various quantities related to stochastic variables or stochastic processes. Our main challenge is that we do not know the distributions of the stochastic variables or processes. Even if we may assume their distributions, we still have the challenge that we do not know the parameters of those distributions. Our main approach is to collect an ensemble of independent observations or realizations, and to average over the ensemble.

4.14.1 Estimation of mean

Suppose the stochastic variable X has mean $\mu = \mathrm{E}[X]$ and variance $\sigma^2 = \mathrm{Var}[X]$. We make n independent observations $X_1, X_2, \ldots, X_n$, and realize that each observation has the same distribution as the original variable X.

As an estimate of the mean we can use the *sample mean*

$$\bar{\mu} = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

This is an unbiased estimator for the mean

$$\mathrm{E}[\bar{\mu}] = \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n} X_j\right] = \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}[X_j] = \mu.$$

The variability of the estimator can be assessed by its variance

$$\begin{aligned}\mathrm{Var}[\bar{\mu}] &= \mathrm{E}\left[\left(\frac{1}{n}\sum_{j=1}^{n} X_j - \mu\right)^2\right] \\ &= \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{E}[(X_j-\mu)^2] + \frac{1}{n^2}\sum_{j\neq l}\mathrm{E}[X_j-\mu]\,\mathrm{E}[X_l-\mu] = \frac{\sigma^2}{n}\end{aligned}$$

where we have used the assumption that X_j and X_l are independent for $j \neq l$.

The standard deviation of the estimator $\bar{\mu}$ goes to zero asymptotically as the inverse of the square root of the number of observations for large ensembles. This result is obtained without any knowledge about the distribution of X other than assuming its (unknown) variance σ^2 is finite.

4.14.2 Estimation of variance

The variance of X can be estimated either by the *population variance* or by the *sample variance*.

The estimator for *population variance* is

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)^2$$

and requires that the mean μ is know. This is an unbiased estimator for the variance

$$\mathrm{E}[\widehat{\sigma^2}] = \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}[(X_j-\mu)^2] = \sigma^2.$$

The problem with the population variance as an estimator for the variance is that we usually do not know the mean μ.

The biased estimator for *sample variance* is

$$\sigma_x^2 = \frac{1}{n}\sum_{j=1}^{n}(X_j-\bar{\mu})^2$$

and employs the *sample mean* $\bar{\mu}$ rather than the true mean μ. As suggested by its name this is a biased estimator for the variance

$$\mathrm{E}[\sigma_x^2] = \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}[(X_j-\bar{\mu})^2] = \frac{n-1}{n}\sigma^2.$$

The unbiased estimator for *sample variance* can therefore be constructed as

$$s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j-\bar{\mu})^2.$$

In order to find the variability of the estimators for variance we need to assume a distribution for X. Assuming that X is Gaussian, we will need the Gamma function $\Gamma(z)$ and the χ_n^2 distribution (chi square distribution with n degrees of freedom).

The *Gamma function* is defined by

$$\Gamma(z) = \int_0^\infty x^{z-1}\mathrm{e}^{-x}\,\mathrm{d}x$$

It can be show that

- $\Gamma(z+1) = z\Gamma(z)$
- $\Gamma(1) = 1$
- $\Gamma(n) = (n-1)!$ for positive integers n
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Suppose that the stochastic variables X_j are independent and identically Gaussian distributed with mean 0 and variance 1, and let the stochastic variable $Y = \sum_{j=1}^{n} X_j^2$. Then Y is said to have χ_n^2 distribution, *chi squared distribution* with n degrees of freedom, with probability density function

$$f(y) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}y^{\frac{n}{2}-1}\mathrm{e}^{-\frac{y}{2}}H(y)$$

where $H(y)$ is the Heaviside step function.

In order to show that this is the case we can proceed in three steps, first show that the characteristic function for the χ^2_n distribution is $\phi_Y(k) = (1-2\mathrm{i}k)^{-\frac{n}{2}}$, second show that the characteristic function for X_j^2 is $\phi_{X_j^2}(k) = (1-2\mathrm{i}k)^{-\frac{1}{2}}$, and third recall that all the variables X_j are independent.

It can bee shown that the expected value of the χ^2_n distribution is $\mathrm{E}[Y] = n$ and the variance is $\mathrm{Var}[Y] = 2n$.

Now it is trivial to recognize that if we slightly rewrite the expression for the population variance, $\widehat{\sigma^2} = \frac{\sigma^2}{n}\sum_{j=1}^{n}\left(\frac{X_j-\mu}{\sigma}\right)^2$, the sum is recognized as having a χ^2_n distribution and we have the variance $\mathrm{Var}[\widehat{\sigma^2}] = \frac{2\sigma^4}{n}$. Thus the standard deviation of the population variance estimator for the variance is $\sqrt{\frac{2}{n}}\sigma^2$.

After some more work it can be shown that the variance of the unbiased sample variance estimator is $\mathrm{Var}[s^2] = \frac{2\sigma^4}{n-1}$.

4.14.3 Estimation of spectrum

Taking inspiration from equation (9) we may anticipate that the spectrum of a periodic process can be estimated by

$$\hat{S}_n = \frac{1}{\Delta\omega}|\hat{X}_n|^2. \tag{14}$$

From equation (9) we already know that this is an unbiased estimator

$$\mathrm{E}[\hat{S}_n] = S_n$$

In order to find the variability of this estimator for the spectrum, we need to assume a distribution for the underlying process. Assume

$$X(t) = A\cos(\omega_p t) + B\sin(\omega_p t)$$

for $0 \leq t \leq T$, with $\Delta\omega = 2\pi/T$ and $\omega_p = p\Delta\omega$ for some integer p, and where A and B are statistically independent equally distributed Gaussian variables with mean 0 and variance σ^2. We find the mean $\mathrm{E}[X(t)] = 0$ and the autocorrelation function

$$R(\tau) = \mathrm{E}[X(t+\tau)X^*(t)] = \sigma^2\cos(\omega_p\tau).$$

We find the spectrum to be

$$S_n = \frac{\sigma^2}{2\Delta\omega}(\delta_{n,p} + \delta_{n,-p}).$$

On the other hand we have the Fourier coefficients

$$\hat{X}_n = \frac{A - \mathrm{i}B}{2}\delta_{n,-p} + \frac{A + \mathrm{i}B}{2}\delta_{n,p}$$

and the spectrum estimator

$$\hat{S}_n = \frac{A^2 + B^2}{4\Delta\omega}(\delta_{n,-p} + \delta_{n,p})$$

which is the sum of the squares of two Gaussian variables. We therefore recognize that $\hat{S}_n$ has a χ^2_2 distribution which is the exponential distribution. From the properties of the χ^2_2 distribution (see above) we therefore conclude that

$$\mathrm{Var}[\hat{S}_n] = S_n^2$$

or alternatively that the estimator has standard deviation equal to its expected value. Sometimes we introduce the so-called coefficient of variation (C.O.V.) as the ratio of the standard deviation and the expected value, in the present case the coefficient of variation is unity.

4.14.4 Estimation with Matlab etc.

In Matlab and Octave we have the following routines which do not follow the same pattern for default behavior:

- `var(X)` = `var(X,0)` is the unbiased estimator for population variance, dividing by `N-1` if `N>1` where `N` is the sample size. `var(X,1)` divides by `N` and thus produces the second moment of the sample about its mean, which is the biased sample variance.
- `std(X)` = `std(X,0)` is the unbiased estimator for population standard deviation, dividing by `N-1` if `N>1` where `N` is the sample size. `std(X,1)` normalizes by `N` and produces the square root of the second moment of the sample about its mean, which is the biased sample standard deviation.
- `skewness(X)` = `skewness(X,1)` is the biased sample skewness. `skewness(X,0)` is the bias-corrected skewness returning an unbiased estimate of the population skewness.
- `kurtosis(X)` = `kurtosis(X,1)` is the biased sample kurtosis. `kurtosis(X,0)` is the bias-corrected kurtosis returning an unbiased estimate of the population kurtosis for normal populations.
- `pwelch(...)` estimates the power spectral density.

4.14.5 Exercises

1. Derive the properties of the Gamma function: $\Gamma(z+1) = z\Gamma(z)$, $\Gamma(1) = 1$, $\Gamma(n) = (n-1)!$ for positive integers n, and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

2. Let the stochastic variables X_j be independent and identically Gaussian distributed with mean 0 and variance 1, and let the stochastic variable $Y = \sum_{j=1}^{n} X_j^2$. Then Y is said to have χ^2_n distribution, *chi squared distribution* with n degrees of freedom.

   Carry out the following three steps in order to derive the χ^2_n distribution: First show that the characteristic function for the χ^2_n distribution is $\phi_Y(k) =$

$(1-2\mathrm{i}k)^{-\frac{n}{2}}$. Second show that the characteristic function for X_j^2 is $\phi_{X_j^2}(k) = (1-2\mathrm{i}k)^{-\frac{1}{2}}$. Third recall that all the variables X_j are independent.

Shown that the expected value of the χ_n^2 distribution is $\mathrm{E}[Y] = n$ and the variance is $\mathrm{Var}[Y] = 2n$.

3. Show that the χ_2^2 distribution is an exponential distribution. If Y is χ_n^2-distributed, then we say that $Z = \sqrt{Y}$ is χ_n-distributed, *chi distributed* with n degrees of freedom. Show that the χ_2 distribution is a Rayleigh distribution.

4. The real stochastic process $X(t) = a\cos(\omega_p t + \Theta)$ has been measured over a time interval T. Here Θ is uniformly distributed on $[0, 2\pi]$, $\omega_p = p\Delta\omega$ for some integer p, and $\Delta\omega = 2\pi/T$. We attempt to estimate the power spectral density $S(\omega)$ with the estimator $\hat{S}_n = \frac{1}{\Delta\omega}|\hat{X}_n|^2$ where $\hat{X}_n$ are the Fourier coefficients of $X(t)$. Compute the coefficient of variation (C.O.V.) of the spectrum estimate.

References

Goda, Y. 2000 *Random seas and design of maritime structures*. World Scientific, 443 pp.

Kinsman, B. 1984 *Wind waves*. Dover, 676 pp.

Lindgren, G. 2013 *Stationary stochastic processes*. CRC Press, 347 pp.

Naess, A. & Moan, T. 2013 *Stochastic dynamics of marine structures*. Cambridge University Press, 410 pp.

Ochi, M. K. 1998 *Ocean waves*. Cambridge University Press, 319 pp.

Papoulis, A. & Pillai, S. U. 2002 *Probability, random variables, and stochastic processes*, 4th edn. McGraw-Hill, 852 pp.

Tayfun, M. A. 1980 Narrow-band nonlinear sea waves. *J. Geophys. Res.* **85**, 1548–1552.

Tucker, M, J. & Pitt, E. G. 2001 *Waves in ocean engineering*. Elsevier.

Øksendal, B. 2003 *Stochastic differential equations*. Springer, 360 pp.