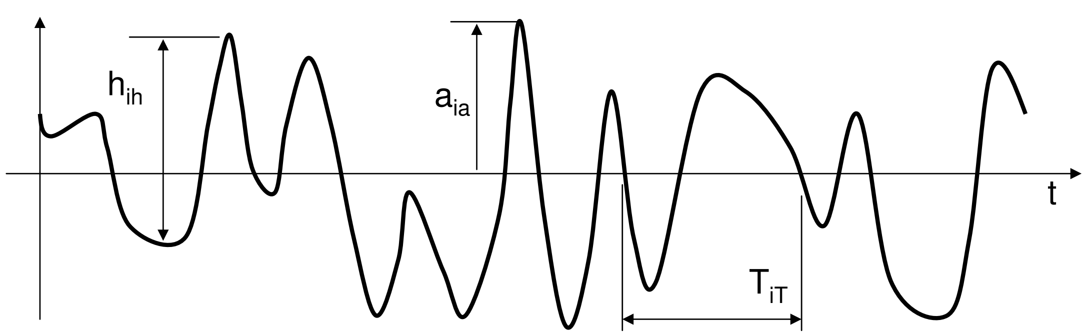
5 SHORT-TERM STATISTICS

The spectrum contains information about the magnitude of each frequency component in a stationary and ergodic random process. A summation of harmonic functions with random phase satisfies ergodicity and stationarity, and this will be a dominant model of a random process in our discussion. Also, the central limit theorem provides that a random process of sufficient length and ensemble size has a Gaussian distribution.

The primary calculation is the frequency-domain multiplication of an input spectrum by a transfer function magnitude squared, so as to obtain the output spectrum. Further, a Gaussian input driving an LTI system will create a Gaussian output. Most importantly, the input and output spectra have statistical information that we will be able to integrate into the system design process. In this section, we focus on short-term statistics, namely those which will apply to a random process that is truly stationary. An easy example is a field of ocean waves: over the course of minutes or hours, the process is stationary, but over days the effects of distant storms will change the statistics.

Considering specific "events" within a random process, several of the most important are the amplitude a_{i_a}, the height h_{i_h}, and the period T_{i_T}. The index here is counting through the record the number of amplitude measurements, height measurements, and period measurements. In the figure below, the period is measured specifically between zero downcrossings, and the amplitude is the maximum value reached after an upcrossing and before the next downcrossing. The height goes from the minimum after a zero downcrossing to the maximum after the following zero upcrossing. These definitions have to be applied consistently, because sometimes (as shown) there are fluctuations that do not cross over the zero line.



We will focus on statistics of the amplitude a; the spectrum used below is that of a. Let us define three even moments:

$$
\begin{aligned}
M_0 &= \int_0^\infty S^+(\omega) d\omega \\
M_2 &= \int_0^\infty \omega^2 S^+(\omega) d\omega \\
M_4 &= \int_0^\infty \omega^4 S^+(\omega) d\omega.
\end{aligned}
$$

We know already that M_0 is related to the variance of the process. Without proof, these are combined into a "bandwidth" parameter that we will use soon:

$$\epsilon^2 = 1 - \frac{M_2^2}{M_o M_4}.$$

Physically, ϵ is near one if there are many local minima and maxima between zero crossings (*broadband*), whereas it is near zero if there is only one maxima after a zero upcrossing before returning to zero (*narrow-band*).

5.1 Central Role of the Gaussian and Rayleigh Distributions

The Central Limit Theorem - which states that the sum of a large number of random variables approaches a Gaussian - ensures that stationary and ergodic processes create a data trace that has its samples normally distributed. For example, if a histogram of the samples from an ocean wave measurement system is plotted, it will indicate a normal distribution. Roughly speaking, in any given cycle, the trace will clearly spend more time near the extremes and less time crossing zero. But for the random process, these peaks are rarely repeated, while the zero is crossed nearly every time. It is recommended to try a numerical experiment to confirm the result of a normal distribution:

$$p(y) = \frac{1}{\sqrt{2\pi}\sigma_y} e^{-y^2/2\sigma_y^2},$$

where the standard deviation is σ_y and the mean is zero. As indicated above, the standard deviation is precisely the square root of the area under the one-sided spectrum.

In contrast with the continuous trace above, heights are computed only once for each cycle. Heights are defined to be positive only, so there is a lower limit of zero, but there is no upper limit. Just as the signal y itself can theoretically reach arbitrarily high values according to the normal distribution, so can heights. It can be shown that the distribution of heights from a Gaussian process is Rayleigh:

$$p(h) = \frac{h}{4\sigma_y^2} e^{-h^2/8\sigma_y^2},$$

where σ here is the standard deviation of the *underlying* normal process. The mean and standard deviation of the height itself are different:

$$\bar{h} = \sqrt{2\pi}\sigma_y \simeq 2.5\sigma_y$$
$$\sigma_h = \sqrt{8 - 2\pi}\sigma_y \simeq 1.3\sigma_y.$$

Notice that the Rayleigh pdf has an exponential of the argument squared, but that this exponential is also multiplied by the argument; this drives the pdf to zero at the origin. The cumulative distribution is the simpler Rayleigh cpf:

$$p(h < h_o) = 1 - e^{-h_o^2/8\sigma_y^2};$$

$P(h)$ looks like half of a Gaussian pdf, turned upside down! A very useful formula that derives from this simple form is that

$$p(h > h_o) = 1 - p(h < h_o) = e^{-2h_o^2/(\bar{h}^{1/3})^2}.$$

This follows immediately from the cumulative probability function, since $\bar{h}^{1/3} = 4\sigma_y$. It is confirmed that $p(h > \bar{h}^{1/3}) = e^{-2} \simeq 0.13$.

5.2 Frequency of Upcrossings

The first statistic we discuss is the frequency with which the process exceeds a given level; we consider upcrossings only of the positive value A. Now let $\bar{f}(A)$ be the average frequency of upcrossings past A, in upcrossings per second. Then $\bar{f}(0)$ is the average frequency of zero upcrossing, or $1/\bar{T}$, the inverse of the average period, $E(T)$. The formulas are

$$\begin{aligned} \bar{f}(0) &= \frac{1}{2\pi}\sqrt{\frac{M_2}{M_0}} \\ \bar{f}(A) &= \frac{1}{2\pi}\sqrt{\frac{M_2}{M_0}} e^{-A^2/2M_0}. \end{aligned}$$

With M_0 equal to the variance, the exponential here clearly is of the Gaussian form. Here is an example of the use of these equations in design. An fixed ocean platform is exposed to storm waves of standard deviation two meters and average period eight seconds. How high must the deck be to only be flooded every ten minutes, on average?

This problem does not involve any transfer function since the platform is fixed. If it were floating, there would be some motion and we would have to transform the wave spectrum into the motion spectrum. All we have to do here is invert the equation to solve for A, given that $\bar{f}(A) = 1/(60 \times 10)$, $M_0 = 4$ and $\bar{T} = 8$ or $\bar{f}(0) = 1/8$:

$$A = \sqrt{-2M_0 \ln\left(\bar{T}\bar{f}(A)\right)} = 5.87m.$$

This result gives a flavor of how valuable these statistics will be - even though the standard deviation of wave elevation is only two meters, every ten minutes we should expect a six-meter amplitude!

5.3 Maxima At and Above a Given Level

Now we look at the probability of any maximum amplitude a_{ia} reaching or exceeding a given level. We normalize the amplitude with the random process variance, i.e., $\eta = a/\sqrt{M_0}$ and

$\bar{\eta} = A/\sqrt{M_0}$. The results are very useful for calculating extreme loads. First,

$$\begin{aligned} p(\eta = \bar{\eta}) &= \frac{\epsilon}{\sqrt{2\pi}} e^{-\bar{\eta}^2/2\epsilon^2} + \phi(\bar{\eta}q/\epsilon)\frac{\bar{\eta}q}{\sqrt{2\pi}} e^{-\bar{\eta}^2/2} \text{ where} \\ q &= \sqrt{1-\epsilon^2}, \\ \phi(\xi) &= \int_{-\infty}^{\xi} e^{-u^2/2} du \text{ (related to the error function erf).} \end{aligned}$$

With large amplitudes being considered and small ϵ (a narrow-banded process), we can make some approximations to find:

$$\begin{aligned} p(\eta = \bar{\eta}) &\approx \frac{2q}{1+q} \bar{\eta} e^{-\bar{\eta}^2/2} \longrightarrow \\ p(\eta > \bar{\eta}) &\approx \frac{2q}{1+q} e^{-\bar{\eta}^2/2}. \end{aligned}$$

The second relation here is the more useful, as it gives the probability that the (nondimensional) amplitude will *exceed* a given value. It follows directly from the former equation, since (roughly) the cumulative distribution is the derivative of the probability density.

5.4 1/N'th Highest Maxima

We next define a statistic that is the average lowest value of the 1/N'th highest peaks. The construction goes this way: From the record, list all of the highest peaks. Rank and then collect the highest 1/N fraction of these numbers. The average lowest value of this set is the 1/N'th highest maxima. For instance, let the peaks be [6 7 11 5 3 4 8 5 9 4 2 5]. There are twelve numbers here, and the one-third highest maxima $a^{1/3}$ is around 6.5, because [7 11 8 9] exceed it. We use the superscript $1/N$ to denote the 1/N'th highest maxima of a quantity.

Building on the previous section, we have that

$$\begin{aligned} p(a > a^{1/N}) = \frac{1}{N} &\approx \frac{2q}{1+q} \exp\left(-(a^{1/N})^2/2M_0\right), \text{ so that} \\ a^{1/N} &\approx \sqrt{2M_0 \ln\left(\frac{2q}{1+q}N\right)}. \end{aligned}$$

5.5 1/N'th Average Value

We can similarly define a statistic that is the average of the 1/N'th highest peaks. In this case, we are after the average of this collection of 1/N peaks:

$$\begin{aligned} \bar{a}^{1/N} &= E(a | a > a^{1/N}) \\ &= \int_{a^{1/N}}^{\infty} a \, p(a = a_m | a_m > a^{1/N}) da. \end{aligned}$$

Note that we use the dummy variable a_m. We have then the conditional probability

$$p(a = a_m | a_m > a^{1/N}) = \frac{p[(a = a_m) \cap (a_m > a^{1/N})]}{p(a_m > a^{1/N})}.$$

Working in nondimensional form, we have

$$\begin{aligned} \bar{\eta}^{1/N} &= \int_{\eta^{1/N}}^{\infty} \frac{1}{1/N} \eta p(\eta = \eta_m) d\eta \\ &= \frac{2qN}{1+q} \int_{\eta^{1/N}}^{\infty} \eta^2 e^{\eta^2/2} d\eta. \end{aligned}$$

Here are a few explicit results for amplitude and height:

$$\begin{aligned} \bar{a}^{1/3} &= 1.1\sqrt{M_0} \text{ to } 2\sqrt{M_0} \\ \bar{a}^{1/10} &= 1.8\sqrt{M_0} \text{ to } 2.5\sqrt{M_0}. \end{aligned}$$

The amplitudes here vary depending on the parameter ϵ - this point is discussed in the Principles of Naval Architecture, page 20. Here are some 1/N'th average heights:

$$\begin{aligned} \bar{h}^{1/3} &= 4.0\sqrt{M_0} \\ \bar{h}^{1/10} &= 5.1\sqrt{M_0}. \end{aligned}$$

The value $\bar{h}^{1/3}$ is the significant wave height, the most common description of the size of waves. It turns out to be very close to the wave size reported by experienced mariners.

Finally, here are expected highest heights in N observations - which is not quite either of the 1/N'th maximum or the 1/N'th average statistics given above:

$$\begin{aligned} \bar{h}100) &= 6.5\sqrt{M_0} \\ \bar{h}(1000) &= 7.7\sqrt{M_0} \\ \bar{h}10000) &= 8.9\sqrt{M_0}. \end{aligned}$$

5.6 The 100-Year Wave: Estimate from Short-Term Statistics

For long-term installations, it is important to characterize the largest wave to be expected in an extremely large number of cycles. We will make such a calculation here, although as indicated in our discussion of Water Waves, the Rayleigh distribution does not adequately capture extreme events over such time scales. Spectra and the consequent Rayleigh height distribution are *short-term properties only.*

The idea here is to equate $p(h > h_o)$ from the distribution with the definition that in fact $h > h_o$ once in 100 years. Namely, we have

$$p(h > h_{100yr}) = \frac{1}{100\text{years}/\bar{T}} = \mathrm{e}^{-2h_{100yr}^2/\bar{h}^{1/3}},$$

where $\bar{T}$ is the average period. As we will see, uncertainty about what is the proper $\bar{T}$ has little effect in the answer. Looking at the first equality, and setting $\bar{T} = 8$ seconds and $\bar{h}^{1/3} = 2$ meters as example values, leads to

$$\begin{aligned} 2.5 \times 10^{-9} &= \mathrm{e}^{-2h_{100yr}^2/\bar{h}^{1/3}}; \\ \log(2.5 \times 10^{-9}) &= -2h_{100yr}^2/4; \\ h_{100yr} &= 6.3 \text{ meters, or } 3.1\bar{h}^{1/3}. \end{aligned}$$

According to this calculation, the 100-year wave height is approximately three times the significant wave height. Because $\bar{T}$ appears inside the logarithm, an error of twofold in $\bar{T}$ changes the extreme height estimate only by a few percent.

2.017J Design of Electromechanical Robotic Systems
Fall 2009