EXTREME VALUE ANALYSIS, NONLINEAR RANDOM OSCILLATORS, AND APPLICATIONS TO SHIP MOTIONS IN IRREGULAR SEAS

Dylan Glotzer

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2018

Approved by:

Vladas Pipiras

Amarjit Budhiraja

Jan Hannig

Ross Leadbetter

Kai Zhang

ABSTRACT

DYLAN GLOTZER: Extreme value analysis, nonlinear random oscillators, and applications to ship motions in irregular seas
(Under the direction of Vladas Pipiras)

This dissertation concerns several statistical problems arising in Naval Engineering which relate to statistical uncertainty and characterizing rare events, and naturally involve a stochastic component to be accounted for through statistical methods. Chapter 1 has been adapted from Glotzer and Pipiras [43] and provides an overview of the problems to be discussed in later chapters.

In Chapter 2, statistical inference of a probability of exceeding a large critical value is studied in the peaks-over-threshold (POT) approach. The focus is on assessing the performance of various confidence intervals for the exceedance probability, and several approaches to uncertainty reduction are considered. This chapter has been published as Glotzer et al. [44].

Chapter 3 concerns the study of a single-degree-of-freedom random oscillator with a piecewise linear restoring force (experiencing softening after a certain point value of the response, called a "knuckle" point), with the goal of understanding the structure of the distribution tail of its response or (local) maximum. The random oscillator considered here serves as a prototypical model for ship roll motion in beam seas. A theoretical analysis is carried out first by focusing on the maximum and response after crossing the knuckle point, where explicit calculations can be performed assuming standard distributions for the derivative at the crossing; and second by considering the white noise random external excitation, where the stationary distribution of the response is readily available from the literature. Both approaches reveal the structure of the distribution tails where a Gaussian core is followed by a heavier tail, possibly having a power-law form, ultimately resulting in a tail having a finite upper bound. The extent of the light tail region is shown to be the result of conditioning on the system not reaching the unstable equilibrium (associated with "capsizing").

In Chapter 4, the distributions of certain response rates of the above random oscillator are investigated analytically and numerically. These include the minimum response rate leading to capsizing, referred to as the critical response rate, and the split-time metric, which measures the closeness of an observed response rate to the critical response rate and is used to assess the probability of capsizing. Three nonlinear restoring forces are considered: piecewise linear (experiencing linearly softening stiffness above a knuckle point), doubly piecewise linear (experiencing piecewise linearly softening stiffness above a knuckle point), and the cubic restoring force of the Duffing oscillator. Numerical simulation of the critical response rate and split-time metric is proposed from a derived distribution in the first two cases; in the latter case, the density of the critical response rate is approximated assuming white noise excitation. The distribution of the split-time metric is found to have a "light" tail, motivating the use of an exponential distribution or a Weibull distribution tail rather than the generalized Pareto distribution for exceedances above threshold in the POT approach.

Finally, Chapter 5 considers inference for the mean and variance of a stationary continuous-time stochastic process, for which the so-called long-run variances of the process and its square play a central role. The well-known problem of estimating the non-zero long-run variances is revisited here in the context of random oscillatory processes such as random (non-)linear oscillators and related models. The less studied case of the zero long-run variance is also considered. The approaches are extended to the situation where multiple independent records of the stochastic process are available, by introducing an estimator of the long-run variance which improves on other natural candidate estimators. A simulation study is provided to assess the performance of the proposed methods in estimating the long-run variances and constructing confidence intervals.

To my wife and life, Lindsey.

ACKNOWLEDGEMENTS

I would first and foremost like to thank my advisor, Professor Vladas Pipiras, for his incredible support and mentorship. Neither this dissertation nor the work contained therein would have been possible without Prof. Pipiras' guidance and counsel. His thoughtfulness in contemplating questions along with his keen attention to detail have shown me how to engage with problems and to enjoy thinking about them. I will forever be grateful for such a lesson.

I would also like to extend a sincere thank you to each of the other members of my dissertation committee, all of whom have contributed to my education in significant ways: Professors Amarjit Budhiraja, Jan Hannig, Ross Leadbetter, and Kai Zhang. For each of them, I have had the pleasure of receiving instruction in the classroom, along with vital advice and direction outside of the classroom. They each have given me a perspective on, and appreciation for, the value of deep scholarly thought.

Much of the work presented in this dissertation was initiated while visiting the Carderock Division of the Naval Surface Warfare Center (NSWCCD), as a summer intern under the Naval Research Enterprise Internship Program under mentor Timothy Smith, who made significant contributions to the work presented in Chapter 2. Additionally, I would like to thank Dr. Vadim Belenky for his time spent working closely with me at NSWCCD, for his consistent good spirits, and for his essential contributions to the work presented in Chapters 2 - 5. Additional contributors at NSWCCD include Bradley Campbell (Chapter 2), and Kenneth Weems (Chapters 4, 5). Professor Themis Sapsis, from the Massachusetts Institute of Technology, also contributed to the work presented in Chapters 3 and 4.

Thank you to my fellow UNC graduate students. I am forever indebted to the culture of geniality and industriousness found in the Hanes Hall basement, which has allowed me (and others) to thrive.

Finally, thank you to my parents, Jane and Steve, and my brother, Harlan, for making me believe I could accomplish anything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CDF	Cumulative Distribution Function
DPWL	Doubly Piecewise Linear
EVT	Extreme Value Theory
FPK	Fokker-Planck-Kolmogorov
GPD	Generalized Pareto Distribution
LAMP	Large Amplitude Motion Program
pdf	Probability Density Function
POT	Peaks-Over-Threshold
PWL	Piecewise Linear

CHAPTER 1

Statistical perspectives on some problems arising in Naval Engineering

1.1 Introduction

Statistical methods play important roles for many problems arising in Naval Engineering. They are naturally called for when analyzing collected data, when incorporating random components into physical models, or when understanding potentially intricate relationships between variables of interest. The goal of this introductory chapter is to discuss several such problems of relevance in Naval Engineering, which will be studied in greater detail in further chapters.

1.2 Extreme values

1.2.1 Probabilities of rare events

A general class of problems arising in Naval Engineering concerns extreme values (rare events) and estimation of their probabilities. Two examples of rare events of interest are the following:

Case 1. A ship motion (e.g. roll) exceeding a certain large angle (in either direction);

Case 2. Capsizing of a ship.

The goal is, generally speaking, to estimate the probabilities of such rare events given a time history of a ship motion and perhaps additional information. As outlined in Chapter 2, this can be done under the umbrella of the statistical methodology known as Extreme Value Theory (EVT; e.g. Coles [23]). Additionally, qualitative stochastic dynamics models for ship motions justify or refine the use of EVT, and interact with ship geometry and other relevant parameters. This will be discussed in Chapters 3 and 4.

The first observation to make is that many of the problems on estimation of rare probabilities could be recast into an estimation problem for the exceedance probability

$$\mathbb{P}(Y > y^*), \tag{1.1}$$

where Y is some variable related to the extreme event of interest and y^* is some large critical value. Moreover, this should be done from observations $Y_1, \ldots, Y_n$ of Y that can be assumed to be independent. No values of $Y_1, \ldots, Y_n$ are typically larger than y^*. In Case 1 above, Y can be taken as a suitable envelope peak of a ship motion, thus "declustering" possible time dependent peak values of the motion – see Chapter 2 for details. In Case 2 above, Y is taken as the so-called split-time metric (see Belenky et al. [14]), defined as

$$Y = \dot{X} - \dot{X}_{cr} + 1, \tag{1.2}$$

where $\dot{X}$ is the crossing rate at some fixed intermediate threshold and $\dot{X}_{cr}$ is the critical rate, that is, the smallest rate that would lead to capsizing. The split-time metric will be discussed in more detail in Chapter 4. These quantities are computed through a ship motion simulator (see Belenky et al. [14, 16] for more details). This corresponds to the exceedance probability (1.1) with $y^* = 1$, since capsizing is associated with the event that $\dot{X} > \dot{X}_{cr}$. Estimating the probability in (1.1) thus concerns extrapolating into the tail where no data are available.

The extrapolation problem of estimating the probability in (1.1) is commonly tackled with the peaks-over-threshold (POT) approach of EVT. In this approach, the probability in (1.1) is expressed as

$$\begin{aligned} \mathbb{P}\left(Y > y^*\right) &= \mathbb{P}\left(Y > u\right) \mathbb{P}\left(Y > y^* | Y > u\right) \\ &=: \mathbb{P}\left(Y > u\right) \mathbb{P}_u\left(Y > y^*\right), \end{aligned} \tag{1.3}$$

where u ($u < y^*$) is an intermediate threshold with sufficiently many Y_i's above u and $\mathbb{P}_u$ denotes a conditional probability. The probability $\mathbb{P}\left(Y > u\right)$ is estimated as the proportion of data $Y_i > u$, with the associated confidence interval set by binomial calculations. The conditional probability $\mathbb{P}_u\left(Y > y^*\right)$, on the other hand, is approximated by the tail of the

generalized Pareto distribution (GPD), namely, as

$$\mathbb{P}_u\left(Y > y^*\right) \approx \overline{F}_{u,\xi,\sigma}(y^*), \tag{1.4}$$

where

$$\overline{F}_{\mu,\xi,\sigma}(x) := \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)_+^{-1/\xi} := \begin{cases} \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}, & \mu < x, \quad \text{if } \xi > 0, \\ e^{-\frac{x-\mu}{\sigma}}, & \mu < x, \quad \text{if } \xi = 0, \\ \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}, & \mu < x < \mu - \frac{\sigma}{\xi}, \quad \text{if } \xi < 0, \end{cases} \tag{1.5}$$

with a threshold u, a scale parameter $\sigma > 0$ and a shape parameter $\xi \in \mathbb{R}$. This approximation is justified by the so-called second extreme value theorem (Pickands-Balkema-de Haan theorem; Pickands [81]), which essentially states that the difference between the two probabilities in (1.4) approaches 0 uniformly for $y^* > u$ as $u \to \infty$. The right-hand side of (1.4) is used for extrapolation into the tail.

Considerable literature in Statistics is devoted to estimating the parameters of the GPD (whose estimation uncertainty is then propagated onto estimation of the tail probability in (1.4)) as well as on setting a threshold. A typical method to estimate the GPD parameters ξ, σ (for fixed u) is through maximum likelihood, and a variety of confidence intervals for the exceedance probability are available (see Chapter 2 for details and further references). Among threshold selection methods, visual tools are still commonly used. For example, Figure 1.1 depicts the shape parameter estimates $\widehat{\xi}$ against the varying threshold u for a Pareto random variable Y satisfying $\mathbb{P}\left(Y > y\right) = y^{-3}$, $y > 1$, and hence (1.4), with $\xi = 1/3$; the normal confidence intervals for ξ are also plotted. A threshold u in the approximation in (1.16) can be selected as a threshold above which the estimates stabilize. The choice of a threshold is critical and is the subject of ongoing research; see Scarrott and MacDonald [88] for a review. Details are also provided in Chapter 2.

Several further observations need to be made regarding the approximation in (1.4). Which random variables Y (or the corresponding probability distributions) are associated with which shape parameter values ξ is well known in theory (e.g. de Haan and Ferreira [27]). More specifically, the correspondence is:

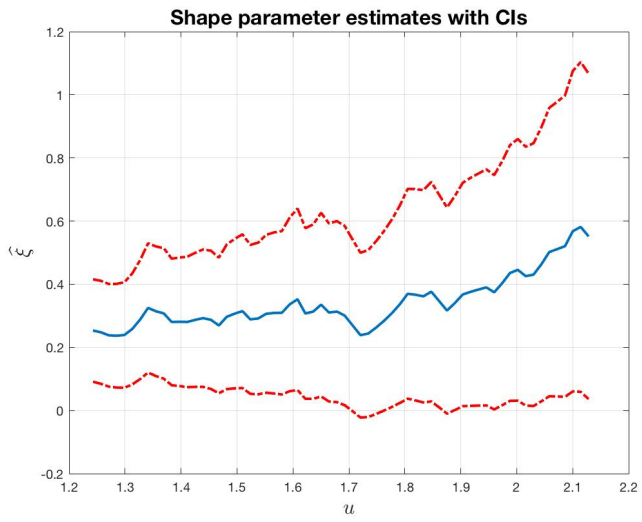

Figure 1.1: Shape parameter estimates and confidence intervals for a Pareto distribution as a function of threshold.

- $\xi > 0$: distributions having "heavy" (power-law) tails; essentially, $\mathbb{P}(Y > y) \sim cy^{-1/\xi}$, as $y \to \infty$;
- $\xi < 0$: distributions having "light" tails with finite upper bound y_0; essentially, such that $\mathbb{P}(Y > y) \sim c(y_0 - y)^{-1/\xi}$, as $y \to y_0$;
- $\xi = 0$: most distributions having "light" tails with infinite upper bound, such as exponential, Gaussian, Weibull and others.

From a practical perspective, the GPD provides a convenient parametric family that allows modeling tails raging from light with finite upper bound to heavy. This could also be seen by expressing the GPD tail as

$$\overline{F}_{u,\xi,\sigma}(y) = e^{-h_{u,\xi,\sigma}(y)}, \tag{1.6}$$

where $h_{u,\xi,\sigma}(y)$ depicted in Figure 1.2 allows for deviations from the straight line and hence the exponential distribution ($\xi = 0$) towards heavier ($\xi > 0$) and lighter ($\xi < 0$) tails. Along similar lines, for example, a negative value of ξ would be suggested for the lighter-than-exponential Gaussian distribution (see Figure 1.3), even if in theory this distribution is associated with the shape parameter $\xi = 0$.

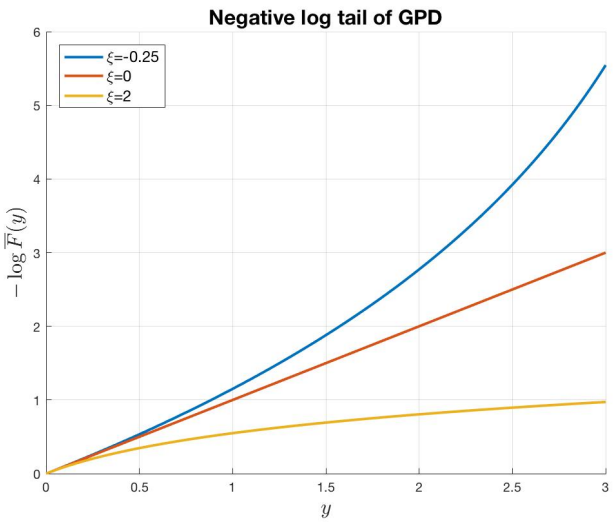

Figure 1.2: The function $h_{u,\xi,\sigma}(y)$ in (1.6) with $u = 0, \sigma = 1$ and several choices of ξ.

In practice, assessing the performance of the above POT methodology involves repeated runs (replications), estimating the rare probability and constructing the associated confidence interval for each run, as depicted in Figure 1.4. The true probability is added to the plot as a horizontal line, after computing it analytically if possible or calculating it numerically from a much larger data set. The methodology performs well if the confidence intervals include the true probability for the proportion of times corresponding to the chosen confidence level (e.g. 95%).

One feature of these confidence intervals is their size, especially the average size above the true probability. A related question of interest is whether the size of confidence intervals and hence statistical uncertainty can be reduced. In some cases, this can be achieved by taking the physical properties of the system into account. For example, in Chapter 2, the statistical uncertainty will be reduced by fixing an upper bound in the situations when negative shape parameter values were dominant.

1.2.2 Stochastic dynamics perspective

Despite some justification of the use of the POT approach discussed above and around (1.4) in particular, the methodology might appear to be a black box, without any physical insight. The latter will be provided in Chapters 3 and 4, with the methodology justified further

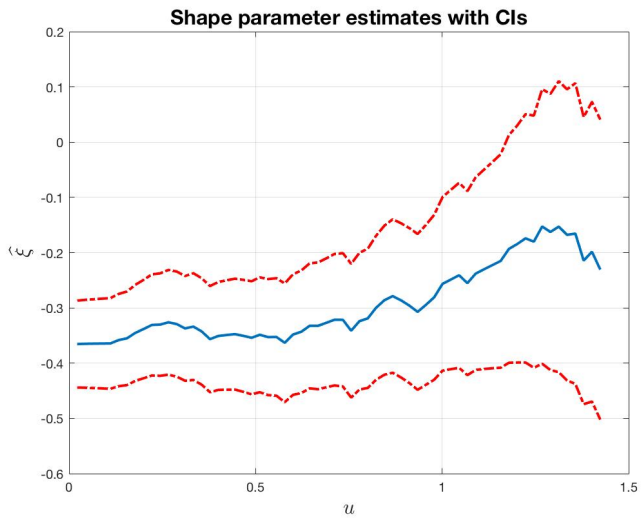

Figure 1.3: Shape parameter estimates and confidence intervals against threshold for a Gaussian distribution.

by examining its performance on and interactions with physical models for ship motions. One such prototypical model for ship roll motion in beam seas is a random oscillator with a nonlinear restoring force and random excitation, satisfying

$$\ddot{x}(t) + 2\delta\dot{x}(t) + r(x(t)) = z(t), \tag{1.7}$$

where $\delta > 0$ is the damping parameter, $r(x)$ is the restoring force and $z(t)$ is the zero-mean random excitation. The excitation $z(t)$ is commonly assumed to be a Gaussian stationary process, with the spectral density suggested by e.g. the Bretschneider spectrum for wave elevations. White noise excitation $z(t)$ has also been assumed for analytic tractability and because inertia plays a dominant role in ship motion, so high frequency content in the excitation is "filtered out" naturally by the system.

A further simplification can be made by assuming a piecewise linear restoring force $r(x)$, given by

$$r(x) = \begin{cases} -k_1\omega_0^2\left(x + x_m\right) - \omega_0^2 x_m, & \text{if } x < -x_m, \\ \omega_0^2 x, & \text{if } -x_m \leq x \leq x_m, \\ -k_1\omega_0^2\left(x - x_m\right) + \omega_0^2 x_m, & \text{if } x > x_m, \end{cases} \tag{1.8}$$

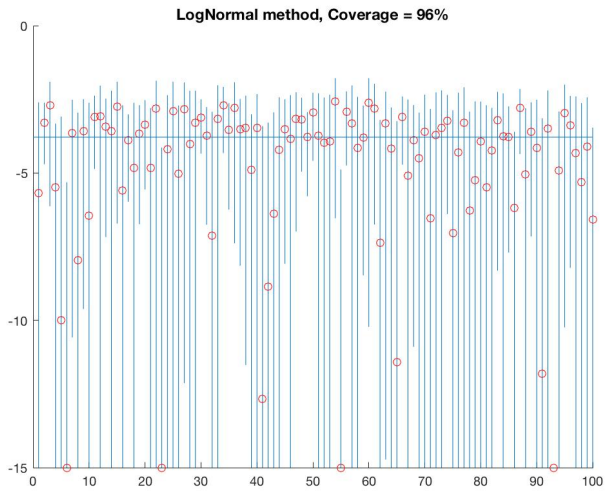

Figure 1.4: Probability estimates (on a log vertical scale) with confidence intervals for 100 runs.

where ω_0 is the natural frequency of the system, x_m is referred to as the knuckle point (separating the linear and nonlinear regimes) and $k > 0$ enters into the negative slope of the nonlinear part. The point $x_v > 0$ such that $r(x_v) = 0$ is called the point of vanishing stability.

A yet simpler case is obtained by assuming that the excitation is switched off above the knuckle point. In this case, the solution of (1.7) above the knuckle point can be expressed as

$$x(t) = Ae^{\lambda_1 t} + Be^{\lambda_2 t} + x_v, \tag{1.9}$$

where x_v is the point of vanishing stability, $\lambda_1 = -\delta + \sqrt{k\omega_0^2 + \delta^2} > 0$, $\lambda_2 = -\delta - \sqrt{k\omega_0^2 + \delta^2} < 0$, and the constants

$$A = \frac{\dot{x}_1 + \lambda_2(x_v - x_m)}{\lambda_1 - \lambda_2}, \quad B = \frac{-\lambda_1(x_v - x_m) + \dot{x}_1}{\lambda_1 - \lambda_2}, \tag{1.10}$$

depend on the rate $\dot{x}_1$ at the upcrossing of the knuckle point x_m. Capsizing is associated with $A > 0$ and hence corresponds to $\dot{x}_1 > \lambda_2(x_v - x_m)$. Otherwise, the path of $x(t)$ has a peak x_{max} with its value in the interval $[x_m, x_v]$ given by a nonlinear transformation

$$G(\dot{x}_1) = x_v - C|B|^{\frac{\lambda_1}{\lambda_1 - \lambda_2}}|A|^{\frac{-\lambda_2}{\lambda_1 - \lambda_2}} \tag{1.11}$$

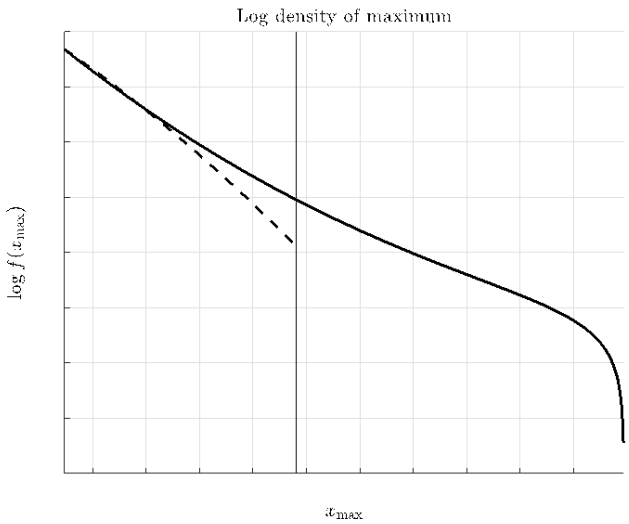

Figure 1.5: The density of the maximum value on the log vertical scale, with that for the linear system in dashed.

for some constant C. This transformation will tend to have a "stretching" effect for the corresponding probability distribution, compared to the linear case, which should not be surprising given that the restoring force experiences softening after the knuckle point and hence the paths of $x(t)$ will take larger values for the same energy after the upcrossing. Figure 1.5 illustrates this effect for parameter values motivated by a ship roll application, where the distribution of $\dot{x}_1$ is taken as Rayleigh in analogy to the corresponding distribution for a linear system, except truncated from above to ensure that no capsizing occurs (that is, $\dot{x}_1 < \lambda_2 (x_v - x_m)$ from the discussion above).

In fact, the stretching effect can be so pronounced that a positive shape parameter of the GPD would be estimated for the peaks data generated by the above system. Vice versa, the distribution of the maximum collapses around the end point x_v, so a negative shape parameter of the GPD can be estimated as well, depending on the choice of the threshold and the availability of the data in the collapsing tail. Since both positive and negative shape parameter estimates are expected, the GPD seems a rational choice for a parametric distribution when working with peaks of ship motion data.

The split-time metric, on the other hand, has properties that are quite different from those of the extreme motion (peaks). This will be discussed in Chapter 4. By the discussion following (1.10), the critical roll rate for the above system is constant and given by $\lambda_2(x_v - x_m)$. The

split-time metric is then

$$Y = \dot{x}_1 - \lambda_2(x_v - x_m) + 1 \tag{1.12}$$

and hence is a linear function of the rate $\dot{x}_1$ of the upcrossing. As noted above, the distribution of $\dot{x}_1$ is taken as Rayleigh. From the perspective of the GPD, the Rayleigh distribution is in the domain of attraction of the GPD with the shape parameter $\xi = 0$, though in practice negative shape parameters would be estimated since the Rayleigh tail is lighter than the exponential tail.

The analogous conclusion about the GPD distribution with the shape parameter $\xi = 0$ can be obtained for a number of systems extending the simplest case described above, including the piecewise linear system with the excitation switched on above the knuckle point and quite general nonlinear oscillators (e.g. the Duffing oscillator) driven by white noise excitation. Motivated by these findings, the POT methodology has been applied to extrapolate the distribution of the split-time metric by using the exponential distribution or the tail of the Weibull distribution rather than the GPD to fit exceedances over a large intermediate threshold.

In both Cases 1 and 2 (see the beginning of Section 1.2.1), the distribution tail of the variable of interest is related directly to the form of the restoring force, which itself is determined by ship geometry.

1.3 Statistical uncertainty in parameter estimation

A common problem arising with random processes observed in Naval Engineering (such as motions, loads, etc.) is estimation of quantities (parameters) of interest and setting related confidence intervals that characterize statistical uncertainty in estimation. Parameters of interest could be the mean, the standard deviation, the autocovariance and so on, as well as quantities derived from these (such as single significant amplitude and others). A quite general statistical methodology to deal with this problem is outlined next and discussed in further detail in Chapter 5. The basic difficulty in characterizing statistical uncertainty is to account for time dependency in observed random processes that is naturally present due to e.g. ship motion inertia and irregular seas.

The first observation to make is that many parameters of interest could be expressed in terms of means (expected values) of random processes. Here are a few examples. Let $\{X_t\}_{t\in\mathbb{R}}$ be a stationary process of interest. Suppose that its standard deviation $\sigma_X = \sqrt{\mathrm{Var}(X_t)}$ needs to be estimated. Then,

$$\sigma_X = \sqrt{\mathbb{E}(X_t^2) - (\mathbb{E}X_t)^2}, \tag{1.13}$$

that is, it involves the means $\mathbb{E}(X_t^2)$ and $\mathbb{E}X_t$ of the processes X_t^2 and X_t, respectively. Similarly, suppose the autocovariance function $\gamma_X(h) = \mathrm{Cov}(X_t, X_{t+h})$ at lag $h > 0$ is of interest. Then,

$$\gamma_X(h) = \mathbb{E}(X_t X_{t+h}) - (\mathbb{E}X_t)^2, \tag{1.14}$$

that is, it involves the means $\mathbb{E}(X_t X_{t+h})$ and $\mathbb{E}X_t$ of the processes $X_t X_{t+h}$ and X_t, respectively. More generally, many quantities of interest can be expressed as

$$Q = g(\mathbb{E}X_{1,t}, \ldots, \mathbb{E}X_{k,t}) =: g\left(\mathbb{E}X_t^{(k)}\right), \tag{1.15}$$

where $g(x_1, \ldots, x_k)$ is a known "smooth" function and $X_t^{(k)} = (X_{1,t}, \ldots, X_{k,t})$ is a k-variate stationary process, expressed in terms of the observed process X_t. For example, $Q = \gamma_X(h)$ in (1.14) can be written as (1.15) with $g(x_1, x_2) = x_1 - x_2^2$ and $X_{1,t} = X_t X_{t+h}$, $X_{2,t} = X_t$.

A quantity Q expressed by (1.15), is estimated through

$$\widehat{Q} = g\left(\overline{X}_{1,T}, \ldots, \overline{X}_{k,T}\right) =: g\left(\overline{X}_T^{(k)}\right), \tag{1.16}$$

where $\overline{X}_{1,T}, \ldots, \overline{X}_{k,T}$ are the sample means of the processes $X_{1,t}, \ldots, X_{k,t}$, computed from observing the original process X_t on the time interval $[0, T]$. It is expected that the sample means are asymptotically normal, that is,

$$\begin{aligned}\sqrt{T}\left(\overline{X}_T^{(k)} - \mathbb{E}X_t^{(k)}\right) &= \sqrt{T}\left(\overline{X}_{1,T} - \mathbb{E}X_{1,t}, \ldots, \overline{X}_{k,T} - \mathbb{E}X_{k,t}\right) \\ &\xrightarrow{d} N(0, \Pi^{(k)}),\end{aligned} \tag{1.17}$$

where $N(0, \Pi^{(k)})$ indicates a k-variate Gaussian (normal) distribution with a zero-mean vector and a $k \times k$ covariance matrix $\Pi^{(k)} = (\pi_{ij})_{i,j=1,\ldots,k}$, and $\xrightarrow{d}$ indicates the convergence in dis-

tribution. The covariance matrix $\Pi^{(k)}$ will be discussed again below. But supposing that the relation (1.17) holds, the δ-method then yields

$$\sqrt{T}(\widehat{Q} - Q) \xrightarrow{d} N(0, \sigma_Q^2), \tag{1.18}$$

where

$$\sigma_Q^2 = \nabla g(\mathbb{E}X_t^{(k)})' \Pi^{(k)} \nabla g(\mathbb{E}X_t^{(k)}) \tag{1.19}$$

with $\nabla g\,(x_1, \ldots, x_k) = (\partial g/\partial x_1, \ldots, \partial g/\partial x_k)'$ and the prime indicating a transpose. A normal approximate confidence interval for Q is then

$$\left(\widehat{Q} - \frac{z_\alpha \widehat{\sigma}_Q}{\sqrt{T}}, \widehat{Q} + \frac{z_\alpha \widehat{\sigma}_Q}{\sqrt{T}}\right), \tag{1.20}$$

where $\widehat{\sigma}_Q$ is an estimator for σ_Q and z_α is a critical value for the desired confidence level α (e.g. $z_\alpha = 1.96$ for $\alpha = 95\%$). In view of (1.19), set

$$\widehat{\sigma}_Q^2 = \nabla g(\overline{X}_T^{(k)})' \widehat{\Pi}^{(k)} \nabla g(\overline{X}_T^{(k)}), \tag{1.21}$$

where $\widehat{\Pi}^{(k)}$ is an estimator for $\Pi^{(k)}$. For example, for $Q = \gamma_X(h)$ in (1.14) that can be written as (1.15) with $g\,(x_1, x_2) = x_1 - x_2^2$ and $X_{1,t} = X_t X_{t+h}$, $X_{2,t} = X_t$, the quantity ∇g appearing in (1.19) is $\nabla g\,(x_1, x_2) = (1, -2x_2)'$ and hence $\sigma_Q^2 = \pi_{11} - 4\mathbb{E}X_{2,t}\pi_{12} + 4(\mathbb{E}X_{2,t})^2\pi_{22}$.

In fact, the most delicate step in the approach described above is constructing an estimator $\widehat{\Pi}^{(k)}$ for the covariance matrix $\Pi^{(k)}$. The covariance matrix $\Pi^{(k)}$ is defined as

$$\Pi^{(k)} = (\pi_{ij})_{i,j=1,\ldots,k} = \left(\int_{-\infty}^{\infty} \gamma_{X_i,X_j}\,(u)\,du\right)_{i,j=1,\ldots,k}, \tag{1.22}$$

where

$$\gamma_{X_i,X_j}\,(u) = \mathrm{Cov}(X_{i,t}, X_{j,t+u}) \tag{1.23}$$

is the (cross) autocovariance function of the processes $\{X_{i,t}\}$ and $\{X_{j,t}\}$ at lag u. The matrix $\Pi^{(k)}$ given by (1.22) is known as the long-run variance (matrix) of the k-variate stationary process $\{X_t^{(k)}\} = \{(X_{1,t}, \ldots, X_{k,t})\}$. Possibly up to a multiplicative constant factor, the long-

run variance $\Pi^{(k)}$ is also the spectral density matrix at zero frequency of the k-variate stationary process $\{X_t^{(k)}\}$.

A considerable literature in Statistics on developing suitable estimators for the long-run variance $\Pi^{(k)}$ goes back at least to Parzen [78], in connection to estimation of a spectral density. Time-domain approaches for discrete time series have been revisited in the seminal work of Andrews [3]. The continuous-time framework presented here borrows from Lu and Park [64]. Thus, the long-run variance $\Pi^{(k)}$ is commonly estimated in the literature through an estimator of the type

$$\widehat{\Pi}^{(k)} = (\widehat{\pi}_{ij})_{i,j=1,...,k} \tag{1.24}$$

with

$$\widehat{\pi}_{ij} = \int_{-T}^{T} K\left(\frac{u}{S_T}\right) \widehat{\gamma}_{X_i,X_j}(u)\, du, \tag{1.25}$$

where S_T (smaller than the sample size T) is referred to as a bandwidth; K is a function, known as a kernel function, which is symmetric, satisfies $K(0) = 1$ and may have additional properties; and $\widehat{\gamma}_{X_i,X_j}(u)$ is the sample (cross) autocovariance function. The latter is defined in a standard way as a sample autocovariance, although a multiplicative factor of $1/T$ (instead of $1/(T-u)$ or another form) is used in the definition. An example of a commonly used kernel is the Bartlett (or triangular) kernel defined as $K(v) = 1 - |v|$ if $|v| < 1$ and $K(v) = 0$ if $|v| \geq 1$.

The choice of the bandwidth S_T in (1.25) is motivated by the so-called bias-variance trade-off. The essential point here is that the bandwidth S_T that balances the bias and the variance depends on the underlying process itself and is commonly selected either by assuming a particular model for the underlying process or through a data-driven approach. See Chapter 5 for details.

CHAPTER 2

Confidence intervals for exceedance probabilities with application to extreme ship motions

2.1 Introduction

We describe first the real-life application which sets the directions and frames the questions pursued in this and the following chapters (Section 2.1.1). We then outline the contributions and the structure of this chapter (Section 2.1.2).

2.1.1 Motivation

The work in this chapter is motivated by applications to ship motions and, more specifically, their stability in irregular seas. See, for example, Lewis [57], Benford [18], Belenky and Sevastianov [11], Neves et al. [76] for more information on this research area. When it comes to ship stability, the focus is on several variables characterizing the ship motion including *roll* and *pitch* angles, which are, respectively, the rotational movements around longitudinal (stern-to-bow) and lateral (starboard-to-port side or right-to-left side) axes, as well as vertical and lateral *accelerations* in various locations on the ship. See Figure 2.1. The ship stability failures are related directly to the exceedance of certain values by these variables. For example, the exceedance of a certain roll angle can lead to a cargo shift (which then can lead to capsizing), loss or damage of cargo in containers on deck, or down-flooding internal volumes of a ship. A large enough acceleration can lead to serious injuries or even death of a crew and passengers, as well as cargo damage. Such occurrences are well known not only among the researchers working in the area but also often make it to the popular media.[1]

[1]Recent examples of accidents related to intact stability failures include: Ro/Ro Ferry *Aratere* on 3rd March 2006 (Maritime New Zealand, 2007), Cruise ship *Pacific Sun* on 30 July 2008 (Marine Accident Investigation Branch, 2009), Ferry *Ariake* on 13 November, 2009 (Transportation Safety Board, 2011), to name but a few.

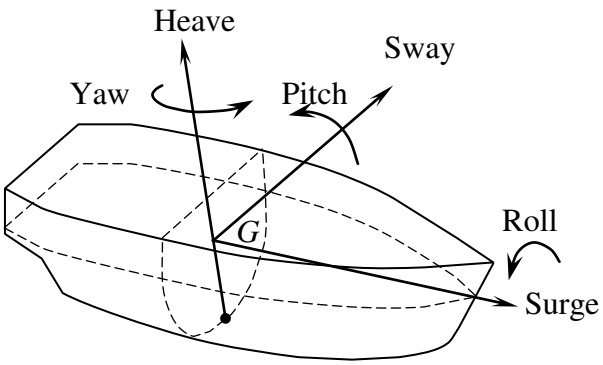

Figure 2.1: The motions of the ship.

The measured variables of interest to stability are understandably affected by the *geometry* and *loading* of the ship, the *operational parameters* and the surrounding *sea*. The operational side includes the *heading* (the angle between the vector of ship speed and predominant direction of wave propagation) and the value of *speed* of the ship. The state of the sea is usually described by a spectrum of wave elevations. Note that a wide range of *conditions* (the values of the above descriptors) are possible. What can be expected under a particular condition is often suggested from the understanding of the dynamics governing the ship motion.

An appealing but also critical feature of the research area is the availability of computer programs simulating ship motions, see the recent state-of-the-art review by Reed et al. [84]. In this chapter, we use a fast volume-based ship motion simulation algorithm developed in Weems and Wundrow [104]. The developed code does not incorporate finer hydrodynamics features of ship motions such as the influence of a ship motion on wave pressure field (i.e. wave diffraction and radiation; cf. Large Amplitude Motion Program or LAMP, see Lin and Yue [61]). But it is considered qualitatively representative of ship motions and their extremes. Moreover, the code is fast enough (in fact, the only such realistic method available) to be used in validation, where very long time histories of ship motions are necessary (see Section 2.3 below).

Figure 2.2 depicts the time series of roll and pitch angles obtained by the above referenced code for a 10 minute time window at 0.5 second measurement intervals. The ship geometry is that of the ONR tumblehome top (Bishop et al. [19]). The heading is at 45 degrees, the speed is 6 knots, the waves are characterized by significant height of 9m and mean zero-crossing period of 10.65s which corresponds to 15s of the modal period, using Bretschneider spectrum in open ocean (Lewis [57]).

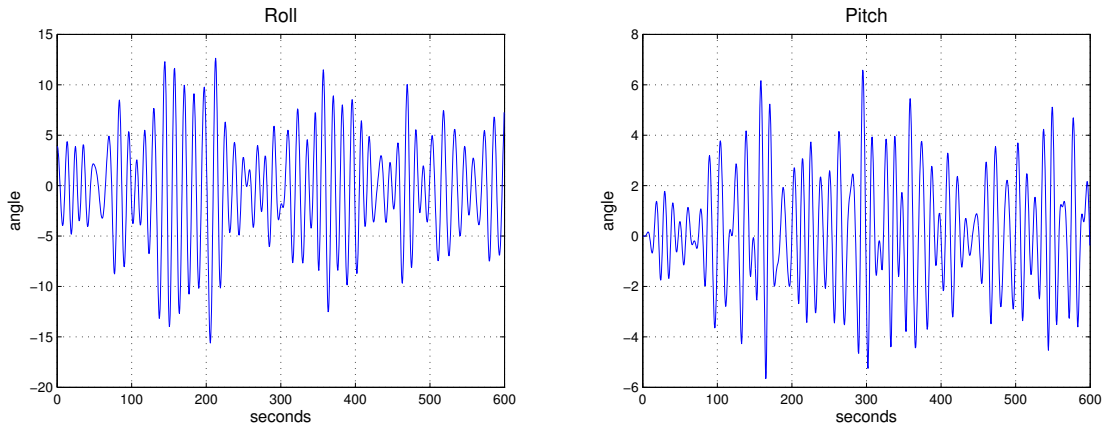

Figure 2.2: The roll and pitch angle series for 10 minutes.

A basic problem is to estimate the probability of roll, pitch or other variable of interest exceeding a critical value, as well as to provide a confidence interval. For example, in the condition of Figure 2.2, one could be interested in the roll angle exceeding 60 degrees (in either positive or negative direction). Inference would have to be made from the roll series of, for example, 100 hours, which would typically not contain such extreme occurrences. Again, the critical angle is often suggested from real-life considerations.

A method suggested for the problem above (and, more specifically, the associated confidence intervals) can be assessed through a validation procedure. The computer code mentioned above can be used to generate millions of hours of ship motion data which would contain exceedances of the target of interest. The "true" exceedance probability can then be estimated directly from this long history of the ship motion. In the validation procedure, the performance of the suggested method could be checked against the "true" exceedance probability at hand. See Section 2.3 for further details and a solution to the estimation problem.

2.1.2 Description of work and contributions

A natural mathematical framework to address the problem of estimating exceedance probabilities described above is the peaks-over-threshold (POT) approach (see, for example, Embrechts et al. [37], Coles [23], Beirlant et al. [7], as well as de Carvalho et al. [25], Ferreira and de Haan [39] for more recent related work). According to this approach, the probability of exceeding a given target of interest is computed as the product of the probability of exceeding

a smaller threshold and the (conditional) probability of exceeding the target above the threshold. The former probability is estimated simply as the proportion of data above the chosen threshold. The peaks over the threshold are modeled using the *generalized Pareto distribution* (GPD), whose complementary distribution function has the form

$$\overline{F}_{\mu,\xi,\sigma}(x) := \left(1+\frac{\xi(x-\mu)}{\sigma}\right)_+^{-1/\xi} := \begin{cases} \left(1+\frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}, & \mu < x, \quad \text{if } \xi > 0, \\ e^{-\frac{x-\mu}{\sigma}}, & \mu < x, \quad \text{if } \xi = 0, \\ \left(1+\frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}, & \mu < x < \mu - \frac{\sigma}{\xi}, \quad \text{if } \xi < 0, \end{cases} \tag{2.1}$$

where ξ is the *shape* parameter, σ is the *scale* parameter and μ is a *threshold*. Note that the GPD has an upper bound $(-\sigma/\xi)$ (above the threshold) for a negative shape parameter $\xi < 0$. When $\xi = 0$, the GPD is the usual exponential distribution.

We are interested here in what confidence intervals should be used for an exceedance probability. As indicated above, in the POT approach, this exceedance probability is the product of two probabilities, one of them being the exceedance probability for GPD. The questions then is what confidence intervals should be used for the exceedance probability in the GPD framework. The probability of the GPD exceeding a fixed *target* c (above the threshold), and its estimator are given by:

$$p_c = p_c(\xi,\sigma) = \left(1+\frac{\xi c}{\sigma}\right)^{-1/\xi}, \quad \widehat{p}_c = p_c(\widehat{\xi},\widehat{\sigma}) = \left(1+\frac{\widehat{\xi} c}{\widehat{\sigma}}\right)^{-1/\widehat{\xi}}, \tag{2.2}$$

where $\widehat{\xi}$ and $\widehat{\sigma}$ are some estimators of the shape and scale parameters, respectively. Somewhat surprisingly, the question of confidence intervals for the exceedance probability in (2.2) has apparently not been considered in much depth in the literature on extreme values. The paper by Smith [90], which laid the mathematical foundations for the ML estimators of the GPD, considers the problem of estimating the exceedance probability and provides the asymptotic normality result for the probability estimator (Section 8 of Smith [90]). This can in turn be used for confidence intervals but the normality assumption is not particularly appropriate (see Section 2.2 below).

Estimation of exceedance probabilities has also been considered by others but with different goals in mind. For example, Smith and Shively [91] are interested in trends for exceedance prob-

abilities. Exceedance probabilities in the spatial context appear in Draghicescu and Ignaccolo [34]. Considerable interest in exceedance (also sometimes referred to as failure) probabilities is when working with multivariate extremes. See, for example, de Haan and Sinha [28], de Haan and de Ronde [26], Heffernan and Tawn [49], Drees and de Haan [35].

Much of the focus in the extreme value analysis, on the other hand, has been on the related inverse problem of quantile estimation (see, for example, Embrechts et al. [37], Coles [23], Beirlant et al. [7]). The quantiles have been of greater practical interest in many applications driving the extreme value analysis, including finance (VaR calculations), insurance and hydrology (1-in-T years event). A closer look at the confidence intervals for quantiles can be found in Hosking and Wallis [51], Tajvidi [100] and also Section 4.3.3 of Coles [23], Section 5.5 of Beirlant et al. [7].

In applications to ship motions, as indicated in Section 2.1.1, it is common to look at the probabilities of exceeding a particular target rather than quantiles. Though perhaps not surprisingly, the two perspectives are also complementary. In fact, one of our findings is that the confidence intervals for exceedance probabilities perform well if constructed from those for quantiles. Another reason to focus on probabilities rather than quantiles is that probabilities can be aggregated naturally into "lifetime" probabilities, when integrated over a set of conditions of interest (as discussed, for example, in Section 1 of Belenky and Sevastianov [11]).

We study a number of ways to construct confidence intervals for the exceedance probability of the GPD and, more generally, in the POT framework in Section 2.2. We consider both direct methods, which are based on the functional form of exceedance probability (2.2) and the sampling distribution of the underlying estimators $\widehat{\xi}, \widehat{\sigma}$, and indirect (inverse) methods, which construct confidence intervals from those for quantiles.

The application of the considered confidence intervals to ship motions can be found in Section 2.3. In the validation procedure, the performance of the confidence intervals is analogous to that found under the idealized GPD framework. In particular, the methods recommended under the GPD framework also perform well and best in the application to ship motions. It should also be noted that the proposed solution is the first to address satisfactorily the estimation problem of the exceedance probabilities in ship stability. Some earlier attempts

include Belenky and Campbell [10] who used the Weibull distribution (instead of the GPD) to fit peaks over threshold, and McTaggart [68].

Finally, in Section 2.4, we discuss the issue of uncertainty (the size of confidence intervals) and its reduction. Conclusions can be found in Section 2.5.

2.2 Confidence intervals for exceedance probabilities

2.2.1 Methods for the GPD

We study and assess here several ways to construct confidence intervals for the exceedance probability p_c of the GPD given in (2.2). The probability is estimated through $\widehat{p}_c$ in (2.2) where we use the ML estimators $\widehat{\xi}$ and $\widehat{\sigma}$ computed from the sample $y_1, \ldots, y_n$ of size n. The large sample asymptotics of the ML estimators (Smith [90]) is the bivariate normal,

$$\sqrt{n}\begin{pmatrix} \widehat{\xi} - \xi_0 \\ \widehat{\sigma} - \sigma_0 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, W^{-1}), \tag{2.3}$$

where ξ_0, σ_0 denote the true values and

$$W^{-1} = \begin{pmatrix} 1+\xi_0 & -\sigma_0 \\ -\sigma_0 & 2\sigma_0^2 \end{pmatrix}. \tag{2.4}$$

In practice, the limiting covariance matrix can be estimated by replacing ξ_0 and σ_0 with their respective estimators $\widehat{\xi}$ and $\widehat{\sigma}$. Another common choice is to approximate nW through the observed information matrix

$$n\widehat{W} = \begin{pmatrix} -\frac{\partial^2}{\partial\xi^2} l(\xi,\sigma) & -\frac{\partial^2}{\partial\xi\partial\sigma} l(\xi,\sigma) \\ -\frac{\partial^2}{\partial\xi\partial\sigma} l(\xi,\sigma) & -\frac{\partial^2}{\partial\sigma^2} l(\xi,\sigma) \end{pmatrix}_{(\xi,\sigma)=(\widehat{\xi},\widehat{\sigma})}, \tag{2.5}$$

where $l(\xi,\sigma) = \sum_{i=1}^{n} \ln f_{\xi,\sigma}(y_i)$ is the log-likelihood and $f_{\xi,\sigma}(y)$ denotes the density of the GPD. Strictly speaking, the asymptotic result (2.3) holds for $\xi > -1/2$ only (Smith [90]). It should also be noted that other estimation methods than the MLE are possible for ξ_0 and σ_0. See, for example, a review paper by de Zea Bermudez and Kotz [29, 30] and references therein. Some of

these estimators outperform the ML estimators for small samples. For the sample sizes relevant to our problem of interest, the ML estimators seem to perform quite well and, in particular, to be approximately normal as stated in (2.3), and will be used throughout this chapter.

The exceedance probability $p_c = p_c(\xi, \sigma)$ in (2.2) is a function of ξ and σ, and is estimated through (2.2) by replacing the two parameters ξ and σ by their ML estimates. A confidence interval for p_c can then naturally be obtained through the standard delta method, using the asymptotic result (2.3). This is the approach seemingly adopted by Smith [90], Section 8. However, we found the delta method to perform poorly, in part because p_c can be very small and the normal approximation of $\widehat{p}_c$ may be sufficiently wide to include negative values. We have also tried the delta method for $\log p_c$ but the normal approximation did not appear to provide a good fit to $\log \widehat{p}_c$. Consequently, we consider below several, potentially more accurate ways to construct confidence intervals for the exceedance probabilities: the normal and lognormal methods, the boundary method, the bootstrap method, the profile likelihood method and the quantile method. The terminology behind the normal, lognormal, boundary and quantile methods are ours.

Normal method: The idea behind the normal method is still to use (2.3), which as mentioned earlier provides a good approximation in practice, but not to linearize the function $p_c(\xi, \sigma)$ (or $\log p_c(\xi, \sigma)$) as in the unsatisfactory delta method. In fact, assuming the bivariate normal approximation for $\widehat{\xi}$ and $\widehat{\sigma}$ according to (2.3), we can derive the exact distribution of $\widehat{p}_c$ as follows. Observe that the distribution function of $\widehat{p}_c$ is: for $0 \le z \le 1$,

$$F_{\widehat{p}_c}(z) = P\Big(\Big(1 + \frac{\widehat{\xi} c}{\widehat{\sigma}}\Big)^{-1/\widehat{\xi}} \le z\Big) = P\Big(\Big(1 + \frac{\widehat{\xi} c}{\widehat{\sigma}}\Big)^{-1/\widehat{\xi}} \le z, 1 + \frac{\widehat{\xi} c}{\widehat{\sigma}} > 0\Big) + P\Big(1 + \frac{\widehat{\xi} c}{\widehat{\sigma}} \le 0\Big),$$

where we use the fact that $\widehat{p}_c = 0$ if $1 + \widehat{\xi} c/\widehat{\sigma} \le 0$. This can further be expressed as

$$F_{\widehat{p}_c}(z) = P\Big(\widehat{\sigma} \le \frac{\widehat{\xi} c}{z^{-\widehat{\xi}} - 1}, \widehat{\sigma} > -\widehat{\xi} c\Big) + \mathbb{P}(\widehat{\sigma} \le -\widehat{\xi} c),$$

if we assume that $\widehat{\sigma}$ takes only positive values. (Note also that $\widehat{\xi}/(z^{-\widehat{\xi}} - 1) > 0$ for both $\widehat{\xi} < 0$ and $\widehat{\xi} > 0$.) Note, however, that it is not possible to have $\widehat{\sigma} > \widehat{\xi} c/(z^{-\xi} - 1)$ and $\widehat{\sigma} \le -\widehat{\xi} c$. Indeed, this is certainly not possible if $\widehat{\xi} > 0$, since then $-\widehat{\xi} c < 0$ and $\widehat{\xi} c/(z^{-\widehat{\xi}} - 1) > 0$. If

$\widehat{\xi} < 0$, on the other hand, this is not possible since $-\widehat{\xi}c \leq \widehat{\xi}c/(z^{-\widehat{\xi}}-1)$ or, equivalently, $z^{-\widehat{\xi}} < 1$. Hence, we also have

$$F_{\widehat{p}_c}(z) = P\Big(\widehat{\sigma} \leq \frac{\widehat{\xi}c}{z^{-\widehat{\xi}}-1}\Big) = \int_{\sigma \leq \xi c/(z^{-\xi}-1)} g_{\widehat{\xi},\widehat{\sigma}}(\xi,\sigma)d\xi d\sigma, \tag{2.6}$$

where $g_{\widehat{\xi},\widehat{\sigma}}(\xi,\sigma)$ denotes the bivariate normal density of the limit law (2.3) (replacing ξ_0 and σ_0 by $\widehat{\xi}$ and $\widehat{\sigma}$). In practice, the distribution function $F_{\widehat{p}_c}(z)$ is computed numerically and the $100(1-\alpha)\%$ confidence interval is set as (z_1, z_2) where $z_j = \inf\{z : F_{\widehat{p}_c}(z) \geq \alpha_j\}$, $j = 1, 2$, where $\alpha_1 = \alpha/2$ and $\alpha_2 = 1 - \alpha/2$. We use the generalized inverse in the last expression since $F_{\widehat{p}_c}(z)$ can have a discontinuity (mass) at $z = 0$.

Lognormal method: In the normal method above, we assumed that $\widehat{\sigma}$ does not take negative values or that, from a practical perspective, the probability of $\widehat{\sigma}$ being negative according to (2.3) is negligible. This may not be the case for smaller values of σ and sample sizes n. A natural way to address this is by parameterizing the GPD through ξ and $\ln\sigma$, instead of σ. The difference is that $\ln\sigma$ now takes possibly negative values. The asymptotic normality result then becomes

$$\sqrt{n}\Big(\begin{array}{c}\widehat{\xi} - \xi_0 \\ \widehat{\ln\sigma} - \ln\sigma_0\end{array}\Big) \xrightarrow{d} \mathcal{N}(0, W_1^{-1}), \tag{2.7}$$

where

$$W_1^{-1} = \mathrm{diag}\{1, \sigma_0^{-1}\}W^{-1}\mathrm{diag}\{1, \sigma_0^{-1}\}. \tag{2.8}$$

Arguing as in the normal method above, we have

$$F_{\widehat{p}_c}(z) = P\Big(\widehat{\ln\sigma} \leq \ln\frac{\widehat{\xi}c}{z^{-\widehat{\xi}}-1}\Big) = \int_{\ln\sigma \leq \ln(\xi c/(z^{-\xi}-1))} g_{\widehat{\xi},\widehat{\ln\sigma}}(\xi,\ln\sigma)d\xi d\ln\sigma, \tag{2.9}$$

where $g_{\widehat{\xi},\widehat{\ln\sigma}}(\xi,\ln\sigma)$ denotes the bivariate normal density of the limit law (2.7). The confidence interval can then be computed as in the normal method above. We shall refer to this as the lognormal method. A nice feature of the normal and lognormal methods is that they provide confidence intervals even in the case when $\widehat{\xi} < 0$ and the target is beyond the estimated support bound $(-\widehat{\sigma}/\widehat{\xi})$.

Boundary method: The normal and lognormal methods described above involve a relatively intensive numerical computation of the integrals (2.6) and (2.9). An approximate confidence interval which is fast to compute and easy to implement, can be constructed through the following boundary method. That is, take the confidence interval as

$$\Big(\min_{j,k=1,2} p_c(\xi_j, \sigma_k), \max_{j,k=1,2} p_c(\xi_j, \sigma_k) \Big), \tag{2.10}$$

where ξ_1, ξ_2 and σ_1, σ_2 are suitable critical values of the distributions of $\widehat{\xi}$ and $\widehat{\sigma}$, respectively. If $\widehat{\xi}$ and $\widehat{\sigma}$ were asymptotically uncorrelated, it would be natural to consider $\xi_j = \widehat{\xi} \pm C_{\sqrt{\alpha}} \mathrm{se}_{\widehat{\xi}}$ and $\sigma_k = \widehat{\sigma} \pm C_{\sqrt{\alpha}} \mathrm{se}_{\widehat{\sigma}}$, where se stands for standard error, C_β denotes the $100(\beta/2)\%$ quantile of the standard normal distribution and $(1-\alpha)\%$ is the confidence level sought. To account for the correlation between $\widehat{\xi}$ and $\widehat{\sigma}$, we take

$$\begin{pmatrix} \xi_j \\ \sigma_k \end{pmatrix} = V \begin{pmatrix} \xi_{0,j} - \widehat{\xi} \\ \sigma_{0,k} - \widehat{\sigma} \end{pmatrix} + \begin{pmatrix} \widehat{\xi} \\ \widehat{\sigma} \end{pmatrix}, \tag{2.11}$$

where $n^{-1}W^{-1} = VDV'$ with a diagonal $D = \mathrm{diag}\{d_1, d_2\}$ and $\xi_{0,j} = \widehat{\xi} \pm C_{\sqrt{\alpha}} \sqrt{d_1}$ and $\sigma_{0,k} = \widehat{\sigma} \pm C_{\sqrt{\alpha}} \sqrt{d_2}$. Note that the confidence intervals obtained by the boundary method are expected to be conservative. Indeed, the region determined by the points (ξ_j, σ_k) can be thought as the $100(1-\alpha)\%$ confidence region for the parameters ξ_0 and σ_0. But since $p_c(\xi, \sigma)$ is not a one-to-one function, there are points (ξ, σ) outside the confidence region for which the value $p_c(\xi, \sigma)$ falls inside the confidence interval (2.10).

Bootstrap method: The bootstrap method is somewhat standard with the confidence interval determined by the $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ quantiles of the bootstrap distribution of the exceedance probability.

Profile (likelihood) method: The profile (likelihood) method refers to another standard method to construct confidence intervals based on the profile likelihood. This is achieved by first expressing σ as a function of ξ and the exceedance probability p_c,

$$\sigma = \frac{\xi c}{p_c^{-\xi} - 1},$$

then parameterizing the likelihood in terms of ξ and p_c (instead of σ), and finally constructing the confidence interval based on the profile likelihood in a standard way. (See Coles [23] for the same approach when estimating a return level, instead of an exceedance probability.) Since the exceedance probability is constrained to be nonnegative, the use of the profile likelihood may be questionable.

Quantile method: Finally, the quantile method actually refers to a set of methods. The basic idea is the following. Exceedance probabilities p (p_c above) are associated with respective return levels (quantiles) x_p (c above) of the GPD distribution. A return level x_p can be estimated with a confidence interval $\widehat{x}_p \pm m_p$. Any of the methods discussed above (normal, lognormal, boundary, bootstrap and profile) can be adapted to construct a confidence interval for x_p – the difference being that the function (2.2) is now the return level

$$x_p = x_p(\xi, \sigma) = \frac{\sigma}{\xi}\Big(p^{-\xi} - 1\Big). \tag{2.12}$$

Moreover, the plot of $(-\ln p, \widehat{x}_p)$ with added confidence intervals is known as a return level plot (e.g. Coles [23]). To indicate the underlying method used to set confidence intervals for return levels, we will refer to the quantile method as quantile-boundary, quantile-lognormal, etc. A natural way to set a confidence interval for the exceedance probability p_c of the level c is then

$$(p_1, p_2), \tag{2.13}$$

where $p_1 = \inf\{p : \widehat{x}_p + m_p \geq c\}$ and $p_2 = \inf\{p : \widehat{x}_p - m_p \geq c\}$ (with $\inf\{\emptyset\} = 0$). See Figure 2.3. For the parameter values considered below, the functions $\widehat{x}_p + m_p$ and $\widehat{x}_p - m_p$ are increasing and continuous in the argument $(-\ln p)$. The quantile approach is appealing in that it makes estimation of exceedance probabilities and return levels consistent.

In the reliability context and for a location-scale family of distributions, the quantile approach was studied in Hong et al. [50] (see also Section I-C therein for earlier uses of connections between confidence intervals for quantiles and exceedance probabilities).

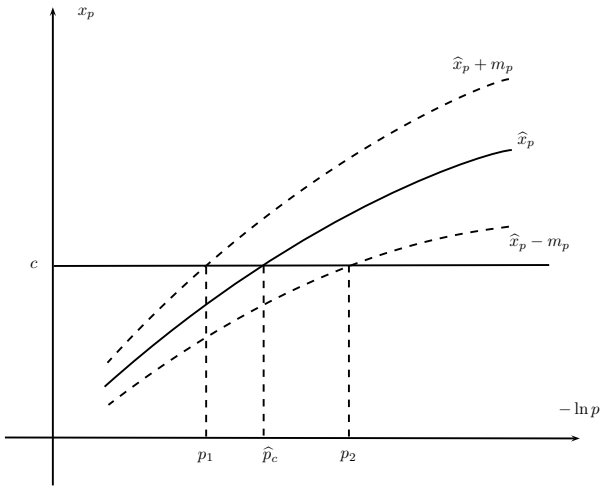

Figure 2.3: The quantile method to set confidence intervals for exceedance probability.

2.2.2 Simulation study for the GPD

We examine here the confidence intervals proposed in Section 2.2.1 through a simulation study. The empirical coverage frequencies of the confidence intervals (based on 500 Monte Carlo replications) are reported in Tables 2.1 and 2.2 for the sample sizes $n = 100$ and $n = 50$, respectively. The sample size of approximately $n = 100$ is a typical value that we encounter in the application to ship motions described in Section 2.3 below. The results are also presented for the smaller sample size $n = 50$, since in practice, one does not expect many peaks over a threshold for which the GPD is used as a model.

The first four columns in the tables present the true values of the parameters ξ_0, σ_0, and also the target c and the corresponding exceedance probability p_c. The values of $\xi_0 = \pm.1$ are some of the typical values encountered in our application of interest. When $\xi_0 = .6$, the GPD has infinite variance but finite mean. σ_0 is just a scale parameter, which we fix at 1. For the other two true parameters, we fix the exceedance probability p_c and compute the respective target c.

The other columns of the tables correspond to the methods considered. The normal, log-normal and boundary methods use the limiting covariance matrix W^{-1}/n in (2.3). It is approximated by the inverse of the observed information matrix (2.5), which we found to yield better results than using, for example, the expression (2.4) (with ξ_0, σ_0 replaced by $\widehat{\xi}, \widehat{\sigma}$). The bootstrap method is based on 500 bootstrap replications. Finally, for the quantile methods, we

consider three ways to construct confidence intervals for the return levels: lognormal, boundary and profile.

A number of observations can be drawn from Tables 2.1 and 2.2. The normal and lognormal methods are slightly anti-conservative, with the lognormal method preferred. The reason for the methods being anti-conservative is the estimation of the limiting covariance matrix W^{-1}/n in (2.3). The intervals have the expected coverage probability if the true covariance matrix (2.4) is taken (the exact coverage probabilities not reported here). As claimed in Section 2.2.1, the boundary method yields slightly conservative confidence intervals. The bootstrap and profile methods do not work well, especially for the value of ξ_0 close to zero or negative. Again, we suspect that this is due to the fact that the probability cannot be negative. Issues with bootstrap for the GPD were also reported and studied in Tajvidi [100].

Turning to the quantile methods, the quantile-lognormal method is slightly anti-conservative, as is the direct lognormal method. The quantile-boundary method is, on the other hand, slightly conservative. The quantile-profile method seems to perform best, with the coverage probabilities consistently close to the nominal level. Note that the profile-likelihood method for return levels does not have such pronounced limitation of the same method for exceedance probabilities – although it is true that a return level cannot be negative, the confidence interval would rarely reach zero. Note also that the results for $n = 100$ and $n = 50$ are comparable. One notable difference is that the quantile methods become slightly more anti-conservative when the sample size is reduced from $n = 100$ to $n = 50$.

In conclusion, the quantile method based on profile likelihood seems to perform best among the methods considered. The (log)normal and boundary methods, for both direct and indirect (quantile) approaches, can also be recommended but keeping in mind their (anti)conservative nature. Finally, we also note that the direct (log)normal and boundary methods are computationally less intensive compared to the indirect (quantile) methods.

2.2.3 The POT framework

Suppose now that $x_1, \ldots, x_N$ are i.i.d. observations of a general (i.e. non-GPD) random variable X, and that we are interested in estimating the probability $\mathbb{P}(X > x_{cr})$ of the variable X exceeding a critical value x_{cr}. Again, in the peaks-over-threshold (POT) approach, the

true values				direct methods					quantile methods		
ξ_0	σ_0	c	p_c	norm	logn	bound	boot	profl	logn	bound	profl
$-.1$	1	6.02	10^{-4}	90.4	90.8	96.2	68.2	76.2	92	97	95
		6.84	10^{-5}	95.2	95.6	97.6	65.6	78.8	88.8	96.6	94.6
		7.49	10^{-6}	94	94.6	96	74.6	80.8	91.2	96.8	94.2
.1	1	15.12	10^{-4}	92.6	93.2	98	87	98.4	89.6	97	94.2
		21.62	10^{-5}	90.6	91.2	97.2	82.8	97.6	92.4	98.6	95.2
		29.81	10^{-6}	91.2	92.6	97.6	81.2	97.8	91.2	98.4	94.6
.3	1	49.5	10^{-4}	91.8	92.4	98	89	97.2	89.4	97.4	92.2
		102.08	10^{-5}	88.8	89.2	98.4	86.6	97.6	92.2	97.8	95
		206.99	10^{-6}	93.4	94.2	99	91.8	98.6	92.6	98.4	94.4
.6	1	416.98	10^{-4}	90.8	90.8	97.8	91	93.8	92.4	98.6	94.2
		1665	10^{-5}	92.8	93.2	98	92.6	95.6	92.6	98.4	94
		6633.45	10^{-6}	94	93.8	98.6	92	95.8	93.4	99.4	95.2

Table 2.1: Coverage frequencies for confidence intervals when $n = 100$

true values				direct methods					quantile methods		
ξ_0	σ_0	c	p_c	norm	logn	bound	boot	profl	logn	bound	profl
$-.1$	1	6.02	10^{-4}	90.2	91.8	96.2	65.8	78	83.6	94	92.8
		6.84	10^{-5}	96	96	94.4	69.8	71.8	89.6	95.4	93.4
		7.49	10^{-6}	95.4	95.8	93.8	74.4	75.4	88.2	94.4	92.2
.1	1	15.12	10^{-4}	91.4	92.2	97.6	75.6	98.4	89.9	97.4	92.6
		21.62	10^{-5}	87.6	89	96.6	67	98.2	88.4	96.8	94
		29.81	10^{-6}	90.8	92.2	96.2	70	98.2	88.8	96.4	93
.3	1	49.5	10^{-4}	88.2	91	96	88	99	90	97.2	95.8
		102.08	10^{-5}	85.8	87.8	94.4	82.6	98.6	90.2	96.8	93.4
		206.99	10^{-6}	88.4	90	96.8	85.4	98.2	90.8	96.6	93.8
.6	1	416.98	10^{-4}	80.4	90.6	98.2	88	96.4	89.8	98.4	93.4
		1665	10^{-5}	80.8	91.4	97	89.8	98.2	89.4	96.8	93
		6633.45	10^{-6}	83.4	90.2	97.8	89.4	98	92.6	97.4	92.8

Table 2.2: Coverage frequencies for confidence intervals when $n = 50$

probability is written as

$$\begin{aligned} \mathbb{P}(X > x_{cr}) &= \mathbb{P}(X > u)\mathbb{P}(X > x_{cr}|X > u) \\ &= \mathbb{P}(X > u)\mathbb{P}(X - u > x_{cr} - u|X > u) =: \mathbb{P}_{nr} \cdot \mathbb{P}_r, \end{aligned} \tag{2.14}$$

where u stands for an intermediate threshold, and the subscripts nr and r refer to the non-rare and rare problems, respectively. The non-rare probability is estimated directly from the data as the proportion of data above the threshold u, $\widehat{\mathbb{P}}_{nr} = \sum_{j=1}^{N} 1_{\{x_j > u\}}/N$, with the respective confidence interval based on standard binomial calculations. The rare probability is estimated supposing that the peaks over threshold $Y = X - u$ follow a GPD, and setting

$$\widehat{\mathbb{P}}_r = \widehat{p}_{x_{cr}-u},$$

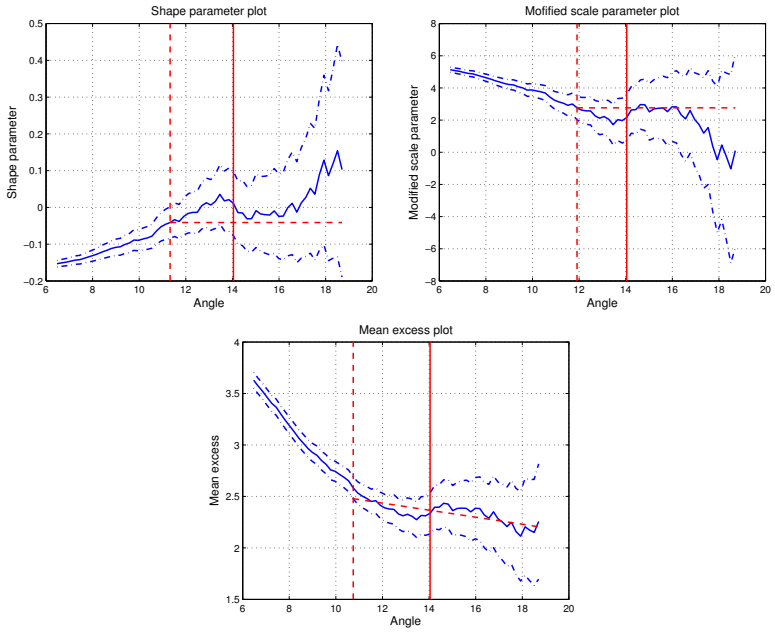

Figure 2.4: Shape parameter, modified scale parameter and mean excess plots. The vertical dashed line indicates the thresholds chosen with the corresponding (horizontal or arbitrary) lines passing through the confidence intervals for the larger thresholds. The vertical solid line indicates the threshold choice using the Reiss and Thomas method.

where $\widehat{p}_c$ is the exceedance probability (2.2) in the GPD framework, estimated from the data $y_i = x_{i'} - u$ of the peaks exceeding the threshold. The confidence intervals for $\mathbb{P}_r = p_{x_{cr}-u}$ are constructed by one of the methods of Section 2.2.1. The confidence interval for the original exceedance probability $\mathbb{P}(X > x_{cr})$ is obtained by multiplying the respective endpoints of the confidence intervals of $\mathbb{P}_{nr}$ and $\mathbb{P}_r$.

Threshold selection has been discussed and studied by many authors (for example, a review is given in Scarrott and MacDonald [88]) and is not the focus here. A special feature of the application to ship motions discussed in Section 2.3 is that the threshold selection should be automated, but with the possibility of closer examination if needed. The automatic selection is naturally sought in the ship motion application because multiple records need to be analyzed for the accuracy that is meaningful for practical applications.

In the automatic selection that we use, the threshold u is selected as the maximum of the thresholds u_{sh}, u_{ms}, u_{me} and u_{rt} chosen by the following four automatic procedures. The thresholds u_{sh}, u_{ms} and u_{me} are selected automatically from the commonly used shape parameter, modified scale parameter and mean excess plots, respectively. For example, the plot of the

non-GPD						direct		quantile		
model	parameters	N	n	c	p_c	logn	bound	logn	bound	profl
Weibull	$(\lambda, \tau) = (1, 1/2)$	2000	126	132.5	10^{-5}	90.0	99.2	97.0	99.6	95.0
			123	190.9	10^{-6}	92.2	99.4	94.8	98.8	93.4
	$(\lambda, \tau) = (1, 2)$	2000	194	3.4	10^{-5}	94.6	96.4	90.0	96.4	94.4
			195	3.7	10^{-6}	94.4	97.2	86.8	94.2	93.2
Burr	$(\beta, \tau, \lambda) = (1, 2, 2)$	2000	221	17.8	10^{-5}	95.6	99.2	90.4	97.2	92.8
			210	31.6	10^{-6}	96.4	99.6	87.4	94.8	92.4
Reverse	$(\beta, x_+) = (0.1, 10)$	2000	156.5	9.8	10^{-5}	96.8	93	83.4	92.4	90.8
Burr	$(\tau, \lambda) = (2, 2)$		150	9.9	10^{-6}	98.2	92.2	80.4	90.2	89.0

Table 2.3: Empirical coverage frequencies in the non-GPD context using the POT approach

estimated shape parameters with confidence intervals (against thresholds) should be about constant over the range where GPD fit is appropriate. The threshold u_{sh} is chosen as the smallest threshold for which the horizontal line drawn from the corresponding estimate passes through the confidence intervals of the shape parameter for all the larger thresholds. The thresholds u_{ms} and u_{me} are chosen similarly except that the line in the mean excess plot does not need to be horizontal. The choice of the three thresholds is illustrated in Figure 2.4, for one of the data sets considered in Section 2.3 below.

The threshold u_{rt}, on the other hand, is selected following the Reiss and Thomas [86], p. 137, automatic procedure (see also Neves and Fraga Alves [75]). Let $\xi_{k,n}$ be the estimates of the shape parameter ξ based on the k largest values of y_i (by using the moment estimation for computational efficiency). Choose k^* as the value that minimizes

$$\frac{1}{k}\sum_{i\le k} i^{\beta}|\xi_{k,n} - \text{med}(\xi_{1,n}, \ldots, \xi_{k,n})|,$$

where $\beta = 1/2$ (though other values of $\beta < 1/2$ can be considered as well) and med denotes the median. In practice, after the suggestion of Reiss and Thomas, the function above is slightly smoothed. The threshold u_{rt} is then chosen as the k^* largest value of y_i. It is depicted as a vertical solid line in Figure 2.4 and probably better corresponds to a visually desired choice of threshold. In our experience, the Reiss and Thomas choice most often provides the largest (most conservative) value among the methods considered.

Table 2.3 presents the empirical coverage frequencies of the confidence intervals constructed through the above POT approach for several non-GPDs. The distributions considered are: the

Weibull distribution with the CDF

$$f_{x_{\max}}) = 1 - e^{-\lambda x^{\tau}}, \quad x > 0,$$

with parameters $\lambda > 0$, $\tau > 0$; the Burr distribution with the CDF

$$f_{x_{\max}}) = 1 - \Big(\frac{\beta}{\beta + x^{\tau}}\Big)^{\lambda}, \quad x > 0,$$

with parameters $\lambda > 0$, $\tau > 0$, $\beta > 0$; and the reverse Burr distribution with the CDF

$$f_{x_{\max}}) = 1 - \Big(\frac{\beta}{\beta + (x_+ - x)^{-\tau}}\Big)^{\lambda}, \quad x < x_+,$$

with parameters $\lambda > 0$, $\tau > 0$, $\beta > 0$. Two choices of the parameter τ are considered for the Weibull distribution, with $\tau = 1/2$ ($\tau = 2$, resp.) providing heavier (lighter, resp.) tails than exponential (but both associated with the shape parameter $\xi = 0$ in the POT framework). The Burr distribution has a power-law tail, corresponding to the shape parameter $\xi = 1/(\tau\lambda)$ in the POT framework. Similarly, the reverse Burr distribution has a finite upper bound x_+, and corresponds to the negative shape parameter $\xi = -1/(\tau\lambda)$ in the POT framework.

Under the direct approach in Table 2.3, the coverage probabilities are reported only for the lognormal and boundary methods. The quantile methods use the proportion of data above the threshold to estimate $\mathbb{P}_{nr}$ but do not take the estimation uncertainty of $\mathbb{P}_{nr}$ into account. Two of the columns also give the sample size N and the average number of peaks over threshold n. As before, p_c is the exceedance probability and c is the corresponding critical target.

Our goal with Table 2.3 is not to provide an exhaustive study of the POT approach in the non-GPD framework, but rather to make a few general comments. First, note from the table that the approach works quite well. Second, note that the performance of the considered methods is not as uniformly good as in the GPD context. Thus, the performance of the methods for non-GPDs depends not only on the way to produce confidence intervals above a threshold but also on the non-GPD itself, as well as the (automatic) choice of the threshold.

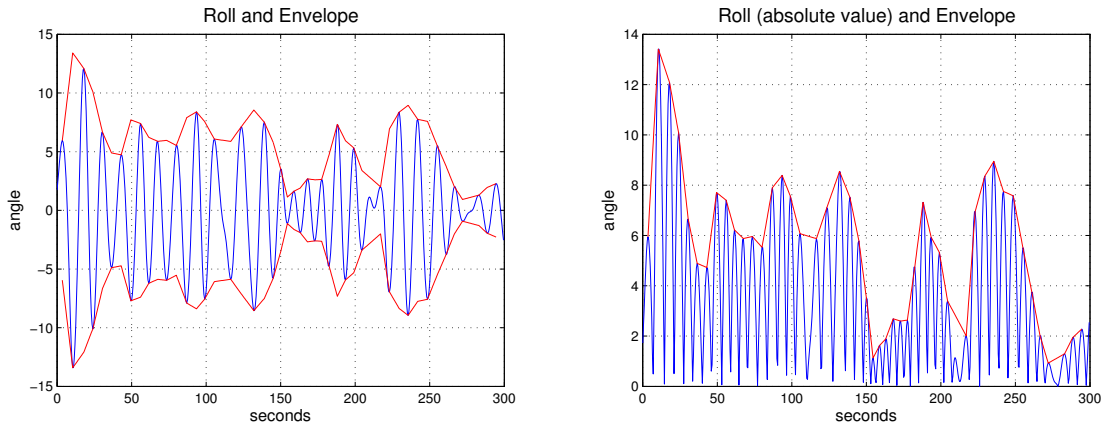

Figure 2.5: The roll angle series with envelope for 5 minutes. Left: original roll series. Right: roll series in absolute value.

2.3 Application to extreme ship motions

We shall use the POT approach outlined in Section 2.2.3 to estimate the probability of roll and pitch angle exceeding a critical value. Several issues need to be addressed before we can apply the methods for constructing confidence intervals discussed in Section 2.2.3. An important and pressing issue is the presence of temporal dependence as clearly seen from Figure 2.2. A related issue is also what is meant by an exceedance probability and how it relates to time.

The issue of temporal dependence is addressed through the following envelope approach. Motivated by the periodic nature of a ship motion, the maxima and minima are first found between consecutive zero crossings of the series. These are the positive and negative peaks in the series of interest. The absolute values of the peaks are then connected by a piecewise linear function producing an *envelope* of the series. This is depicted in Figure 2.5. The left plot includes the original roll series for 5 minutes, with the positive and negative envelope. The right plot depicts the absolute values of the roll and the positive envelope connecting linearly the absolute values of the peaks.

After the envelope is found for the whole roll time series (not just the 5 minutes shown), its average value is computed. Next, the maxima and minima are found in the envelope between consecutive crossings of the average envelope value. These are the envelope peaks above/below the envelope average. This is illustrated in Figure 2.6, where the envelope average is plotted

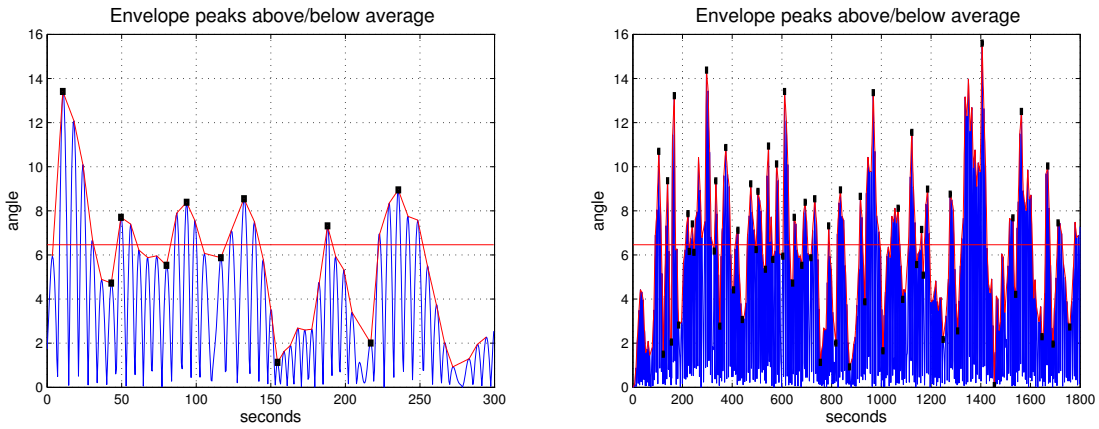

Figure 2.6: Envelope peaks above/below envelope average. Left: 5 minutes. Right: 30 minutes.

as a horizontal line and the envelope peaks above/below the envelope average are indicated by small black marks.

Note from Figure 2.6 that focusing on the envelope peaks (above the average) deals, at least qualitatively, with temporal dependence. That is, the larger values close in time are "clustered" and only the largest values in clusters are recorded as envelope peaks. (A closer look at the decorrelation properties of the envelope peak series can be found in a report by Belenky and Campbell [10].) In what follows, we shall work only with the envelope peaks. It is also important to note that the envelope approach is automated. This is particularly convenient when dealing with multiple conditions and many records.

Focusing on the envelope peaks also simplifies the notion of exceedance and the associated exceedance probabilities. Note that the series of interest will exceed a large target when an envelope peak will exceed the target. It is then natural to think of an exceedance probability as that for the envelope peaks. This is the perspective adopted throughout this chapter.

We should also clarify what we mean by probabilities, which are now related to the envelope peaks. Suppose a series contains $1,000$ envelope peaks of which 45 exceed a given threshold. Then, the estimated probability is $45/1000 = .045$ of exceeding the threshold. This probability is not informative without a reference to time. Suppose the series is actually recorded over 15 minutes or $15 \cdot 60 = 900$ seconds. It is then more informative to consider the (probability) rate of $45/900 = .05$ envelope peaks (over the threshold) per second. Though we will continue

referring to probabilities below, the results will be reported in terms of (probability) rates, rather than probabilities themselves.

If $x_1, \ldots, x_N$ are the envelope peaks of the series at hand, the exceedance probability is then estimated with a confidence interval as explained in Section 2.2.3. The performance of the confidence intervals can be assessed through a validation procedure as follows. The computer code (discussed in Section 2.1) can be used to generate significantly more series of ship motions, which contain rare events of interest and from which exceedance probabilities can be estimated by direct counting. More specifically, for the same condition used in Figures 2.2–2.6, the code was used to generate $115,000$ hours of the ship motion. With the target roll angle of $x_{cr} = 60$ degrees, the probability rate of exceedance obtained by direct counting based on rare events from the available records is 7.25×10^{-8} envelope peaks per second (that is, 30 envelope peaks above 60 degrees in $115,000$ hours). This "true" rate estimate can be supplemented by the confidence interval obtained by a standard binomial argument.

A typical given series (record) to make inference from covers only 100 hours and would not contain rare events of interest. For each record, confidence intervals for exceedance probabilities can be computed as in Section 2.2.3. The confidence intervals can then be assessed by their coverage frequencies of the "true" exceedance probability. This could be examined graphically as in Figure 2.7 where the lognormal, boundary, quantile-lognormal and quantile-profile confidence intervals are presented for 100 records of the total length of 100 hours. The critical value of interest is the roll of 60 degrees as above. Note that the vertical axis for the probability rate is in the log scale, and that we truncated the confidence intervals and the probability (rate) estimates at a practically negligible probability rate of 10^{-15}. The horizontal dashed lines indicate the confidence bounds for the "true" probability. The small circles are the probability rate estimates.

For the roll and pitch motion at 45 and 30 degree headings, we also report the coverage frequencies for the methods of Section 2.2 in Table 2.4, based on the results in 100 records. The columns under $\widehat{\xi}$ and n provide the average estimates of the shape parameter and the number of peaks over threshold. The standard errors are given in parentheses. In the parentheses under the coverage probabilities, we provide the average of the sizes of the suggested confidence

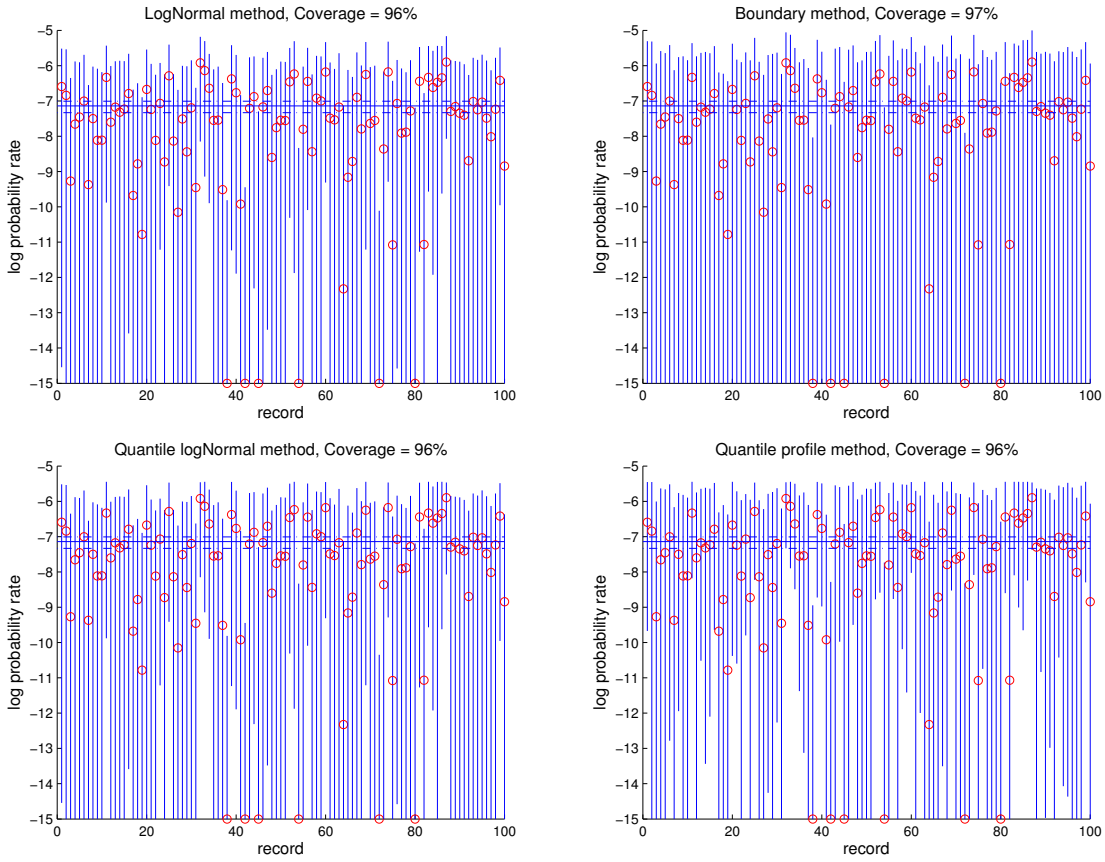

Figure 2.7: Confidence intervals for 100 records of length 100 hours. Roll series for 45° heading and with critical roll angle of 60°. Top left: lognormal method. Top right: boundary method. Bottom left: quantile-lognormal method. Bottom right: quantile-profile method.

intervals above the true value (supposing it is contained), which will be discussed further in Section 2.4 below.

Note from Table 2.4 that the performance of the confidence intervals is similar to those in Sections 2.2.2 and 2.2.3. Target values are chosen based on available rare events in the large set of records. The performance seems also satisfactory, validating the approach from a practical perspective. The point of using such validation is to show that the approach works on the ship motion data generated by a qualitatively correct computer code, before applying the methods to real or experimental data (where a large number of records are naturally not available). Or, put differently, had the methods not passed the validation, no applied researcher would be confident in using them.

The approach to estimate the exceedance probabilities certainly works in part because of the mathematical justification as outlined in Section 2.2.3. But this is not the whole story!

series				direct methods				quantile methods		
varble	head	$\widehat{\xi}$	n	logn	bound	boot	profl	logn	bound	profl
roll	45	0.19	96.06	96	97	92	100	96	97	96
		(0.13)	(31.76)	(1.19)	(1.46)	(1.15)	(1.52)	(1.18)	(1.41)	(1.40)
	30	0.04	105.03	84	91	76	99	84	91	89
		(0.13)	(46.96)	(0.82)	(1.13)	(0.87)	(1.37)	(0.82)	(1.12)	(1.07)
pitch	45	−0.06	107.06	99	100	95	98	99	100	100
		(0.11)	(50.92)	(0.62)	(0.73)	(0.57)	(0.73)	(0.62)	(0.73)	(0.74)
	30	−0.08	107.63	97	98	94	96	97	98	98
		(0.11)	(46.61)	(0.43)	(0.49)	(0.41)	(0.51)	(0.43)	(0.48)	(0.51)

Table 2.4: Headings of 30 and 45 degrees. Roll: target is 60 at 45° and 35 at 30°. Pitch: target is 10.

Another important component to success is related to the length of the record and the physics of the ship motion. The 100–hour records are typical for Naval Architecture purposes. Our results show that these records have sufficiently enough physics to allow one to extrapolate into the tail using the POT framework.

2.4 Uncertainty reduction

An interesting but also practically important question is whether the uncertainty of the estimators or, equivalently, the size of of the confidence intervals can be reduced. For example, in Figure 2.7, the right (top) endpoints of the confidence intervals are about one order of magnitude above the true value. One order seems acceptable from a practical perspective. But we also encounter conditions where the uncertainty could be as high as two or three orders of magnitude.

Can the uncertainty (or the size of confidence intervals) be reduced? It surely depends on the approach and model used (that is, the POT approach with the two parameter GPD above threshold), the sample size (that is, the number of exceedances above threshold), and the efficiency of the estimation method used. Efficiency cannot be improved considerably since the ML estimators of the GPD parameters are used. But several directions could be explored when it comes to the first two points.

More specifically, in Section 2.4.1, we study the situation where it may be meaningful to fix a right upper bound when a negative shape parameter is expected. A substantial uncertainty reduction is achieved with this approach but it may not be promising to search for extensions to positive shape parameters, or ways of fixing a bound. Section 2.4.2 contains a short and, in

our view, informative account of several other possibilities that we tried but which did not lead to much of the uncertainty reduction.

2.4.1 Fixing the upper bound

When the shape parameter of a GP distribution is negative, the distribution has a finite upper bound. One direction for uncertainty reduction is to fix this upper bound before estimation based on some physical considerations, e.g. limiting angle for roll after which ship capsizes. Fixing the bound reduces the number of parameters from 2 to 1, so that the reduction of uncertainty is expected.

In applications to ship stability, the pitch motion typically yields a negative shape parameter, as can already be seen from Table 2.4 (3rd column). There are physical reasons for this phenomenon which, in technical terms, have to do with the form of the stiffness of the pitch motion. Moreover, again for physical reasons, an upper bound for the pitch motion may be expected at about 15°–20°, as roll stiffness of ONR Tumblehome becomes flat and does not support any resonance excitation. Details of the physics of the pitch motion go beyond the scope of this chapter.

From a statistical standpoint, deriving the GPD framework with a fixed upper bound is straightforward. Suppose for notational simplicity that the threshold μ is 0, and denote a fixed upper bound by $y_{\max}$. When the shape parameter ξ of the GPD (2.1) is negative, the upper bound is given by $(-\sigma/\xi)$. Setting $y_{\max} = -\sigma/\xi$, solving for $\xi = -\sigma/y_{\max}$ and substituting this into (2.1) when $\xi < 0$, we obtain the complementary GPD function with the upper bound $y_{\max}$,

$$\overline{F}_\sigma(y) = \Big(1 - \frac{y}{y_{\max}}\Big)^{y_{\max}/\sigma}, \quad 0 < y < y_{\max}. \tag{2.15}$$

Note that the function (2.15) depends only on the scale parameter σ (with the shape parameter of the GPD being $\xi = -\sigma/y_{\max}$).

The parameter σ in (2.15) can be estimated using ML. Given observations $y_1, \ldots, y_n$ (all smaller than $y_{\max}$), optimizing the log-likelihood

$$\ell(\sigma) = \sum_{i=1}^{n} \log\Big(\frac{1}{\sigma}\Big(1 - \frac{y_i}{y_{\max}}\Big)^{y_{\max}/\sigma - 1}\Big)$$

leads to the ML estimator

$$\widehat{\sigma} = -\frac{y_{\max}}{n}\sum_{i=1}^{n}\log\Big(1-\frac{y_i}{y_{\max}}\Big). \tag{2.16}$$

The inverse of the observed information matrix can easily be checked to be

$$\Big(-\frac{\partial^2\ell}{\partial\sigma^2}\Big)^{-1}\Big|_{\sigma=\widehat{\sigma}} = \frac{\widehat{\sigma}^2}{n}. \tag{2.17}$$

A confidence interval for an exceedance probability $p_c = \overline{F}_{\sigma_0}(c)$ can then be given by the boundary method as $(\overline{F}_{\sigma_1}(c), \overline{F}_{\sigma_2}(c))$, where $\sigma_1 = \widehat{\sigma} - C_\alpha\widehat{\sigma}/\sqrt{n}$ and $\sigma_2 = \widehat{\sigma} + C_\alpha\widehat{\sigma}/\sqrt{n}$ are two critical values for the distribution of $\widehat{\sigma}$ based on (2.17) (with as before, C_α denoting the $100(\alpha/2)\%$ quantile of the standard normal distribution).

Figure 2.8 compares the confidence intervals for the exceedance probability of the pitch motion at the 30° heading (under the same condition as earlier) obtained through the lognormal method as in Section 2.3, and the boundary method with the upper bound fixed at 15° as explained above. The left plot in Figure 2.8 corresponds to the entry of Table 2.4 under "pitch", "30" degree heading and "logn" method, with the uncertainty measure of 0.43 in the parentheses. The same measure for the right-plot of Figure 2.8 is 0.34. The reduction of uncertainty is also evident from Figure 2.8 itself, with smaller variability of the estimators (red circles) and the sizes of confidence intervals in the right plot.

It should also be noted that the results with the fixed upper bound are not sensitive to the choice of the bound (suggested by physical considerations). For example, fixing the bound at 17° and 20° leads to the same coverage frequency of 99%, with the exception that the uncertainty measure above becomes slightly larger, 0.36 and 0.38, respectively. The conclusions are the same for the pitch motion at the 30° heading (not reported here).

Remark 2.4.1. Whether a similar approach can be developed for a positive shape parameter remains an open question. One idea we entertained was to experiment with truncated GPD models in the spirit of, for example, Aban et al. [1], Beirlant et al. [6]. (Truncation seems natural because, for example, the roll and pitch angles are bounded by 180 degrees.) But the truncated GPD models did not appear to fit the data well. In Chapter 3, we study extreme

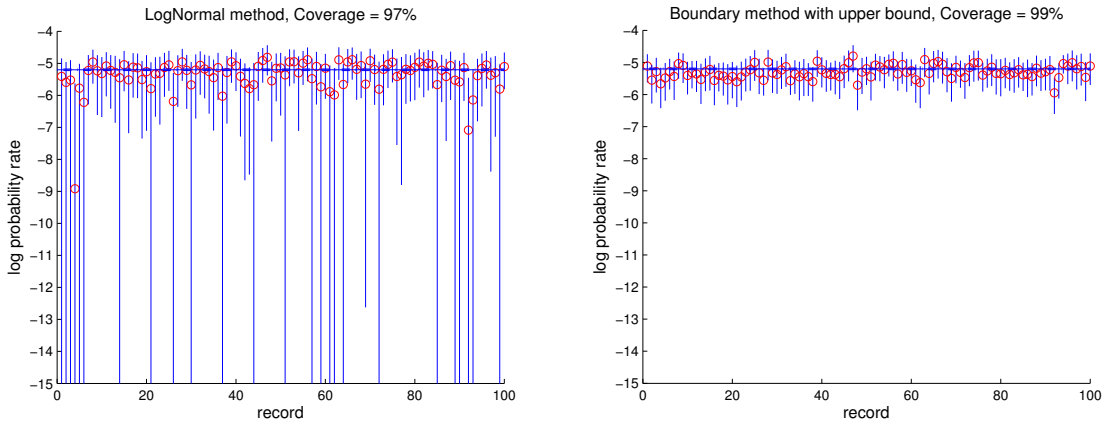

Figure 2.8: Confidence intervals for 100 records of length 100 hours. Pitch 30°. Left: lognormal method. Right: boundary method with fixed upper bound at 15°.

value methods on mathematically tractable physical models mimicking ship motion dynamics and gain further insight into the above issues from this approach.

2.4.2 Other possibilities

We explored or thought about several other possibilities for uncertainty reduction. One natural possibility would be to view the variables describing different conditions as covariates and then pool the data across different conditions by modeling covariates to reduce uncertainty. This idea is particularly relevant in the application of interest here since naval engineers have to take measurements regularly across a range of conditions. The idea also has a sound statistical footing, as developed in Davison and Smith [24] and described, for example, in Chapter 6 of Coles [23].

Following this approach, we have modeled records across a number of headings (e.g. 15°, 22.5°, 30°, 37.5°, 45° degrees). But we generally found the reduction in uncertainty small if any. Some of this is due to a small reduction of uncertainty even under ideal situations (when the model incorporating the covariates is known). The uncertainty in the underlying model for the covariates (entering the POT framework) also plays a role.

Finally, another possibility might be to use some of the more advanced approaches in modeling dependent peaks over threshold, as in e.g. Smith et al. [92]. The idea here is that this would seemingly allow for a larger sample size to be considered. Even if the dependence structure is captured correctly by these approaches, we also expect them to lead to little uncertainty

reduction. As with the covariates above, we view these approaches as serving different purposes and used to answer different questions.

2.5 Conclusions

In this chapter, we studied the various methods to construct confidence intervals for exceedance probabilities in the peaks-over-threshold approach. The performance of the confidence intervals was assessed through several simulation studies, pointing to the superior performance of some of the considered methods. The developed methods were applied to build confidence intervals for the probabilities of extreme ship motions, leading to satisfactory results overall. Finally, several uncertainty reduction approaches were considered, with a promising solution when a negative shape parameter is expected. Whether uncertainty reduction can be achieved in the case of a positive shape parameter remains an open question.

CHAPTER 3

Distribution tail structure and extreme value analysis of constrained piecewise linear oscillators

3.1 Introduction

Piecewise linear oscillators might seem to be exotic models for engineering applications as most of the real-world forces are smooth. Nevertheless, they have proved useful on several occasions, e.g. as in the classical problem of a dynamical system with dry friction (Andronov et al. [4]). Another application area concerning ship motions will be discussed in more detail below. Thus, consider a single-degree-of-freedom random oscillator given by

$$\ddot{x} + 2\delta\dot{x} + r(x) = y(t), \tag{3.1}$$

where $\delta > 0$ is a damping parameter, $r(x) = \nabla V(x)$ is a nonlinear restoring force (stiffness) associated with a potential function V and $y(t)$ is an external random excitation (forcing). The potential function is characterized by the existence of a stable center and two symmetric unstable saddle equilibrium points. The stiffness function associated with such potential is assumed to have a piecewise linear form given by

$$r(x) = \begin{cases} -k_1\omega_0^2\,(x + x_m) - \omega_0^2 x_m, & \text{if } x < -x_m, \\ \omega_0^2 x, & \text{if } -x_m \le x \le x_m, \\ -k_1\omega_0^2\,(x - x_m) + \omega_0^2 x_m, & \text{if } x > x_m, \end{cases} \tag{3.2}$$

where ω_0 is a natural frequency in the linear regime $(-x_m, x_m)$, $-k_1\omega_0^2 < 0$ is a negative slope in the nonlinear regime $|x| > x_m$ and x_m, called the "knuckle" point, defines the threshold above which the system behaves nonlinearly, i.e. the point above which the restoring force is

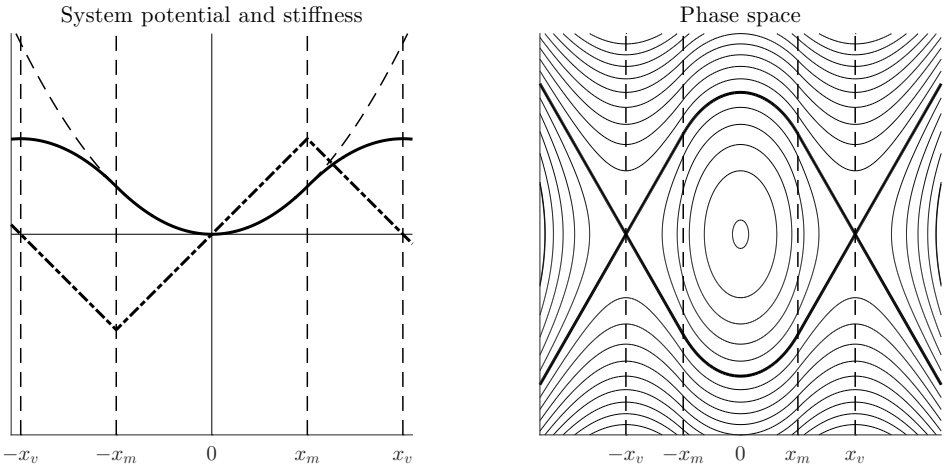

Figure 3.1: Left: a piecewise linear stiffness function $r(x)$ (dash-dotted curve) and its associated potential function $V(x)$ (solid curve). The potential function for the linear stiffness function is also shown (dashed curve), as are $-x_m$, x_m, $-x_v$, x_v (dashed vertical lines). Right: A phase portrait of the system on the left.

decreasing. The corresponding potential for this case is given by

$$V(x) = \begin{cases} -\frac{1}{2}k_1\omega_0^2x^2 - (k_1+1)\,\omega_0^2x_mx - \frac{1}{2}\,(k_1+1)\,\omega_0^2x_m^2, & \text{if } x < -x_m, \\ \quad \frac{1}{2}\omega_0^2x^2, & \text{if } -x_m \le x \le x_m, \\ -\frac{1}{2}k_1\omega_0^2x^2 + (k_1+1)\,\omega_0^2x_mx - \frac{1}{2}\,(k_1+1)\,\omega_0^2x_m^2, & \text{if } x > x_m. \end{cases} \tag{3.3}$$

Figure 3.1 provides the plots of a generic piecewise linear stiffness function (3.2), the associated potential function (3.3) and the phase space (for the unforced, unperturbed system). The point $x_v > 0$ such that $r(x_v) = 0$ corresponding to the unstable equilibrium is referred to as the point of vanishing stability, and will play an important role below.

The model (3.1)–(3.3) is an attractive tool for a qualitative consideration of large roll motions of a ship in waves, including capsizing (understanding the latter as a transition to motion near another stable equilibrium). Indeed, the piecewise linear function (3.2) is a schematic model of an actual ship roll stiffness resulting from hydrostatic and hydrodynamic pressures

over the submerged portion of the ship hull. While this model can be seen as a very simplistic one, it retains most known nonlinear properties of an oscillator with a similar smooth stiffness, including loss of isochronism, fold and flip bifurcations, as well as fractalization of the safe basin (Belenky [9]). It also allows for a closed-form solution to the probability of capsizing within a given time (Belenky [8]). The model has seen some further development and applications, e.g. by Paroka and Umeda [77]. Another important outcome of the study of piecewise linear oscillators was the so-called "split-time" approach to find the probability of capsizing by a novel numerical simulation scheme; see Campbell et al. [21] for a review. This reference also contains an updated and rectified closed-form solution for capsizing with a piecewise linear stiffness.

Ship stability accidents are not limited to capsizing. Encountering large (extreme) roll angles can also have catastrophic impact, in terms of both human, cargo or machinery loss. Indeed, probabilistic characterization of extreme values of ship roll has attracted much attention in Naval Architecture. A Weibull distribution is employed for wave loads, e.g. for vertical bending moment (Lewis [56]). Significant nonlinearity of roll motion, however, prevents this conventional approach for probabilistic characterization of dynamic stability. Some of the approaches, accounting for nonlinearity of roll motions, have been reviewed in Belenky et al. [13]. Recent works include further development of the critical wave group method by Malara et al. [67], Anastopoulos et al. [2], and path integration method by Kougioumtzoglou and Spanos [53], Chai et al. [22]. Modern simulation tools for probabilistic characterization are reviewed in Reed et al. [84]. See also Reed and Zuzick [85] who focus specifically on accreditation for regulatory applications.

Another standard statistical approach to characterizing extremes is the peaks-over-threshold (POT) method based on fitting a generalized Pareto distribution (GPD) to data above a suitably chosen threshold (e.g. Beirlant et al. [7], Coles [23], Embrechts et al. [37]). The POT approach has been applied to ship motions and loads data (e.g. Guha et al. [47], Campbell et al. [20, 21], Chapter 2 of this dissertation) and has generally been found to per-

form well in characterizing extremes. The GPD density is expressed as

$$f_{\mu,\xi,\sigma}(x) = \begin{cases} \frac{1}{\sigma}\left(1+\frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi-1}, & \mu < x, \qquad \text{if } \xi > 0, \\ \frac{1}{\sigma}e^{-\frac{x-\mu}{\sigma}}, & \mu < x, \qquad \text{if } \xi = 0, \\ \frac{1}{\sigma}\left(1+\frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi-1}, & \mu < x < \mu - \frac{\sigma}{\xi}, \quad \text{if } \xi < 0, \end{cases} \tag{3.4}$$

where ξ is the *shape* parameter, σ is the *scale* parameter and μ is a *threshold*. The GPD has an upper bound $(-\sigma/\xi)$ (above the threshold) for a negative shape parameter $\xi < 0$. This case shall be referred to as that of a light tail. When $\xi = 0$, the GPD is the usual exponential distribution. When $\xi > 0$, on the other hand, the GPD density has a power-law tail behavior $Cx^{-1/\xi-1}$ with constant $C > 0$ and exponent $-1/\xi - 1$. In this case, the tail is heavy, especially compared to the tail of a Gaussian distribution. Unless specified otherwise, a heavy tail will refer to a distribution tail that is heavier than that of a normal distribution, and the term power-law tail or power-law heavy tail will refer to the tail behavior $Cx^{-1/\xi-1}$ with $\xi > 0$.

The GPD is an asymptotic distribution, while ship roll data arise as a response of a strongly nonlinear dynamical system. The available data does not necessarily provide sufficient information about the system, and the approximation of the tail with GPD may not necessarily reflect real limiting behavior of the nonlinear system. For example, if the available data on the response of the system (3.1) does not include any transitions to another stable equilibrium (capsizing), the POT method will not be able to predict it, unless some additional information is provided. Thus, the solution has to be sought in complementing statistical analysis of the response with mathematical models reflecting physics of particular nonlinear dynamical systems.

A practical step towards this objective is to try to relate a character of nonlinearity and a type of statistical estimate of the response. Belenky and Sevastianov [11] described an influence of the initial form of the roll stiffness on the deviation of roll distribution from normal. Mohamad and Sapsis [71] and Mohamad et al. [70] introduced a probabilistic decomposition method to describe the influence of instabilities on the heavy-tailed statistics of general dynamical systems. Consideration of an intermittent resonance allows applying this approach to parametric roll (Mohamad and Sapsis [72]), where it successfully reproduces a shape of a distribution observed in high-fidelity numerical simulation (Belenky and Weems [12]). Another

example is the introduction of a limiting upper bound to pitch motions (Chapter 2), reflecting the fact that a ship loses sensitivity to wave excitation once pitch angle exceeds a certain value. The result is a dramatic shrinkage of confidence intervals, i.e. the statistical uncertainty of an extrapolated estimate is decreased by introducing additional physical information.

The goals of this study are, in broad terms, to examine the properties of the distribution and its tail of the response (and local maxima) of the piecewise linear oscillator, to interpret them from a physical standpoint whenever possible, and to understand their implications for available extreme value techniques such as the POT approach. A more concrete question of interest, for example, is why Campbell et al. [20, 21] systematically observed positive shape parameters while fitting GPD to the roll peak data. At first glance, a negative shape parameter and the distribution having an upper bound might be expected: observing a roll peak means a ship returning to its stable equilibrium, and a limit is expected beyond which a ship would not return (i.e. she will capsize).

The goals above are achieved by deriving and interpreting the distribution of the response (and local maxima) of the piecewise linear oscillator in two complementary approaches: the first approach for correlated excitation by taking advantage of the piecewise linear form of the stiffness (see Section 3.2 below), and the second approach for white noise excitation based on the Fokker-Planck-Kolmogorov equation (see Section 3.3 below). Both approaches reveal the structure of the distribution tails where a Gaussian core is followed by a heavier tail, possibly power-law heavy, which ultimately turns into a light tail with a finite upper bound. Some implications of the findings on extreme value analysis using GPD are also discussed (Section 3.4 below). The study ends with a summary and conclusions (Section 3.5 below), and an appendix (Appendix A) is included with details on the synthetic processes employed here. The chapter extends earlier study in Belenky et al. [16].

3.2 Probabilistic response for system with piecewise linear stiffness and correlated Gaussian excitation

The focus here is on the distributions of the response of the oscillator (3.1)–(3.2) and its largest values (local maxima). Excitation is assumed to be a stationary ergodic correlated Gaussian process (see Appendix A). The consideration is limited to the values exceeding the

knuckle point $x > x_m$. As the slope of the piecewise linear stiffness is negative for $x \in [x_m, x_v]$, the excitation may be "switched off" for $x > x_m$. This assumption is based on the notion that lightly-damped dynamical systems receive most of their energy for the excitation through the resonance, while the latter is not possible when the slope of stiffness is negative (Belenky and Sevastianov [11]).

3.2.1 Solution in the nonlinear regime and the case of absent excitation

When the excitation is switched off above the knuckle point, the solution of (3.1) above the knuckle point $x > x_m$ is given by

$$x(t) = Ae^{\lambda_1 t} + Be^{\lambda_2 t} + x_v, \tag{3.5}$$

where

$$\lambda_1 = -\delta + \sqrt{k_1 w_0^2 + \delta^2} > 0, \quad \lambda_2 = -\delta - \sqrt{k_1 w_0^2 + \delta^2} < 0 \tag{3.6}$$

are the two eigenvalues associated with the linear oscillator (3.1) in the regime $x > x_m$, and A, B are constants determined by the initial conditions $x_m, \dot{x}_1$ of $x, \dot{x}$ at the upcrossing of x_m by x, through

$$A = \frac{\dot{x}_1 + \lambda_2(x_v - x_m)}{\lambda_1 - \lambda_2}, \quad B = -\frac{\lambda_1(x_v - x_m) + \dot{x}_1}{\lambda_1 - \lambda_2}. \tag{3.7}$$

The absence of capsizing now corresponds to the case $A < 0$ (see also Belenky [8]), and hence $\dot{x}_1 < \dot{x}_{\mathrm{cr}}$ with the critical value of the derivative at the upcrossing given by

$$\dot{x}_{\mathrm{cr}} = -\lambda_2(x_v - x_m). \tag{3.8}$$

The constant B is always negative.

If no capsizing occurs, the solution (3.5) can also be expressed through the hyperbolic cosine as

$$x(t) = He^{-\delta t}\cosh(\omega_1 t + \epsilon) + x_v, \tag{3.9}$$

where $\omega_1 = \sqrt{k_1 w_0^2 + \delta^2}$, and the "magnitude" H and the "phase" ϵ are constants defined as

$$H = -\frac{1}{\omega_1}\sqrt{\omega_1^2(x_m - x_v)^2 - (\dot{x_1} + \delta(x_m - x_v))^2}, \tag{3.10}$$

$$\epsilon = \tanh^{-1}\Big(\frac{\dot{x_1} + \delta(x_m - x_v)}{\omega_1(x_m - x_v)}\Big). \tag{3.11}$$

For stationary Gaussian excitations, the probability density function (pdf) of the derivative $\dot{x}_1$ at the upcrossing is expected to be approximated by a Rayleigh distribution with density

$$\frac{z}{\sigma_{\dot{x}}^2} e^{-z^2/2\sigma_{\dot{x}}^2}, \quad z > 0, \tag{3.12}$$

where $\sigma_{\dot{x}}^2 = \mathbb{E}\dot{x}_l(0)^2$ and $x_l(t)$ is the solution of (3.1) supposing the stiffness function of the linear regime throughout the whole domain. Indeed, recall that the equation (3.12) describes the pdf of a value of the first derivative of a stationary Gaussian process taken at the instant of upcrossing of a given level (for example, Leadbetter et al. [54], p. 201; Lindgren [62], Section 8.4; Sólnes [96], pp. 161–162; Campbell et al. [21], Section 2.4). Supposing $A < 0$, the density of $\dot{x}_1$ can then be thought of as

$$f_{\dot{x}_1}(z) = \frac{z e^{-z^2/2\sigma_{\dot{x}}^2}}{\sigma_{\dot{x}}^2(1 - e^{-\dot{x}_{\mathrm{cr}}^2/2\sigma_{\dot{x}}^2})}, \quad 0 < z < \dot{x}_{\mathrm{cr}}. \tag{3.13}$$

In fact, some caution needs to be exercised in using (3.12) for the purposes here. The pdf (3.12) is that of the derivative $\dot{x}_1$ at an upcrossing of a linear system, but this includes *all* upcrossings. A response of the dynamical system is correlated, and one upcrossing is often followed by another upcrossing, so that they appear in clusters. Consider the first upcrossing in each cluster as a way to sample independent upcrossings. Then, the pdf of these independent upcrossings might, in fact, be different from that in (3.12) for all upcrossings. Comparison between an empirical pdf of the first upcrossings and the pdf of all upcrossings is shown in the left plot of Figure 3.2. (The system parameters are given at the end of this section.) A good agreement is obtained with (3.12) in the case of all upcrossings but *not* in the case of first upcrossings. The value of the derivative at an upcrossing is related to the value of the peak to follow this upcrossing. The higher peak comes with a larger derivative at the preceding

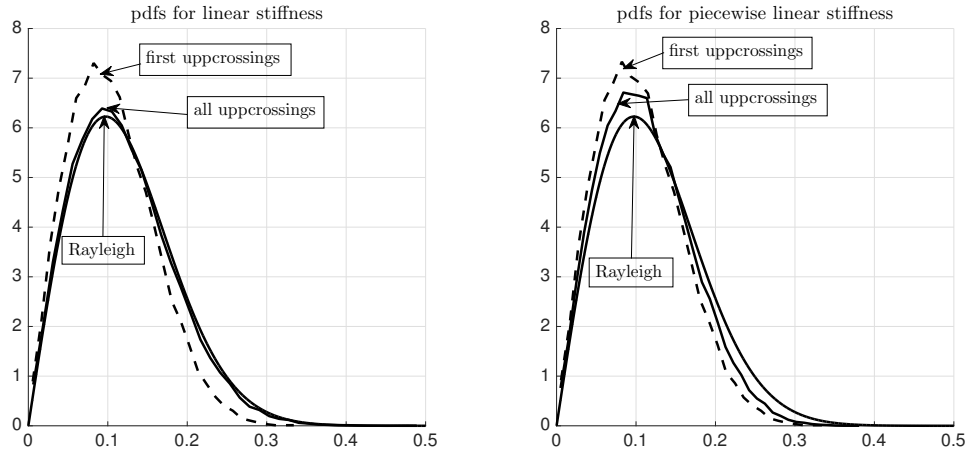

Figure 3.2: Left: The empirical pdfs of $\dot{x}(t)$ from 30,000 hours at the upcrossing of x_m for all upcrossings (dashed curve) and the first upcrossings (solid curve) per 30-minute record. Also, the Rayleigh pdf with parameter $\sigma_{\dot{x}} = 0.13$ rad/s (solid smooth curve) is plotted. Right: The same empirical pdfs of $\dot{x}(t)$ calculated for a piecewise linear system with the same Rayleigh distribution.

upcrossing. The first peak (the derivative at the first upcrossing, respectively) is not usually the largest in a cluster, and the average first peak (the average derivative at the first upcrossing, respectively) in a cluster will not necessarily be equal to the average of all peaks (derivatives, respectively) above the upcrossing threshold. In fact, Figure 3.2, left plot, suggests that the derivatives at the first upcrossings are smaller on average than the derivatives at all upcrossings.

The right plot of Figure 3.2 shows similar quantities for the piecewise linear system. The empirical pdf of the derivatives at all upcrossings is no longer in agreement with the theoretical pdf in (3.12). Each time the knuckle point is crossed, the oscillations (3.5) with natural frequency are generated. If the piecewise linear system is lightly damped for $x \in [-x_m, x_m]$, the oscillations do not die out until the next crossing. As a result, the value of the derivative at the upcrossing is altered compared to the linear system (where the oscillations with natural frequency are present only at the initial transition).

Though no agreement is observed with (3.12) in the right plot of Figure 3.2, the Rayleigh distribution still provides a good fit to the two empirical distributions of that figure. This is illustrated in Figure 3.3. Its left plot shows the two Rayleigh fits (for the Rayleigh parameters $\sigma_{\dot{x}}^2$ in (3.12) chosen through maximum likelihood). The right plot shows the same plot but on the vertical log scale, where a slight disagreement can be seen in the tails of the empirical and fitted distributions for first upcrossings – the empirical pdf seems slightly lighter in the tail

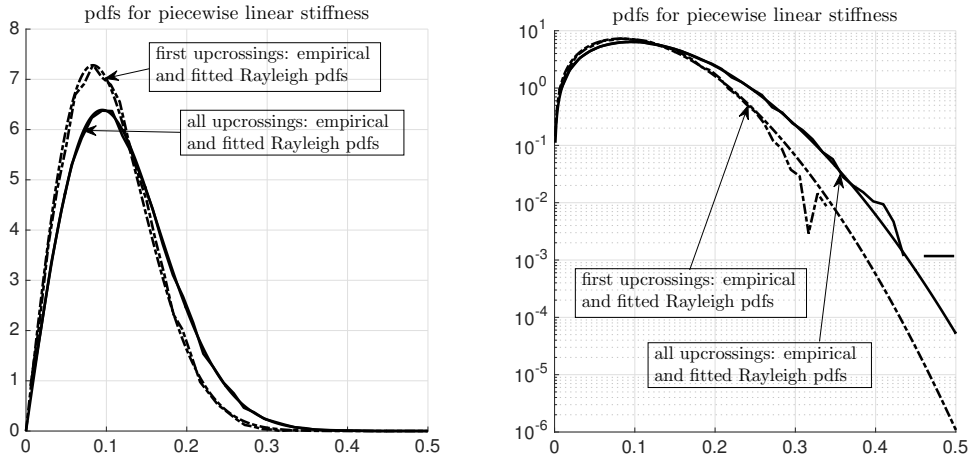

Figure 3.3: Left: The empirical pdfs of $\dot{x}(t)$ and their fitted Rayleigh pdfs. For all upcrossings (solid curves), the fitted Rayleigh parameter is $\widehat{\sigma}_{\dot{x}} = 0.089$ rad/s; for the first upcrossings, it is 0.084 rad/s. Right: The same plot on a vertical log scale.

than the Rayleigh distribution. Since the Rayleigh fits are generally satisfactory in Figure 3.3, the pdf of the derivative at the upcrossing in the case of the piecewise linear system will still be assumed as (3.12), with the understanding that its parameter $\sigma_{\dot{x}}^2$ may need to be adjusted.

The excitation process $y(t)$ and other parameters used in simulations throughout this chapter are defined as follows. The process $y(t)$ is assumed to be a zero mean Gaussian process with the spectral density given and discussed in Appendix A. The plots in Figures 3.2 and 3.3 correspond to the significant wave height $H_s = 9$ m and the mean period $T_1 = 11.595$ s. The other parameter values in the system (3.1)–(3.2) are the damping parameter δ, the natural frequency ω_0, the knuckle point x_m, and the slope parameter k_1. For Figures 3.2 and 3.3, these are $\omega_0 = 0.6$ rad/s, $\delta = 0.15\omega_0$, $x_m = \pi/6 = 0.5236$ rad and $k_1 = 1$ but may change in other simulations below.

3.2.2 The distribution of the maximum value

The solution (3.5) depends on only one random variable: the derivative value at upcrossing $\dot{x}_1$ with its distribution described by (3.12). Thus, the expression (3.5) can be seen as a deterministic function of a random variable. Consider the distribution of the maximum value of (3.5) first, as its derivation is easier than the distribution of all the points in the solution

(3.5). In the absence of capsizing ($A < 0$), the time of maximum of (3.5) can be found by setting the derivative of (3.5) to zero, which results in:

$$t_{\max} = \frac{1}{\lambda_1 - \lambda_2} \log\Big(- \frac{\lambda_2 B}{\lambda_1 A}\Big). \tag{3.14}$$

A zero value of the derivative of (3.5) at $t_{\max}$ allows expressing the constant B as $B = -(\lambda_1/\lambda_2)Ae^{(\lambda_1-\lambda_2)t_{\max}}$, which leads the the following formula for the maximum value:

$$x_{\max} = x(t_{\max}) = x_v - \Big(1 - \frac{\lambda_1}{\lambda_2}\Big)\Big(- \frac{\lambda_2}{\lambda_1}\Big)^{\frac{\lambda_1}{\lambda_1-\lambda_2}} |B|^{\frac{\lambda_1}{\lambda_1-\lambda_2}} |A|^{-\frac{\lambda_2}{\lambda_1-\lambda_2}} =: G(\dot{x}_1), \tag{3.15}$$

where $x_m < G(\dot{x}_1) < x_v$ for $0 < \dot{x}_1 < \dot{x}_{\mathrm{cr}}$. The density of the maximum value (3.15) is then given by

$$f_{x_{\max}}(x) = f_{\dot{x}_1}(G^{-1}(x)) \Big| \frac{d}{dx} G^{-1}(x)\Big|, \quad x_m < x < x_v. \tag{3.16}$$

The function G in (3.15) does not have an inverse expressible in closed form so that the density (3.16) cannot be written in closed form either. The structure of the density can nevertheless be explored in at least two ways: its tail around the unstable equilibrium/endpoint x_v, and its form in the special case of no damping when $\delta = 0$ (above x_m).

3.2.2.1 Behavior of the density around unstable equilibrium

Consider the situation just short of capsizing, i.e. the solution (3.5) when the value of the derivative at the upcrossing is just slightly below the critical value (3.8):

$$\dot{x_1} = \dot{x}_{\mathrm{cr}} - \Delta\dot{x}, \tag{3.17}$$

where $\Delta\dot{x}$ is small. The constants A and B can be expressed in terms of $\Delta\dot{x}$ as

$$A = \frac{\Delta\dot{x}}{\lambda_1 - \lambda_2}, \quad B = \frac{\Delta\dot{x}}{\lambda_1 - \lambda_2} - (x_v - x_m) \approx -(x_v - x_m). \tag{3.18}$$

Substitution of (3.17) and (3.18) into (3.15) yields the approximation of the distribution of the maximum near capsizing, i.e. at the tail: as $\dot{x}_1 \uparrow \dot{x}_{\rm cr}$ (or $\Delta\dot{x} \downarrow 0$),

$$G(\dot{x}_1) \approx x_v - C_0(\dot{x}_{\rm cr} - \dot{x}_1)^{-\frac{\lambda_2}{\lambda_1-\lambda_2}}, \tag{3.19}$$

where

$$C_0 = \Big(1 - \frac{\lambda_1}{\lambda_2}\Big)\Big(\frac{\lambda_2}{\lambda_1}(x_m - x_v)\Big)^{\frac{\lambda_1}{\lambda_1-\lambda_2}}(\lambda_1 - \lambda_2)^{\frac{\lambda_2}{\lambda_1-\lambda_2}}. \tag{3.20}$$

The function (3.19) can be inverted in closed form: as $x \uparrow x_v$,

$$G^{-1}(x) \approx \dot{x}_{\rm cr} - \Big(\frac{x_v - x}{C_0}\Big)^{-\frac{\lambda_1-\lambda_2}{\lambda_2}} \tag{3.21}$$

and hence

$$f_{x_{\max}}(x) \approx \frac{(\lambda_1 - \lambda_2) f_{\dot{x}_1}(x_{\rm cr})}{(-\lambda_2) C_0^{-(\lambda_1-\lambda_2)/\lambda_2}}(x_v - x)^{-\frac{\lambda_1-\lambda_2}{\lambda_2}-1} = C_1(x_v - x)^{-\frac{\lambda_1-\lambda_2}{\lambda_2}-1}. \tag{3.22}$$

Analysis of the tail structure based on the formula (3.22) is further discussed below.

3.2.2.2 Special case of no damping

The oscillator (3.1)–(3.2) is only an approximate qualitative model of large ship rolling. While the roll damping actually increases at large roll angles because sharp edges of deck structures enter water, for the purposes of the present analysis, this can be neglected. The topology of the phase plane is defined by the shape of stiffness. Also, the shape of the distribution near the mean value is mostly defined by the initial shape of stiffness (Belenky and Sevastianov [11]), so that the shape of the tail is also influenced by the stiffness more than by the damping.

In the case of no damping $\delta = 0$, the functions G, G^{-1} and the density $f_{x_{\max}}$ can be expressed in closed form. Use of the hyperbolic form of the solution (3.9) is more convenient here. When $\delta = 0$, $\omega_1 = \sqrt{k_1}\omega_0$ and hence

$$x(t) = H\cosh(\omega_1 t + \epsilon) + x_v, \tag{3.23}$$

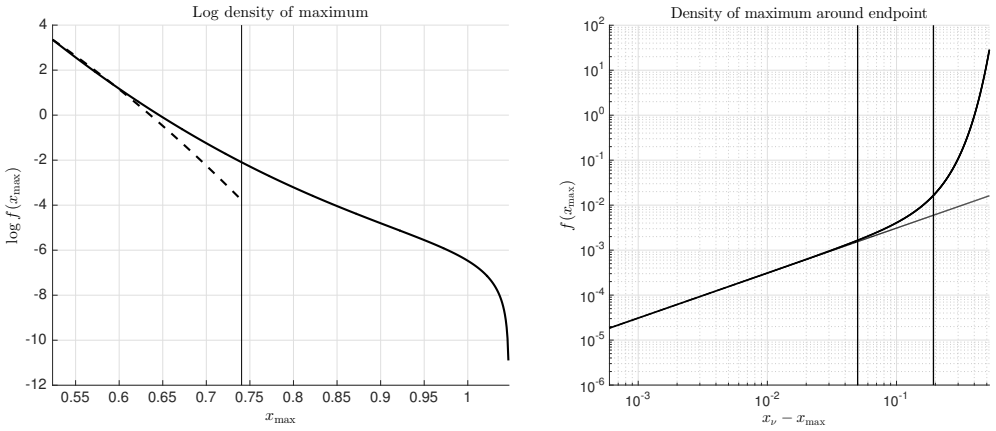

Figure 3.4: Left: The density $f_{x_{\max}}$ of the maximum value on the log vertical scale. Right: The density $f_{x_{\max}}$ around the endpoint x_v on the log-log scale.

where the magnitude and the phase shift are

$$H = -\frac{1}{\omega_1}\sqrt{\omega_1^2(x_m - x_v)^2 - \dot{x_1}^2}, \quad \epsilon = \tanh^{-1}\Big(\frac{\dot{x_1}}{\omega_1(x_m - x_v)}\Big). \tag{3.24}$$

The maximum value is then

$$x_{\max} = x_v + H =: G(\dot{x}_1) \tag{3.25}$$

and hence

$$G^{-1}(x) = \omega_1\sqrt{(x_v - x_m)^2 - (x_v - x)^2}, \quad x_m < x < x_v. \tag{3.26}$$

Substituting (3.26) into (3.16) and using (3.13) leads to the density of the maximum value (3.25) given by

$$f_{x_{\max}}(x) = C(x_v - x)e^{\frac{\omega_1^2}{2\sigma_{\dot{x}}^2}(x_v - x)^2}, \quad x_m < x < x_v, \tag{3.27}$$

where

$$C^{-1} = \frac{\sigma_{\dot{x}}^2}{\omega_1^2}(e^{\frac{\omega_1^2}{2\sigma_{\dot{x}}^2}(x_v - x_m)^2} - 1)$$

is a normalizing constant. Since $(\lambda_1 - \lambda_2)/\lambda_2 = -2$ in the case of no damping, the form (3.27) is consistent with (3.22) as $x \uparrow x_v$.

3.2.2.3 Tail without damping

The left plot, solid line, of Figure 3.4 depicts the density (3.27) on the log vertical scale, for the parameter values $\delta = 0$, $\sigma_{\dot{x}}^2 = 0.0066$, $x_m = \pi/6$, $x_v = \pi/3$, $w_0 = 0.6$ and $k_1 = 1$, which are typical values for ship dynamics. For comparison, the density (in dashed line) of the maximum value for the corresponding linear system is also included, supposing that the derivative at the upcrossing is smaller than $\dot{x}_{\text{cr}}$ in (3.8). (The calculation of the distribution in this case is similar to that above and is omitted.) The density (3.27) has a heavier tail than that of the linear system before collapsing at the endpoint x_v. At the same time, the right upper bound of the distribution is evident around $x_v = \pi/3$.

The right plot of Figure 3.4 depicts the same density (3.27) on the log-log scale around the unstable equilibrium/endpoint x_v. For reference, a straight line is plotted whose slope is 1 as predicted by (3.22). The first vertical line at $x_v - x_{\max} = 5 \times 10^{-2}$ is approximately where the the density is no longer linear in the log-log plot. The percentile corresponding to the value $x_{\max} = x_v - 5 \times 10^{-2}$ is as high as 99.996. For example, the second vertical line corresponds to the 99.9th percentile. These lines are meant to indicate that the power-law behavior around the endpoint can begin very far into the tail of the density.

Figure 3.4 demonstrates the effects that motivated this study: observation of a heavier tail of the peaks distribution of a response of a nonlinear dynamical system, while the presence of an unstable equilibrium (and possibility of capsizing) indicates a right finite upper bound and light tail. Analysis of peaks of the piecewise linear system with similar phase plane topology reproduces this effect. The tail actually has both properties – it is heavy for most of the distance to the unstable equilibrium, but it becomes light in the immediate vicinity of the end point. The reasons why this is occurring and how topology of the phase plane defines the tail structure is addressed in Section 3.3.

3.2.2.4 Transition to flat stiffness

Some aspects of dynamics of piecewise linear system with a flat portion of stiffness (i.e. trapezoid) was examined by Belenky et al. [15] following the discussion by Reed [83].

Consider the influence of a flat portion of the stiffness on the tail structure (the flat portion of stiffness is extended to infinity after the knuckle point). Assume no damping and excitation after the knuckle point. Take the density (3.27) and let k_1 approach 0, while keeping x_m fixed. Note first that

$$x_v = x_m \frac{1+k_1}{k_1} \quad \text{and} \quad x_v - x_m = \frac{x_m}{k_1}. \tag{3.28}$$

Moreover, $\lambda^2 = w_0^2 k_1$. Then, the density (3.27) becomes

$$\frac{\lambda^2(x_v-x)e^{\frac{\lambda^2}{2\sigma_{\dot{x}}^2}(x_v-x)^2}}{\sigma_{\dot{x}}^2(e^{\frac{\lambda^2}{2\sigma_{\dot{x}}^2}(x_v-x_m)^2}-1)} = \frac{\lambda^2(x_v-x)e^{\frac{\lambda^2}{2\sigma_{\dot{x}}^2}[(x_v-x)^2-(x_v-x_m)^2]}}{\sigma_{\dot{x}}^2(1-e^{-\frac{\lambda^2}{2\sigma_{\dot{x}}^2}(x_v-x_m)^2})}$$

$$= \frac{\lambda^2(x_v-x)e^{\frac{\lambda^2}{2\sigma_{\dot{x}}^2}[(x_v-x_m+x_m-x)^2-(x_v-x_m)^2]}}{\sigma_{\dot{x}}^2(1-e^{-\frac{\lambda^2}{2\sigma_{\dot{x}}^2}(x_v-x_m)^2})} = \frac{\lambda^2(x_v-x)e^{\frac{\lambda^2}{2\sigma_{\dot{x}}^2}[2(x_v-x_m)(x_m-x)+(x_m-x)^2]}}{\sigma_{\dot{x}}^2(1-e^{-\frac{\lambda^2}{2\sigma_{\dot{x}}^2}(x_v-x_m)^2})}$$

$$= \frac{w_0^2(x_m(1+k_1)-xk_1))e^{\frac{w_0^2x_m}{\sigma_{\dot{x}}^2}(x_m-x)}e^{\frac{w_0^2k_1}{\sigma_{\dot{x}}^2}(x_m-x)^2}}{\sigma_{\dot{x}}^2(1-e^{-\frac{w_0^2x_m^2}{2\sigma_{\dot{x}}^2k_1}})}$$

$$\to \frac{w_0^2x_m}{\sigma_{\dot{x}}^2}e^{-\frac{w_0^2x_m}{\sigma_{\dot{x}}^2}(x-x_m)}, \quad x > x_m, \quad \text{as } k_1 \downarrow 0, \tag{3.29}$$

that is, the density converges to the exponential pdf with parameter $\frac{w_0^2x_m}{\sigma_{\dot{x}}^2}$. This is illustrated in Figure 3.5. The slope $-k_1$ of the decreasing part of the stiffness function is changed systematically from -1 to 0, as shown in the left plot of the figure. The right plot of the figure shows the corresponding changes in the distribution of the maximum. The heavy part of the tail becomes lighter, until it reaches the exponential distribution (3.29) for $k_1 = 0$. The "inflection point" moves to the right, until it eventually disappears when the position x_v of the unstable equilibrium goes to infinity.

The changes in the slope coefficient k_1 translate into the changes in the shape of stiffness. Thus, the shape of the stiffness function defines the shape of the tail after the knuckle point, while the position x_v of the unstable equilibrium defines the position of the inflection point. The softening nonlinearity ($k_1 > 0$) thus seems to be responsible for the "two-tails" (heavy and light) structure of the tail. It disappears when k_1 becomes zero.

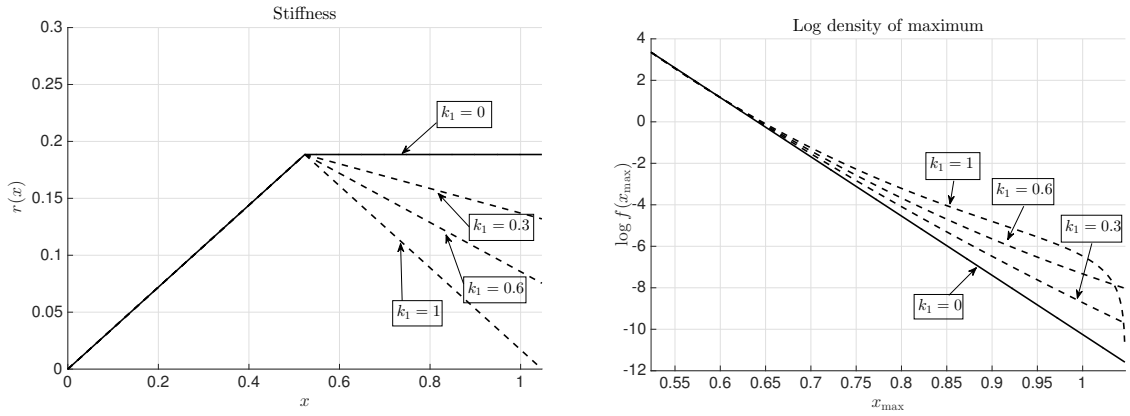

Figure 3.5: Left: The stiffness function for varying choices of k_1. Right: The density $f_{x_{\max}}$ of the maximum value on the log vertical scale for varying choices of k_1.

3.2.3 The distribution of the response

Section 3.2.2 concerned the distribution of the maximum value of the solution (3.5). Here, the focus is on the density of the response (3.5) itself, that is, the distribution of the excursion values of $x(t)$ above x_m over time $0 < t < t_0$, where $x(t_0) = x_m$. This density can be expressed as

$$f_{x(t)}(x) = C_1 \int_{G^{-1}(x)}^{\dot{x}_{\rm cr}} g(x, \dot{x}_1) f_{\dot{x}_1}(\dot{x}_1) d\dot{x}_1, \quad x_m < x < x_v, \tag{3.30}$$

where $\dot{x}_{\rm cr}$ is given by (3.8), $f_{\dot{x}_1}(\dot{x}_1)$ appears in (3.13), G^{-1} is the inverse of the function G in (3.15) and

$$g(x, \dot{x}_1)^{-1} = |x'(t)| \Big|_{t:x(t)=x}, \tag{3.31}$$

emphasizing the dependence on $\dot{x}_1$ in the notation $g(x, \dot{x}_1)$. The role of $g(x, \dot{x}_1)$ is illustrated in Figure 3.6, left plot. The normalizing constant can be expressed as

$$C_1^{-1} = \frac{1}{2} \int_0^{\dot{x}_{\rm cr}} t_0(\dot{x}_1) f_{\dot{x}_1}(\dot{x}_1) d\dot{x}_1, \tag{3.32}$$

where $t_0 = t_0(\dot{x}_1) > 0$ satisfies $x(t_0) = 0$ as above.

3.2.3.1 Special case of no damping

The density (3.30) can be evaluated more explicitly in the case of no damping. In this case, the function $G^{-1}(x)$ is given by (3.26). The function $g(x,\dot{x}_1)^{-1}$ can be expressed as

$$g(x,\dot{x}_1)^{-1} = H\sinh(\lambda t+\epsilon)\Big|_{t=\frac{1}{\lambda}(\cosh^{-1}(\frac{x_v-x}{-H})-\epsilon)} = H\sinh\Big(\cosh^{-1}\Big(\frac{x_v-x}{-H}\Big)\Big)$$

$$= H\sqrt{\Big(\frac{x_v-x}{-H}\Big)^2-1} = \sqrt{\dot{x}_1^2-\lambda^2((x_v-x_m)^2-(x_v-x)^2)} = \lambda^{-1}\sqrt{\dot{x}_1^2-G^{-1}(x)^2},$$

since $-H = 2(AB)^{1/2} = \sqrt{(x_v-x_m)^2-\dot{x}_1^2/\lambda^2}$. Then, letting $G^{-1}(x)=a$, the density (3.30) becomes, after a series of changes of variables,

$$f_{x(t)}(x) = C\int_a^{\dot{x}_{\rm cr}} \frac{ze^{-\frac{z^2}{2\sigma_{\dot{x}}^2}}}{\sqrt{z^2-a^2}}dz = C'\int_{a^2}^{\dot{x}_{\rm cr}^2}\frac{e^{-\frac{w}{2\sigma_{\dot{x}}^2}}}{\sqrt{w-a^2}}dw$$

$$= C'e^{-\frac{a^2}{2\sigma_{\dot{x}}^2}}\int_0^{\dot{x}_{\rm cr}^2-a^2}\frac{e^{-\frac{u}{2\sigma_{\dot{x}}^2}}}{\sqrt{u}}du = C''e^{-\frac{a^2}{2\sigma_{\dot{x}}^2}}\int_0^{\sqrt{\dot{x}_{\rm cr}^2-a^2}} e^{-\frac{v^2}{2\sigma_{\dot{x}}^2}}dv.$$

By recalling that $a^2 = G^{-1}(x)^2 = \lambda^2((x_v-x_m)^2-(x_v-x)^2)$ and using the fact that $\dot{x}_{\rm cr}^2-a^2 = \lambda^2(x_v-x)^2$, the density can be expressed as

$$f_{x(t)}(x) = C_0{\rm erf}\Big(\frac{\lambda(x_v-x)}{\sqrt{2}\sigma_{\dot{x}}}\Big)e^{\frac{\lambda^2}{2\sigma_{\dot{x}}^2}(x_v-x)^2}, \quad x_m<x<x_v, \tag{3.33}$$

where ${\rm erf}(u) = \frac{2}{\sqrt{\pi}}\int_0^u e^{-z^2}dz$ is the error function and C_0 is a normalizing constant.

The difference between the densities of the response and the maximum in (3.33) and (3.27), respectively, is only in the terms ${\rm erf}(\lambda(x_v-x)/\sqrt{2}\sigma_{\dot{x}})$ in (3.33) and x_v-x in (3.27). The two densities are depicted in Figure 3.6, right plot, for the same parameter values as in Figure 3.4.

3.2.4 Power-law tail

Further insight can be gained from Figure 3.7 where the densities of the response and the maximum are depicted in the log-log plot. The densities now appear almost linear, especially that of the maximum density, over a wide range of values. This suggests a power-law behavior of the densities, that is, the behavior $f(x)\approx Cx^{-\alpha-1}$ over a range of values x, with $\alpha>0$.

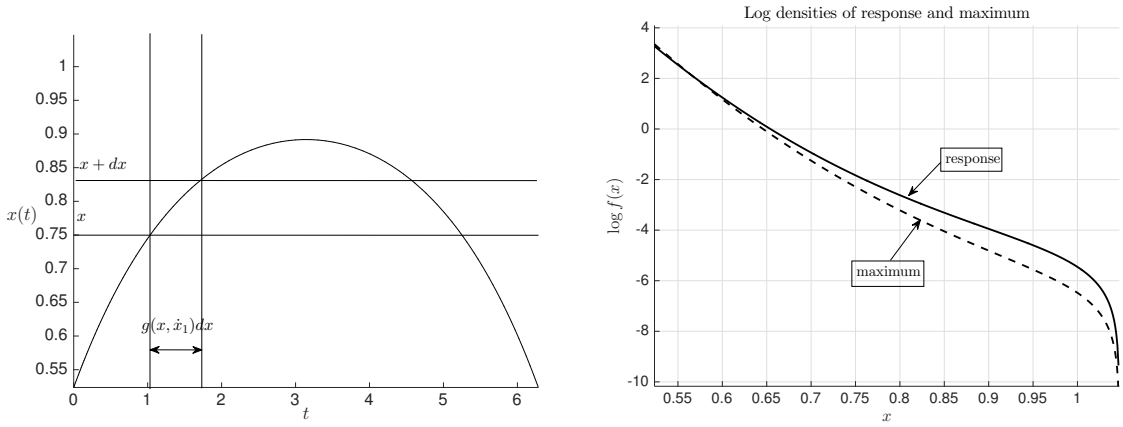

Figure 3.6: Left: Illustration of the role of $g(x, \dot{x}_1)$. Right: The densities $f_{x(t)}$ and $f_{x_{\max}}$ of the response and the maximum value on the log vertical scale.

On the other hand, the power-law behavior is not apparent from the analytic expressions of the densities of the response and the maximum in (3.33) and (3.27). This can nevertheless be explained through the following argument. Consider the case of the density of the maximum in (3.27). This density should appear linear at least around the value $x = x_m$ in the log-log plot (which it does according to Figure 3.7), supposing that a certain condition holds.

Indeed, by considering the log of the density (3.27) around $x = x_m$ and keeping track of the first and second order terms only in the approximations below, observe that

$$\begin{aligned}
\log f_{x_{\max}}(x) &= c_1 + \log(x_v - x) + \frac{\lambda^2}{2\sigma_{\dot{x}}^2}(x_v - x)^2 \\
&= c_2 + \log\Big(1 - \frac{x - x_m}{x_v - x_m}\Big) + \frac{\lambda^2}{2\sigma_{\dot{x}}^2}(x_v - x_m)^2\Big(1 - \frac{x - x_m}{x_v - x_m}\Big)^2 \\
&\approx c_3 - \frac{x - x_m}{x_v - x_m} - \frac{1}{2}\Big(\frac{x - x_m}{x_v - x_m}\Big)^2 - \frac{\lambda^2}{\sigma_{\dot{x}}^2}(x_v - x_m)(x - x_m) + \frac{\lambda^2}{2\sigma_{\dot{x}}^2}(x - x_m)^2 \\
&= c_3 - \Big(\frac{1}{x_v - x_m} + \frac{\lambda^2}{\sigma_{\dot{x}}^2}(x_v - x_m)\Big)x_m\Big(\frac{x}{x_m} - 1\Big) \\
&\quad + \Big(\frac{\lambda^2}{2\sigma_{\dot{x}}^2} - \frac{1}{2(x_v - x_m)^2}\Big)x_m^2\Big(\frac{x}{x_m} - 1\Big)^2 \approx c_3 - (\alpha + 1)\log\frac{x}{x_m},
\end{aligned} \tag{3.34}$$

provided for the last approximation that

$$\alpha + 1 \approx \Big(\frac{1}{x_v - x_m} + \frac{\lambda^2}{\sigma_{\dot{x}}^2}(x_v - x_m)\Big)x_m \approx 2\Big(\frac{\lambda^2}{2\sigma_{\dot{x}}^2} - \frac{1}{2(x_v - x_m)^2}\Big)x_m^2. \tag{3.35}$$

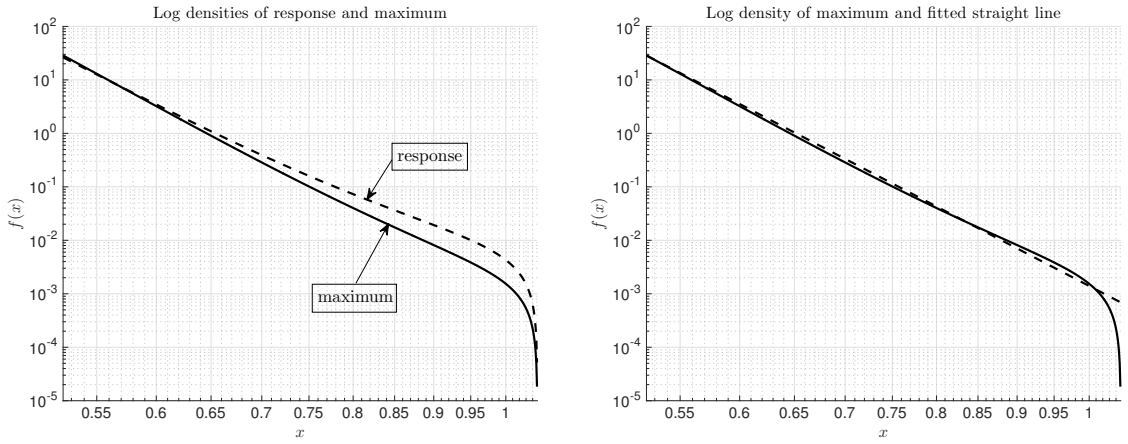

Figure 3.7: The densities $f_{x(t)}$ and $f_{x_{\max}}$ of the response and the maximum value in the log-log plot.

The exponent $\alpha + 1$ then defines the slope of the line in the log-log plot of the density around $x = x_m$. The power-law behavior is analyzed further in Section 3.4.2.

3.3 White noise excitation and interpretation of the tail structure

In this section, the system (3.1) is considered again but now subjected to white noise excitation, $y(t) = s\dot{W}(t)$, where W is a standard Brownian motion and s is the intensity of the white noise excitation. Despite the broadband character of the stochastic forcing, which is not realistic for the description of wave excitation, its mathematical structure allows for the derivation of closed-form expressions for the non-Gaussian pdf of the response. One of the objectives of this chapter is to understand whether the derived results, based on white noise excitation, can be related to the corresponding results when the excitation is correlated. Such a conjecture follows from the observation that in ship motion, inertia plays a dominant role, and therefore high frequency content in the excitation is "filtered out" naturally by the system. That would immediately imply that the primary factor defining the properties of the pdf tail is not the correlation structure of the excitation, but rather the intrinsic dynamics of the system (in this case the restoring force). Such a conclusion will help interpreting the non-Gaussian properties of the tail (for both white and colored noises) and directly relating its form with the phase portrait of the unforced and undamped dynamical system.

For single-degree-of-freedom systems like (3.1), the full (unconditional) pdf can be found for the statistical steady state by direct solution of the associated Fokker-Planck-Kolmogorov (FPK) equation. More specifically, the pdf in the statistical steady state is given by

$$f_\infty(x,\dot{x}) = Ce^{-\frac{4\delta}{s^2}\mathbb{H}(x,\dot{x})}, \tag{3.36}$$

where

$$\mathbb{H}(x,\dot{x}) = \frac{1}{2}\dot{x}^2 + V(x)$$

is the Hamiltonian of the system, and C is a normalizing constant (e.g. Sobczyk [94], Theorem 1.6, p. 34 or p. 334). Based on the unconditional pdf (3.36), the constrained pdf for the response x can now be expressed inside the separatrix, that is, within the two heteroclinical orbits that enclose the stable center and connect the two unstable equilibria in the phase space as in Figure 3.1. Denote the locus of the separatrix points by $(x_s, \dot{x}_s)$. This separatrix is defined implicitly through the Hamiltonian function if the unstable equilibria $\pm x_v$ is known:

$$\mathbb{H}(x_s,\dot{x}_s) = \mathbb{H}(\pm x_v, 0) \equiv \mathbb{H}^*. \tag{3.37}$$

The last equation can be solved explicitly for $\dot{x}_s$:

$$\dot{x}_s(x) = \pm\sqrt{2\left(\mathbb{H}^* - V(x)\right)}. \tag{3.38}$$

In this way, the *conditional pdf* within the separatrix is obtained as

$$f_s\left(x\right) = C\int_{-\dot{x}_s(x)}^{+\dot{x}_s(x)} f_\infty\left(x,\dot{x}\right)d\dot{x}, \quad -x_v < x < x_v, \tag{3.39}$$

where C is another normalizing constant. Although the analysis is valid for a general potential function $V(x)$, for the sake of presentation, the focus is on the special case of the piecewise linear system considered in Section 3.2.

In Figure 3.8 (upper left plot), the potential of the system is presented and compared to the corresponding linear system (i.e. the one characterized by a linear restoring function $r(x) = w_0^2 x$). In the same figure, the pdf of the response (under the condition of non-capsizing,

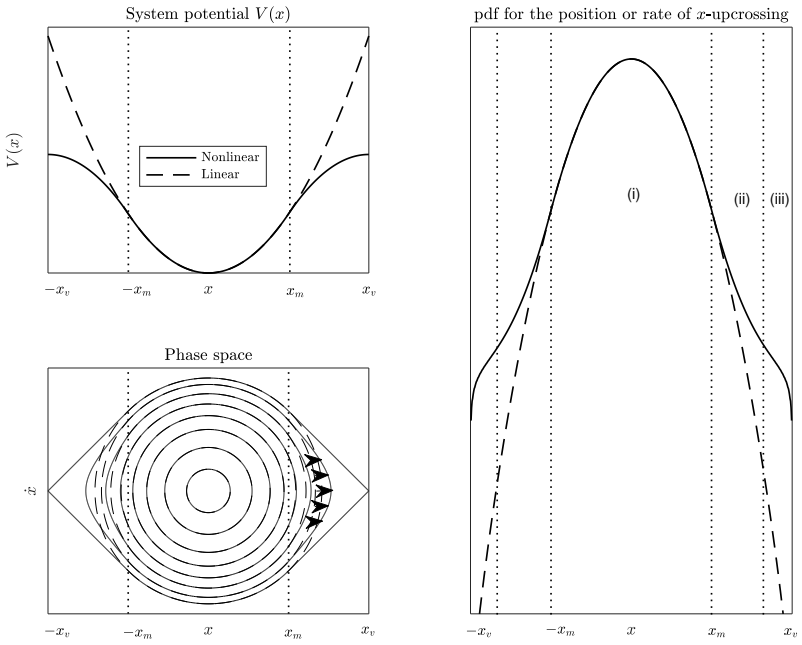

Figure 3.8: Upper left: Potential function for the piecewise linear system (solid curve) and the corresponding linear system (dashed curve); Lower left: Phase portrait for the piecewise linear system and the linear system; Right: Probability density function $f_s(x)$ for both systems considered with the three regions (i), (ii) and (iii). Dotted lines indicate $\pm x_m$ throughout.

i.e. non-crossing of the separatrix) is also presented and compared to the Gaussian pdf that corresponds to the same system but with linear restoring function. As can be observed, the tail of the constrained nonlinear oscillator consists of three different regimes:

(i) a Gaussian core,

(ii) a heavy-tail region, and

(iii) a light-tail region.

A detailed interpretation of this pdf form in connection with the dynamical properties of the system is provided below.

3.3.1 Gaussian core and heavy-tail regime

The phase portraits for the two systems are shown in the lower left plot of Figure 3.8. For $|x| < x_m$ the two systems are identical resulting in (at the statistical level) the presence of a

Gaussian "core" for the nonlinear system, which dominates the response statistics close to the stable equilibrium.

When moving to higher energies, trajectories begin to depart from the linear regime and enter the nonlinear regime, with a deformation of the linear-system phase portrait due to the presence of softening nonlinearity. The softening nonlinearity results in stretching of the trajectories to higher values of x compared with the trajectories of the linear system with equal energy. At the same time, the probability of occurrence of each of those nonlinear trajectories is governed by the Gaussian core, which is fairly identical for the two systems, i.e. trajectories having the same energy have equal probability of occurrence in the two systems. From these two observations, the phase space stretching of the nonlinear system will be reflected in the response pdf through increase of the probability for higher values of x.

Indeed, as observed in the right plot of Figure 3.8 where the pdfs of the two systems are shown, when moving away from the Gaussian core, i.e. for values $|x| > x_m$ the pdf exhibits a heavy tail structure. This heavy tail is a direct manifestation of the deformation (stretching) of the phase space (due to the presence of nonlinearity) which has the following characteristics: it is zero or negligible for small values of $|x|$ so that the Gaussian core of the statistics is not influenced, and it is more pronounced for larger values of $|x|$ leading to larger responses of high energy trajectories compared to the linear system. In this sense, the nonlinearity of the system acts so that it does not change the probability of occurrence of each trajectory (or of each energy level) but only their shape, giving higher probability to larger responses.

3.3.2 Light-tail regime

As discussed in Section 3.3.1, suitable nonlinearity can lead to the formation of heavy tails. However, the heavy tail must turn light at a certain point as more and more trajectories lead to the second stable equilibrium. Thus, for the conditional response pdf, a saturation point exists where the heavy-tail behavior ceases to exist and light tails emerge eventually. This is the transition from the region (ii) to (iii) in the right plot of Figure 3.8. To understand this transition better, the conditional pdf is expressed in terms of the system characteristics. By

the relations (3.36) and (3.39),

$$f_s(x) = C_x e^{-\frac{4\delta}{s^2}V(x)} \int_{-\dot{x}_s(x)}^{+\dot{x}_s(x)} C_{\dot{x}} e^{-\frac{\dot{x}^2}{2\sigma_{\dot{x}}^2}} d\dot{x}, \tag{3.40}$$

where

$$\sigma_{\dot{x}}^2 = \frac{s^2}{4\delta} \tag{3.41}$$

and $C_x, C_{\dot{x}}$ are the normalizing constants for the two corresponding marginals.

The first term in the above product expresses the contribution of the nonlinear restoring force and contains information about the deformation of the phase space. It is the term that results in the heavy tail character of the distribution away from the Gaussian core. The second term is related exclusively to the conditioning that is imposed so that the system response does not cross the separatrix. For system energy where $\sigma_{\dot{x}}$ is small, compared with the vertical separatrix radius, the integral term is approximately equal to 1. This is because the integral extends over the full effective support of the Gaussian distribution for the velocity.

However, when moving closer to the unstable equilibria $\pm x_v$, i.e. the value of x becomes large, the integral term deviates from 1 since the local width of the separatrix becomes comparable with or even smaller than the effective support of the velocity marginal. The integral term becomes much smaller, thereby reducing significantly the pdf value. This is the underlying reason for the eventual formation of light tails when approaching the unstable equilibria.

3.3.2.1 Estimation of the light-tail domain

To quantify a distance from x_v where the influence of the second term in (3.40) becomes important, the local width of the separatrix $\dot{x}_s(x)$ around x_v is compared to the effective support of the $\dot{x}$ marginal, measured through the standard deviation $\sigma_{\dot{x}}$. First, approximate $\dot{x}_s(x)$ around x_v by a Taylor expansion:

$$\dot{x}_s(x) = \dot{x}_s(x^*) + \frac{\partial \dot{x}_s}{\partial x}(x_v)(x - x_v) + O\left((x - x_v)^2\right). \tag{3.42}$$

Clearly, $\dot{x}_s(x_v) = 0$ and now the slope of the separatrix close to the unstable equilibrium has to be estimated. From the relation (3.37) that defines the separatrix,

$$\frac{\partial \mathbb{H}}{\partial x} dx + \frac{\partial \mathbb{H}}{\partial \dot{x}} d\dot{x} = 0,$$

which implies that

$$\frac{\partial \dot{x}_s}{\partial x} = - \lim_{x \uparrow x_v} \frac{\frac{\partial \mathbb{H}}{\partial x}}{\frac{\partial \mathbb{H}}{\partial \dot{x}}}.$$

Substitution of the Hamiltonian and application of L'Hopital's rule yield

$$\frac{\partial \dot{x}_s}{\partial x} = - \lim_{x \uparrow x_v} \frac{\frac{\partial V}{\partial x}}{\dot{x}_s} = - \frac{\frac{\partial^2 V}{\partial^2 x}}{\frac{\partial \dot{x}_s}{\partial x}}.$$

Therefore,

$$\left(\frac{\partial \dot{x}_s(x_v)}{\partial x} \right)^2 = - \frac{\partial^2 V(x_v)}{\partial^2 x}. \tag{3.43}$$

By the expression (3.43), the separatrix in the vicinity of the unstable equilibrium can be approximated as follows. The integral term in the conditional pdf becomes important when the width of the separatrix is comparable (i.e. sufficiently small) with the standard deviation for the velocity $\sigma_{\dot{x}}$. In particular,

$$\dot{x}_s(x) \sim \sigma_{\dot{x}}. \tag{3.44}$$

By the Taylor expansion (3.42) for the separatrix and the relation (3.43), the condition (3.44) takes the form

$$\sqrt{- \frac{\partial^2 V(x_v)}{\partial^2 x}} \Delta x \sim \sigma_{\dot{x}}, \tag{3.45}$$

where $\Delta x = x - x^*$ is the size of the layer over which the light tail is formed. This yields

$$\Delta x \sim \frac{\sigma_{\dot{x}}}{\sqrt{- \frac{\partial^2 V(x_v)}{\partial^2 x}}}. \tag{3.46}$$

The derivation of (3.46) relies on the smoothness of the separatrix as well as the Gaussian distribution of the velocity marginal. To this end, the analysis is valid also for systems of the same form where the excitation is a correlated stochastic process (see Section 3.3.2.2 below). For the special case of white noise considered here, $\sigma_{\dot{x}}$ is given by (3.41) and hence (3.46)

becomes

$$\Delta x \sim \frac{s}{\sqrt{-4\delta \frac{\partial^2 V(x_v)}{\partial^2 x}}}. \tag{3.47}$$

3.3.2.2 Connection to piecewise linear system with correlated excitation

As indicated above, the derivation leading to (3.46) is expected to work for nonlinear systems driven by correlated excitations. Indeed, this is illustrated here on the piecewise linear system (3.1)–(3.3) considered in Section 3.2. For the potential function $V(x)$ of the piecewise linear system given in (3.3), the width of the light-tail regime (3.46) is given by

$$\Delta x \sim \frac{\sigma_{\dot{x}}}{\sqrt{k_1 \omega_0^2}}. \tag{3.48}$$

For the case of no damping and no excitation above the knuckle point, this can be expressed as

$$\Delta x \sim \frac{\sigma_{\dot{x}}}{\lambda} \tag{3.49}$$

On the other hand, consider the pdf of the maximum value given by (3.27). The logarithm of the pdf (3.27) has an inflection point given by

$$x_{\text{infl}} = x_v - \frac{\sigma_{\dot{x}}}{\lambda}. \tag{3.50}$$

Thus, even from the perspective of (3.27) and the location of the inflection point, the light tail width can be thought as $\frac{\sigma_{\dot{x}}}{\lambda}$, which is exactly the same as the right-hand side of (3.49). In the next section, the role of the light-tail region on the POT approach in extreme value analysis is examined.

3.4 Extreme value analysis using generalized Pareto distribution

As the distributions of response and maxima of the piecewise linear oscillator (3.1)–(3.3) are known, a question of interest here is how standard extreme value analysis based on the POT and the GPD (3.4) performs on the data generated by (3.1)–(3.3). Comparing features of the known distribution tail to the results of the POT analysis (Chapter 2 of this dissertation, Campbell et

al. [21]) should lead to a better understanding about the performance of the approach applied to ship motions, as well as to other oscillator-like systems.

The POT approach and the GPD fits are studied below for the random oscillator (3.1)–(3.3) from several angles. In Section 3.4.1, the behavior of the estimates of the shape parameter as a function of threshold for a range of values of the oscillator model is examined. In Section 3.4.2, the power-law behavior of the distribution tail of the oscillator noted in Section 3.2.4 is revisited. In Section 3.4.3, the question is when the point of vanishing stability x_v can be estimated through the POT approach, which relates the findings to the light-tail region of the oscillator discussed in Sections 3.2 and 3.3.

3.4.1 Shape parameter estimates

The definition (3.4) of the GPD and the discussion following it point to the special role played by the shape parameter ξ. In particular, the cases of positive and negative shape parameters are quite different. A negative shape parameter $\xi < 0$ corresponds to the GPD having a finite endpoint at $x = \mu + \sigma/(-\xi)$. The distribution of the maximum or the response of the oscillator (3.1)–(3.3) derived in Section 3.2 naturally has such an endpoint at the point of vanishing stability x_v. Moreover, $\xi > 0$ is associated with the power-law behavior of the GPD pdf with the exponent $-1/\xi - 1$ around the endpoint. In view of (3.22), for the oscillator (3.1)–(3.3),

$$-\frac{1}{\xi} - 1 = -\frac{\lambda_1 - \lambda_2}{\lambda_2} - 1,$$

that is,

$$\xi = \frac{\lambda_2}{\lambda_1 - \lambda_2} \in (-1, 0). \tag{3.51}$$

In the case of no damping above the knuckle point,

$$\xi = -\frac{1}{2}. \tag{3.52}$$

Despite the distribution of the maximum or the response of the oscillator (3.1)–(3.3) always having an endpoint at x_v, a negative shape parameter $\xi < 0$ will not necessarily be estimated from data. As already observed in Section 3.2.4, a power-law tail is expected in some cases.

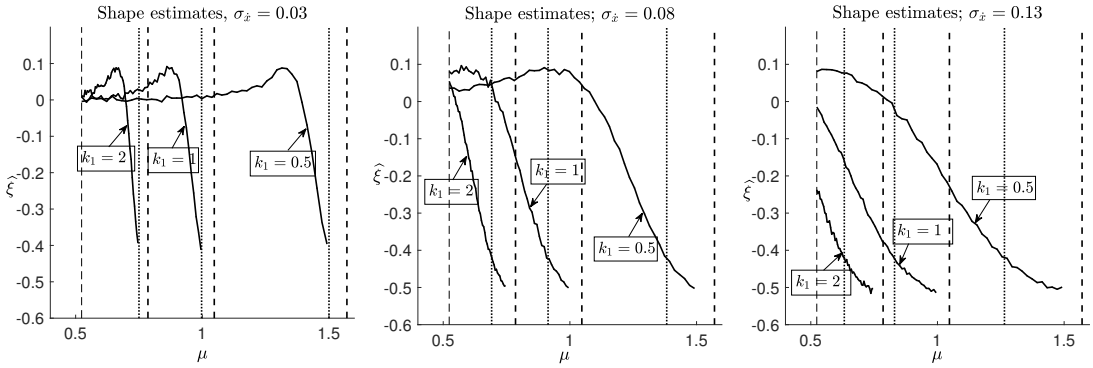

Figure 3.9: Average shape parameter estimates as function of threshold, for $\sigma_{\dot{x}} = 0.03$ rad/s, 0.08 rad/s, 0.13 rad/s and $k_1 = 0.5$, 1, 2.

The power-law tail with the exponent α as in Section 3.2.4 is associated with the positive shape parameter value $\xi = 1/\alpha > 0$ of the GPD. From another perspective, as discussed in Section 3.2.2.2, the region of the power-law behavior around the endpoint can occur too far into the tail from a practical standpoint.

For simplicity, focusing on the pdf (3.27) of the oscillator maximum response in the case of no damping above the knuckle point, characterized by the shape parameter ξ in (3.52), the pdf can be examined through the lens of the GPD as follows. In Figure 3.9, the plots of the average estimated shape parameter values are presented as functions of threshold, as explained in greater detail below. (The estimates are obtained through the maximum likelihood estimation as in Campbell et al. [21] and Chapter 2 of this dissertation.) A varying threshold is considered since in practice it is selected through a data driven method.

The underlying oscillator parameters are the same as those used at the end of Section 3.2.1 except that a range of the parameters $\sigma_{\dot{x}}$ and k_1 is considered: the first plot of Figure 3.9 corresponds to $\sigma_{\dot{x}} = 0.03$ rad/s, the second plot to $\sigma_{\dot{x}} = 0.08$ rad/s, and the third plot to $\sigma_{\dot{x}} = 0.13$ rad/s. In each of these plots, three different values of k_1 are considered: $k_1 = 0.5$, 1 and 2. Thresholds μ are taken above the knuckle point $x_m = \pi/6 = 0.5236$ rad but below $0.95x_v$, where $x_v = x_m(1 + k_1)/k_1$ is the point of vanishing stability in (3.28). For each set of the fixed parameters of the oscillator model and a threshold, 100 datasets of length 400 are generated above the threshold according to the pdf (3.27) and the average value of the shape parameter estimates over these 100 datasets are plotted. The vertical lines in the plots

correspond to the points of vanishing stability x_v (dashed lines) and the inflection points x_{infl} (dotted lines) calculated according to (3.50).

The average value of the shape parameter estimates grows and then falls sharply with the increase of the threshold for $\sigma_{\dot{x}} = 0.03$ rad/s; this tendency is observed for all three values of the slope k_1 when $\sigma_{\dot{x}} = 0.03$ rad/s. The effect of the stretching phase plane in Figure 3.8 is stronger with larger amplitude of the response. Increasing the threshold excludes small peaks and leaves large ones. Thus, the increased influence of the effect of phase plane stretching leads to an increase in the shape parameter estimate until the population of larger peaks starts to drop because of a transition to another stable equilibrium ("capsizing").

Continuing consideration of the case with $\sigma_{\dot{x}} = 0.03$ rad/s, the decrease of the slope coefficient k_1 does not change the shape of the curve, but stretches and shifts it to the right. This is expected as the position of the unstable equilibrium x_v moves to the right with the decrease of the slope coefficient k_1; see Figure 3.5. In the limit of $k_1 = 0$, the distribution of peaks becomes exponential (see (3.29)), from which zero values of the shape parameter estimates are expected for all the thresholds. On the other hand, Figure 3.5 shows a heavier tail when k_1 increases; this makes the shape parameter estimates grow faster for smaller thresholds, which is also seen in the left plot of Figure 3.9.

Consider now the effect of increasing $\sigma_{\dot{x}}$ from 0.03 to 0.08 rad/s on the curve corresponding to $k_1 = 0.5$, shown in the first two plots of Figure 3.9. Increase of the standard deviation of the velocities corresponds to the increase of the excitation, and hence to larger velocities at the crossing of the knuckle point. Naturally, this leads to the growth of the population of larger peaks and is reflected in the faster increase of the shape parameter estimates. Indeed, with the increase of excitation, capsizing is more probable, so its influence is "felt" for smaller thresholds. Regarding the effect of increasing k_1, similarly to above, the curve is shrunk and shifted to the left; however, shrinkage seems to be prevalent. Finally, the case $\sigma_{\dot{x}} = 0.13$ rad/s (the last plot in Figure 3.9) can be seen as a part of this tendency, only the increase has been "shadowed" by the knuckle point. The behavior of the curves corresponding to $k_1 = 1$ and $k_1 = 2$ does not contradict this description.

The inflection point (represented by dotted lines for the various cases) seems to be overestimated by (3.50), as negative parameters are estimated for lower thresholds. This means

that the capsizing influence starts to affect the distribution before the vertical distance to the separatrix decreases to $\sigma_{\dot{x}}$ (see Section 3.3.2.1). The inflection point was searched from the condition of comparability, i.e. with the accuracy of up to a constant. The numerical relationship of the position of the inflection point and $\sigma_{\dot{x}}$ still needs to be studied. For a threshold in the light-tail region of the distribution (that is, larger than the inflection point), the shape parameter estimates are quite close to the value -0.5, which is expected according to (3.52) as discussed above.

In general, the plots of Figure 3.9 illustrate the fact that depending on the underlying oscillator parameters and the likelihood of the transition to another stable equilibrium, the structure of the pdf (3.27) is quite different.

3.4.2 Power-law tail behavior revisited

Recall from Section 3.2.4 that the distribution of the maximum or the response of the oscillator (3.1)–(3.3) can appear to have a power-law tail. In the POT framework, a power-law tail corresponds to the GPD with a positive shape parameter $\xi > 0$, with a larger $\xi > 0$ corresponding to a heavier power-law tail. In particular, Figure 3.9 in Section 3.4.1 shows that depending on the choice of a threshold and the underlying oscillator parameters, a positive shape parameter can be estimated. Several basic questions related to the power-law tail behavior are of interest here. How large $\xi > 0$ can be observed with the oscillator (3.1)–(3.3)? What system parameter values are associated with the largest $\xi > 0$?

To answer these questions, random samples are generated from the pdf (3.27) of the maximum value of the oscillator in the case of no damping above the knuckle point over a range of parameter values $\sigma_{\dot{x}} = 0.03$ rad/s, 0.08 rad/s and 0.13 rad/s, and $0 \leq k_1 \leq 10$ with step-size 0.005. (The rest of the oscillator parameters are the same as at the end of Section 3.2.1.)

Both positive and negative shape parameters can be estimated for many different values of $\sigma_{\dot{x}}$ and k_1 in Figure 3.10. The largest values of $\widehat{\xi}$ in the figure are around 0.1. From the GPD perspective, a distribution tail with $\xi = 0.1$ is not very heavy: such distribution has all its moments finite up to order 10.

Examination of the figure in view of the discussion in Section 3.2.4 is also interesting. The equality of the right two expressions of (3.35) yields a relationship between k_1 and $\sigma_{\dot{x}}$, with the

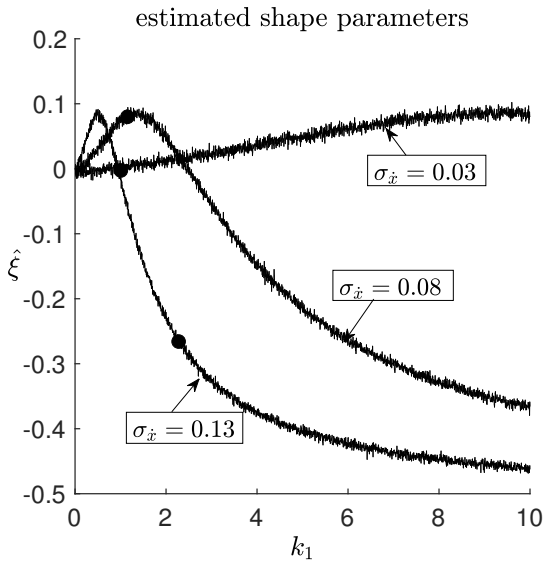

Figure 3.10: Estimated shape parameters from samples of 400 data points from the GPD fitted to $f_{x_{\max}}$ in (3.27), for $\sigma_{\dot{x}} = 0.03$ rad/s, 0.08 rad/s and 0.13 rad/s. Circles correspond to the values of k_1 satisfying the equality of the right two expressions in (3.35).

corresponding value of α in (3.35) and hence $1/\alpha = \xi$. For the three choices of $\sigma_{\dot{x}}$ in Figure 3.10, the corresponding values of k_1 (making the right two expressions of (3.35) equal) are depicted by circles in the figure. Note that the estimated values of $\widehat{\xi}$ at the circles are not necessarily positive (nor necessarily the largest values of $\widehat{\xi}$ observed for a fixed value of $\sigma_{\dot{x}}$). That is, the analysis carried out in Section 3.2.4 is not sufficient to observe a power-law behavior (estimate a positive shape parameter). As argued in the next section, observing a power-law behavior also depends on the scope of the light-tail region. Finally, Figure 3.10 is consistent with Figure 3.9. For example, for $k_1 = 2$, the estimate of ξ in Figure 3.9 with the knuckle point as the threshold is negative for $\sigma_{\dot{x}} = 0.13$ rad/s.

3.4.3 Estimating the point of vanishing stability

As discussed in Section 3.4.1, the distribution of the maximum or the response of the oscillator (3.1)–(3.3) derived in Section 3.2 is associated with the GPD having a negative shape parameter $\xi < 0$ in (3.51) (and $\xi = -0.5$ in (3.52) in the case of no damping above the knuckle point). Moreover, Figure 3.9 of that section shows clearly that a negative shape parameter is not necessarily estimated in practice. The basic question is then: when will a negative shape parameter be in fact estimated from data? As Figure 3.9 already suggests, this might be related

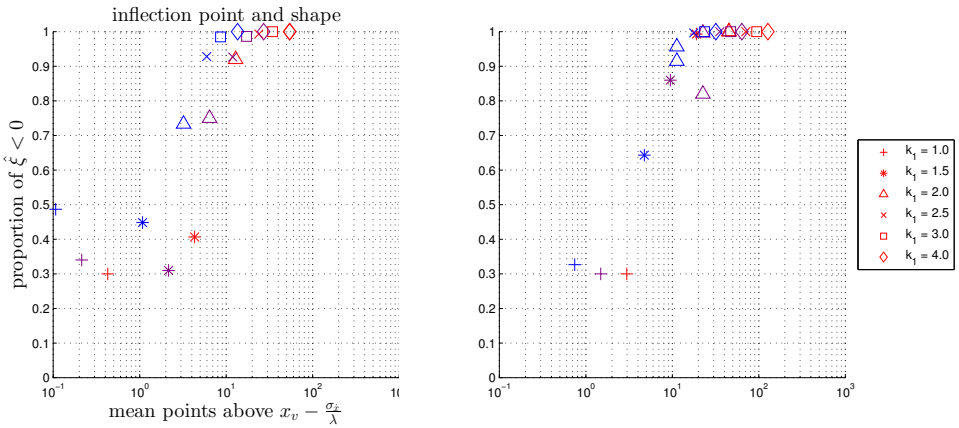

Figure 3.11: Proportion of negatively estimated shape parameters in 150 to 300 records of the oscillator data. Left: $H_s = 9$ m; right: $H_s = 10$ m. Blue: 50 hours; purple: 100 hours; red: 200 hours.

to the light-tail region, which is the region above the inflection point (3.50) discussed in Sections 3.2 and 3.3.

To address these issues, the frequency of the estimation of the negative shape parameters is examined on the data simulated from the oscillator (3.1)–(3.3). For each of the values of the significant wave height $H_s = 6$ m, 9 m and 10 m, and the slope $k_1 = 0.5, 1, 1.5, 2, 2.5, 3$ and 4, 30,000 hours of the data from the oscillator (3.1)–(3.3) are generated. The values of $H_s = 6$ m, 9 m and 10 m correspond to $\sigma_{\dot{x}} = 0.0632$ rad/s, 0.0948 rad/s and 0.105 rad/s, respectively, which are comparable to the values of $\sigma_{\dot{x}}$ considered in Figures 3.9 and 3.10 for the pdf (3.27). (Unlike in Sections 3.4.1 and 3.4.2, the data here were actually generated from the oscillator.) The data were then split into shorter records; either 600 records of 50 hours each, 300 records of 100 hours each, or 150 records of 200 hours each. Within each record, the POT approach was applied to estimate the shape parameter ξ. Threshold values were selected automatically as in Chapter 2 and Campbell et al. [21], and the maximum likelihood estimate for the shape parameter $\widehat{\xi}$ was calculated, as well as the length of the data above the inflection point (3.50). In the latter expression, the theoretical values of $\sigma_{\dot{x}}$ listed above were employed (see Section 3.2.1).

Figure 3.11 summarizes the results by comparing the mean number (across all records of a given set of oscillator parameters) of points above the inflection point to the proportion of records which estimated a negative shape parameter. Notably, enough information in the record to correctly estimate a negative shape parameter is contained at around 10 data points above the

inflection point. For records with very few points above the inflection point, shape parameters are estimated positively and negatively with close to equal proportion, which is consistent with an exponential fit (i.e. with $\xi = 0$). This was the case when $H_s = 6$ m (not shown) and for the lower slopes when $H_s = 9$ m or 10 m. In conclusion, a primary indicator of the ability to estimate a negative shape parameter and, in turn, a finite endpoint of the maximum of the oscillator is the presence of data above the inflection point (that is, in the light-tail region). In particular, parameters which give a pdf that has a shorter region from x_m to the inflection point relative to the region from the inflection point to x_v contribute to estimating a negative shape parameter, due to the greater probability above the inflection point.

3.5 Summary and conclusions

To understand the structure of the tail of a nonlinear dynamical system, the oscillator (3.1)–(3.3) with a piecewise linear restoring force was considered. This oscillator is known to model most principal properties of a dynamical system with softening nonlinearity (Belenky [9]).

The response of (3.1)–(3.3) below or above the knuckle point was expressed in closed form. This allows solving for the response pdf (3.30) above the knuckle point if the excitation above the knuckle point is neglected. The influence of the excitation above the knuckle point is small anyway, as the resonance is not possible there (Belenky and Sevastianov [11]). If damping above the knuckle point is also neglected, the response pdf can be expressed in closed form; see (3.33).

Analogously, a closed form solution was derived for the maxima of the response for the no-damping case; see (3.27). If the damping is present, the maxima pdf can be expressed by (3.16). However, a closed form solution is still available in the vicinity of the unstable equilibrium, described by the relation (3.22).

The other simplification used in this study is neglecting correlation in the excitation, by modeling it as white noise and thus assuming that the response distribution tail structure is mostly the result of nonlinearity of the oscillator rather than internal dependence on the excitation. The pdf of the response becomes available as a solution of the associated FPK equation; see (3.36).

Analysis of the pdf of the response and its maxima above the knuckle point using both approaches has shown that the tail first becomes heavy because of stretching of the phase plane and then becomes light because more and more trajectories lead to capsizing. The following arguments are presented to support this conclusion:

(i) The computed pdf of the response maxima is above the pdf of the corresponding linear system. It is described by a heavy-tail (possibly power-law) region and then a collapse towards the unstable equilibrium (a light-tail region) – see Section 3.2.2 and Figure 3.4, in particular.

(ii) Similar findings apply to the pdf of the response – see Section 3.2.3 and Figure 3.6, in particular.

(iii) Changing the shape of stiffness changes the tail: decrease of the slope after the knuckle point makes the tail lighter, with a limit of exponential tail for a flat restoring term – see Section 3.2.2.4 and Figure 3.5, in particular.

(iv) The distribution of a white noise-excited system shows a heavy tail above the knuckle point due to the phase plane stretching – see Section 3.3.1 and Figure 3.8, in particular.

(v) The probability of a non-capsizing trajectory is dramatically decreased when approaching the unstable equilibrium, where the distance to the separatrix becomes comparable with the conditional standard deviation of the derivative of the response; the latter condition is used to estimate a position of the inflection point, where the tail switches from heavy to light – see Section 3.3.2.

(vi) Monte Carlo simulated samples were used to fit the GPD parameters; the combination of positive and negative values of the shape parameter estimates were found where heavy and light tails were expected – see Section 3.4, Figures 3.9 and 3.10, in particular.

In other words, softening nonlinearity leads to a heavy and light tail combination for the non-capsizing response and its maxima.

What are implications and practical importance of this conclusion?

(i) The question posed at the end of Section 3.1 was answered: observing heavy tails when light tail may be expected while extrapolating roll peak data is natural and related to the fact that ship rolling is described by softening nonlinear system (as a ship usually has a limited range of stability).

(ii) Statistical uncertainty associated with extrapolation of practical volume of ship motion samples is rather large (e.g. Smith and Zuzick [93]). As shown in Chapter 2, introducing a physical information by relating the GPD parameters allows a significant decrease of the confidence intervals. To do this, the structure of the distribution tail should be known.

Where can this work be taken from here?

(i) More study is needed on the inflection point; while the principle is clear, the accuracy of the prediction needs to be improved.

(ii) In Chapter 2, we have related shape and scale parameters for the case of the light tail by postulating the upper bound of the distribution. How can this be done for the heavy tail where an upper bound does not exist?

Another principal issue is related to the modeling of excitation. The spectral-based model (see Appendix A for details), while conventional in Naval Architecture (St. Denis and Pierson [99], Longuet-Higgins [63]), is known to work up to the second moments. The adequacy of this model is not clear for the extreme value problems. One of the alternatives is the autoregressive model (ARM) that has seen wide application in other engineering fields, and while it was considered for modeling waves in the 1980s (Spanos [97]), the interest was renewed recently (Degtyarev and Gankevich [31], Weems et al. [103]). Degtyarev and Reed [32] have shown that ARM reproduces nonlinear properties embedded in its auto-covariance function, so it may work well for extreme events.

CHAPTER 4

Distributions and extreme value analysis of critical response rate and split-time metric in nonlinear random oscillators

4.1 Introduction

The focus of this chapter is on certain questions related to a single-degree-of-freedom nonlinear random oscillator given by

$$\ddot{x}(t) + 2\delta\dot{x}(t) + r(x(t)) = y(t), \tag{4.1}$$

with a softening restoring force (stiffness) $r(x)$ that is characterized by an unstable equilibrium, as e.g. in the special cases of piecewise linear restoring forces depicted in Figure 4.1. In (4.1), $\delta > 0$ is a damping parameter and $y(t)$ is an external random excitation (forcing). Both suitable correlated and white noise excitations will be considered below.

Nonlinear oscillators (4.1), both random and non-random, play a central role in a wide range of areas and applications, as canonical models for oscillatory phenonema. See e.g. Nayfeh and Mook [73], Belenky and Sevastianov [11], or Hayashi [48]. Our primary interest behind (4.1) relates to its use as a prototypical qualitative model of roll motion of a ship in irregular (random) beam seas. The model (4.1) incorporates the possibility of capsizing, understood as the transition through the unstable equilibrium (to another stable equilibrium). See e.g. Belenky [8] for a closed form solution to the capsizing probability of the model (4.1) with a piecewise linear restoring force.

A numerical method to compute (estimate) capsizing probabilities for the model (4.1) and also for more realistic and analytically intractable systems (such as the actual dynamics of a ship) was proposed by Belenky et al. [14] and coined *split-time*. See also Belenky et al. [16] for an up-to-date review of the method. In the split-time approach, calculation of the capsizing probability is reduced to two separate less complex problems: one, the so-called non-

rare problem, involves the rate of the upcrossing of an intermediate level of a roll motion and, second, the so-called rare problem that focuses on capsizing after an upcrossing. In the latter (rare) problem, a roll rate leading to capsizing is computed numerically at each upcrossing, called a *critical response rate*. A split-time metric of the likelihood of capsizing is then formed as the difference between the observed and the critical response rates, with a positive difference corresponding to capsizing. The probability of capsizing after upcrossing can then be found by extrapolating the tail of the distribution of the split-time metric using the generalized Pareto distribution (GPD) as suggested by Extreme Value Theory (see e.g. Coles [23]). The form of the GPD is recalled in (4.11) below. The key advantage of the split-time method is that it lends itself to reproducing a rare phenomenon such as capsizing, which would be too costly to simulate directly with a code of reasonable fidelity. The split-time method was applied to another rare event, the so-called broaching-to in Belenky et al. [17].

One striking feature of the GPD use is that it commonly suggests the distribution of the critical response rate (and the split-time metric) having "light" tails, characterized by a negative or zero shape parameter of the GPD. In contrast, the distribution of the response itself often has a "heavy" right tail, especially in rough seas, characterized by a positive shape parameter of the GPD. The latter phenomenon and, more generally, the tail structure of the distribution of the response itself in the model (4.1) was studied and clarified in Chapter 3, in the case of both correlated and white noise excitations. In particular, a heavier distribution tail was the result of softening nonlinearity. In this chapter, we attempt to provide an analysis similar to that of Chapter 3 but for the distributions of the critical response rate (rather than the response itself) and split-time metric. More specifically, our goals are to: (i) Confirm the light character of the distribution tail of the critical response rate and split-time metric for some forms of the nonlinear oscillator (4.1), for both correlated and white noise excitations, (ii) Provide insight into the structure of the distribution tail of the critical response rate and split-time metric, and (iii) Understand implications of the findings on extreme value analysis of the considered distributions. To the best of our knowledge, this is the first work to take a closer look at the distribution of the critical response rate in a nonlinear oscillator (4.1) and its distribution tails.

Regarding the third point above, our theoretical and empirical analysis suggests that the distribution of the split-time metric is in the domain of maxima attraction of the GPD with zero

shape parameter, that is, the exponential distribution. This motivates the use of the exponential distribution, rather than the GPD, for exceedances above threshold in extreme value analysis. We investigate this approach for the derived distributions of the split-time metric, by using threshold selection methods based on goodness-of-fit tests and the prediction error. In fact, the exponential distribution was used for peaks over threshold in Hydrology before the GPD was adopted as a more flexible model, though threshold selection was made based on other considerations (e.g. Rosbjerg et al. [87], Todorevic and Zelenhasic [102]). Finally, for a more refined examination of a distribution tail in the domain of maxima attraction of the exponential distribution, we suggest the use of the Weibull distribution tail, which was previously considered in the literature for various data fits.

The rest of this chapter is organized as follows. Section 4.2 includes some preliminaries, specifying the exact cases of the model (4.1) considered in this chapter, and also recalling the GPD and its use in Extreme Value Theory and in working with the split-time metric. The considered models are piecewise linear and doubly piecewise linear oscillators with a correlated excitation, and a softening Duffing oscillator with a white noise excitation. The distribution and its tails for the critical response rate (and the split-time metric) in these models are studied in Section 4.3. Section 4.4 contains some numerical results. Section 4.5 concerns the use of the exponential distribution and the distribution with a Weibull tail for peaks over threshold. Section 4.6 concludes.

4.2 Preliminaries

4.2.1 Description of models

We are interested in a single-degree-of-freedom nonlinear oscillator $x(t)$ satisfying (4.1), where $\delta > 0$ is the damping parameter, $y(t)$ is a random, mean zero excitation process and $r(x)$ is the restoring force. More specifically, two forms of random excitation are considered: correlated and white noise. The correlated excitation is assumed to be a mean zero, stationary Gaussian process with the spectral density motivated by the ship rolling application, namely,

$$s_y(\omega) = \omega_0^4 \left(\frac{\omega^2}{g}\right)^2 s_w(\omega), \quad \omega > 0, \tag{4.2}$$

where ω is wave frequency, ω_0 is a natural frequency (both frequencies in rad/s), and $g = 9.807$ rad/s is gravitational acceleration. The Bretschneider spectral density $s_w(\omega)$ is taken for wave elevation:

$$s_w(\omega) = \frac{A}{\omega^5} e^{-\frac{B}{\omega^4}}, \quad \omega > 0, \tag{4.3}$$

where $A = 173 H_s^2 T_1^{-4}$ and $B = 691 T_1^{-4}$ both depend on significant wave height, H_s (in meters), and the period corresponding to mean frequency of waves, T_1 (in seconds). See also Appendix A. The white noise excitation, on the other hand, can be viewed as a (generalized) derivative of the Wiener process, that is, $y(t) = \sigma_f \dot{W}(t)$, and as having a constant spectral density

$$s_y(\omega) = \sigma_f^2, \tag{4.4}$$

where $\sigma_f > 0$ determines the strength of the excitation. Though the broadband nature of the white noise forcing is certainly not realistic for describing wave excitation, its use in the ship rolling application is still relevant, since the ship roll motion is dominated by ship motion inertia. Working with the white noise excitation will allow for more analytic calculations with the model (4.1) that cannot be carried out assuming correlated excitation.

In the case of correlated excitations, two types of the nonlinear restoring force are considered. The *piecewise linear* restoring force is given by

$$r(x) = \begin{cases} -kw_0^2 (x + x_m) - w_0^2 x_m, & \text{if } x < -x_m, \\ w_0^2 x, & \text{if } -x_m \le x \le x_m, \\ -kw_0^2 (x - x_m) + w_0^2 x_m, & \text{if } x > x_m, \end{cases} \tag{4.5}$$

where w_0 is a natural frequency in the linear regime $(-x_m, x_m)$, $-kw_0^2 < 0$ is a negative slope in the nonlinear regime $|x| > x_m$ and x_m, called the "knuckle" point, defines the threshold above which the system behaves nonlinearly, i.e. the point above which the restoring force is decreasing. We shall refer to the oscillator associated with the restoring force (4.5) as piecewise linear (PWL). We also considered the PWL model in Chapter 3.

The doubly piecewise linear restoring force is given by: for $x > 0$,

$$r(x) = \begin{cases} w_0^2 x, & \text{if } 0 < x \le x_m, \\ -kw_0^2 (x - x_m) + w_0^2 x_m, & \text{if } x_m < x \le x_{m,1}, \\ -k_1 w_0^2 (x - x_{m,1}) - kw_0^2 (x_{m,1} - x_m) + w_0^2 x_m, & \text{if } x_{m,1} < x, \end{cases} \tag{4.6}$$

and $r(-x) = r(x)$. It thus makes the decreasing linear part of the piecewise linear restoring force to be piecewise linear itself. We shall refer to the oscillator associated with the restoring force (4.6) as doubly piecewise linear (DPWL).

We denote the respective points of vanishing stability, that is, the points for which the restoring force becomes zero, by x_ν for the PWL oscillator, and by $x_{\nu,1}$ for the DPWL oscillator. The restoring forces and the introduced notation for the PWL and DPWL oscillators are depicted in Figure 4.1.

Though the use of a piecewise linear restoring force may seem unnatural in connection to real-life forces, this turns out to be a useful idealization that allows for analytic arguments in the case of correlated excitation, and also retains most known nonlinear properties of an oscillator with a similar smooth stiffness. See Belenky [9] and, for additional references, Belenky et al. [17]. We consider the DPWL oscillator to assess how our findings for the PWL system are affected by further nonlinearity in the restoring force.

In the case of the white noise excitation (4.4), we shall consider the restoring force

$$r(x) = kx - cx^3, \quad k > 0,\ c > 0, \tag{4.7}$$

associated with a softening Duffing oscillator. Our arguments though are general enough to accommodate other nonlinear restoring forces, but possible at the expense of tractable calculations. An oscillator with a white noise excitation was also considered in Chapter 3.

4.2.2 Split-time metric and its extrapolation through GPD

In the context of ship motions, the split-time metric gives a way to describe how close one came to capsizing, even when capsizing did not occur, conditioned on having crossed an

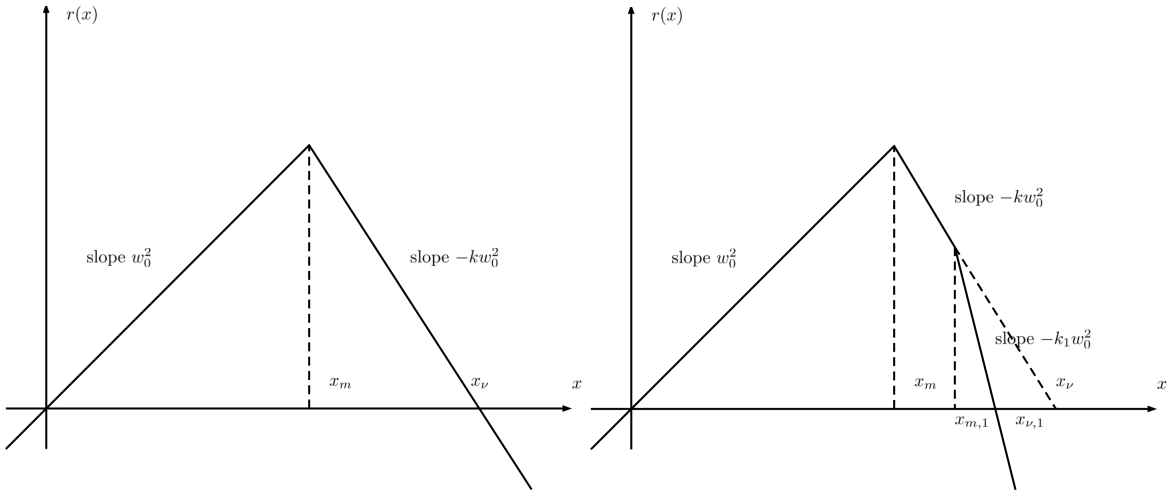

Figure 4.1: The PWL (left plot) and DPWL (right plot) restoring forces.

intermediate threshold (that is commonly taken to be large). It is defined as

$$d_{\text{st}} = 1 + \dot{x}_1 - \dot{x}_{\text{cr}}, \tag{4.8}$$

where $\dot{x}_1$ is an observed response rate at the moment of upcrossing of the threshold, and $\dot{x}_{\text{cr}}$ is the critical response rate which would lead to capsizing. The value of the metric gives a "distance" to capsizing. In particular, $d_{\text{st}} \geq 1$ corresponds to capsizing, and $d_{\text{st}} < 1$ corresponds to not capsizing. The probability of capsizing is then expressed as

$$\begin{aligned} \mathbb{P}(\text{capsizing}) &= \mathbb{P}(\text{crossing threshold})\mathbb{P}(d_{\text{st}} > 1|\text{crossing threshold}) \\ &=: \mathbb{P}(\text{crossing threshold})\mathbb{P}^*(d_{\text{st}} > 1). \end{aligned} \tag{4.9}$$

To estimate $\mathbb{P}$(capsizing), a system of interest (e.g. (4.1)) is simulated until a chosen intermediate threshold is crossed, at which point the simulation is paused and the critical rate is found through a step search, after which the procedure is continued till multiple values of d_{st} are collected and multiple crossings have occurred. The probability $\mathbb{P}$(crossing threshold) can be estimated directly from the crossing frequency. On the other hand, the probability $\mathbb{P}^*(d_{\text{st}} > 1)$ is estimated by applying techniques from Extreme Value Theory, as no observed values of d_{st} would typically be bigger than 1. More specifically, the peaks-over-threshold (POT) approach

(see e.g. Chapter 2 of this dissertation or Coles [23]) is used, by writing first

$$\mathbb{P}^*(d_{\text{st}} > 1) = \mathbb{P}^*(d_{\text{st}} > u)\mathbb{P}^*(d_{\text{st}} > 1 | d_{\text{st}} > u), \tag{4.10}$$

where u is an intermediate threshold of d_{st} (smaller than 1 and for which there are $d_{\text{st}} > u$). The so-called non-rare probability $\mathbb{P}^*(d_{\text{st}} > u)$ is estimated as the observed proportion of d_{st}'s exceeding u. The rare probability $\mathbb{P}^*(d_{\text{st}} > 1 | d_{\text{st}} > u)$ is estimated by fitting a generalized Pareto distribution to d_{st} above u and using it to extrapolate to the capsizing value $d_{\text{st}} = 1$.

It is well known that under general conditions the distribution of exceedances above a sufficiently large threshold from a sequence of i.i.d. random variables is well approximated by a generalized Pareto distribution (GPD; see e.g. Pickands [81] or Coles [23]). Its complementary distribution function is

$$\overline{F}_{\mu,\xi,\sigma}(x) = \begin{cases} \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}, & \mu < x, \qquad \text{if } \xi > 0, \\ e^{-\frac{x-\mu}{\sigma}}, & \mu < x, \qquad \text{if } \xi = 0, \\ \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}, & \mu < x < \mu - \frac{\sigma}{\xi}, \quad \text{if } \xi < 0, \end{cases} \tag{4.11}$$

where the parameters are ξ (shape), and σ (scale), and μ (threshold, equal to u in (4.10)). The sign of the shape parameter determines whether the tail of the GPD is light ($\xi \leq 0$) or heavy ($\xi > 0$). A GPD with a light tail has an upper bound at $\mu + (-\sigma/\xi)$ when $\xi < 0$, while a heavy tail has a power-law form. When $\xi = 0$, the GPD is the exponential distribution.

We also note that in theory, distributions having a power-law tail are approximated (are in the domain of attraction) of the GPD with positive shape parameter $\xi > 0$. Likewise, the domain of attraction of GPD with negative shape parameter $\xi < 0$ are distributions having an upper finite bound and a suitable power-law behavior at that bound. In fact, most other light distributions (normal, Weibull, etc.) are in the domain of attraction of an exponential distribution, that is, the GPD with $\xi = 0$. In practice, the situation is a bit "murkier." For example, estimated shape parameters for a normal distribution would typically be negative, since the normal tail is lighter than the exponential tail. From this more practical perspective, the GPD can be thought as a flexible family of models for distribution tails, whose nature can range from power laws (i.e. very heavy) to bounded from above (i.e. very light). Though we

also emphasize that this view is too simplistic, since the GPD arises naturally in Extreme Value Theory as the distribution to use for exceedances above large thresholds.

4.3 Distributions of response rates

We study here the distributions of the critical response rate and the split-time metric for the models listed in Section 4.2.1: the PWL and DPWL oscillators with correlated excitation (Sections 4.3.1 and 4.3.2 below) and the Duffing oscillator with white noise excitation (Section 4.3.3 below).

4.3.1 PWL oscillator with correlated excitation

We are interested here in a critical response rate, that is, the rate needed to capsize at the upcrossing of the process $x(t)$ of the level x_m. We denote the value of the critical response rate as $\dot{x}_{cr}$. The crossing level is naturally taken as the knuckle point x_m since the system transitions into a nonlinear regime above this point.

For the PWL oscillator, the solution after the upcrossing of the process of the level x_m is given by

$$x(t) = Ae^{\lambda_1 t} + Be^{\lambda_2 t} + x_\nu, \tag{4.12}$$

when the excitation is switched off after the upcrossing, and

$$x(t) = Ae^{\lambda_1 t} + Be^{\lambda_2 t} + x_\nu + p_u(t), \tag{4.13}$$

when the excitation is present. Switching off excitation above the knuckle point is natural since most of the excitation is expected to be received through resonance which is not possible when the stiffness is decreasing (see e.g. Belenky and Sevastianov [11]). Here, A and B are constants determined by the initial conditions $x(0) = x_m$ and $\dot{x}(0) = \dot{x}_1$,

$$\lambda_1 = -\delta + \sqrt{kw_0^2 + \delta^2} > 0, \quad \lambda_2 = -\delta - \sqrt{kw_0^2 + \delta^2} < 0, \tag{4.14}$$

and $p_u(t)$ is a particular solution after the upcrossing, that is, the process satisfying

$$\ddot{p}_u(t) + 2\delta\dot{p}_u(t) - kw_0^2 p_u(t) = y(t) \tag{4.15}$$

and such that at time 0, the process $x(t)$ satisfying the linear equation

$$\ddot{x}(t) + 2\delta\dot{x}(t) + w_0^2 x(t) = y(t) \tag{4.16}$$

upcrosses x_m. For later reference, we also let $p(t)$ be the process satisfying

$$\ddot{p}(t) + 2\delta\dot{p}(t) - kw_0^2 p(t) = y(t), \tag{4.17}$$

without conditioning on the upcrossing of $x(t)$.

With the excitation switched off, we have

$$A = \frac{\dot{x}_1 + \lambda_2(x_\nu - x_m)}{\lambda_1 - \lambda_2}, \quad B = -\frac{\lambda_1(x_\nu - x_m) + \dot{x}_1}{\lambda_1 - \lambda_2} \tag{4.18}$$

and with the excitation on,

$$A = \frac{\dot{x}_1 - \dot{p}_1 + \lambda_2(x_\nu + p_1 - x_m)}{\lambda_1 - \lambda_2}, \quad B = -\frac{\lambda_1(x_\nu + p_1 - x_m) + \dot{x}_1 - \dot{p}_1}{\lambda_1 - \lambda_2}, \tag{4.19}$$

where $p_1 = p_u(0)$ and $\dot{p}_1 = \dot{p}_u(0)$. The system capsizes when $A > 0$, leading to the following critical response rates: with the excitation switched off,

$$\dot{x}_{\text{cr}} = -\lambda_2(x_\nu - x_m) \tag{4.20}$$

and with the excitation on,

$$\dot{x}_{\text{cr}} = -\lambda_2(x_\nu - x_m + p_1) + \dot{p}_1. \tag{4.21}$$

Note that $\dot{x}_{\text{cr}}$ is constant when the excitation is switched off. To understand the distribution of $\dot{x}_{\text{cr}}$ when the excitation is switched on, we need to characterize the joint distribution of p_1 and $\dot{p}_1$. We shall make a simplifying assumption that before the upcrossing of x_m, the process

$x(t)$ behaves as a Gaussian process, satisfying the equation (4.16). This simplifying assumption is often made in the literature (e.g. Mohamad and Sapsis [71]), and has been studied closer in Belenky et al. [17], and in Chapter 3.

By using a standard argument (e.g. Lindgren [62], Section 8.3; Sólnes [96], pp. 161-162), the density of $\dot{x}_1$, p_1 and $\dot{p}_1$ at the upcrossing of x_m is given by

$$C_0 \dot{x} f_{\mu_0,\Sigma_0}(x_m, \dot{x}, p, \dot{p}), \tag{4.22}$$

where C_0 is a normalizing constant, $f_{\mu,\Sigma}$ denotes the multivariate normal density with mean μ and variance Σ, and

$$\mu_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \Sigma_0 = \begin{pmatrix} \sigma_x^2 & \sigma_{x\dot{x}} & \sigma_{xp} & \sigma_{x\dot{p}} \\ \sigma_{\dot{x}x} & \sigma_{\dot{x}}^2 & \sigma_{\dot{x}p} & \sigma_{\dot{x}\dot{p}} \\ \sigma_{px} & \sigma_{p\dot{x}} & \sigma_p^2 & \sigma_{p\dot{p}} \\ \sigma_{\dot{p}x} & \sigma_{\dot{p}\dot{x}} & \sigma_{\dot{p}p} & \sigma_{\dot{p}}^2 \end{pmatrix} \tag{4.23}$$

with the entries in Σ_0 denoting the corresponding covariances of the processes satisfying (4.16) and (4.15). In particular, $\sigma_{x\dot{x}} = 0$, $\sigma_{p\dot{p}} = 0$ and $\sigma_{x\dot{p}} = -\sigma_{\dot{x}p}$. Let $\Sigma_{0,\dot{x}p\dot{p}}$ denote the 3×3 submatrix of Σ_0 associated with the variables $\dot{x}$, p and $\dot{p}$, $\Sigma_{0,\dot{x}p\dot{p}|x}$ denote the 3×1 vector $(\sigma_{x\dot{x}} \ \sigma_{xp} \ \sigma_{x\dot{p}})$ consisting of covariances for $\dot{x}$, p, $\dot{p}$ and x, respectively, $\Sigma_{0,x} = \sigma_x^2$ and $\mu_{0,\dot{x}p\dot{p}}$ denote the 3×1 zero vector of the means associated with $\dot{x}$, p and $\dot{p}$, and $\mu_{0,x} = 0$. Through conditioning on the first variable x (see e.g. Timm [101], Section 3.3a), one can rewrite the density (4.22) as

$$C_1 \dot{x} f_{\mu_1,\Sigma_1}(\dot{x}, p, \dot{p}), \tag{4.24}$$

where

$$\mu_1 = \mu_{0,\dot{x}p\dot{p}} + \Sigma_{0,\dot{x}p\dot{p}|x}\Sigma_{0,x}^{-1}(x_m - \mu_{0,x}) = \begin{pmatrix} 0 \\ \frac{\sigma_{px}}{\sigma_x^2} x_m \\ \frac{\sigma_{\dot{p}x}}{\sigma_x^2} x_m \end{pmatrix} =: \begin{pmatrix} 0 \\ \bar{p} \\ \bar{\dot{p}} \end{pmatrix}, \tag{4.25}$$

$$\Sigma_1 = \Sigma_{0,\dot{x}p\dot{p}} - \Sigma_{0,\dot{x}p\dot{p}|x}\Sigma_{0,x}^{-1}\Sigma'_{0,\dot{x}p\dot{p}|x} = \begin{pmatrix} \sigma_{\dot{x}}^2 & \sigma_{\dot{x}p} & \sigma_{\dot{x}\dot{p}} \\ \sigma_{p\dot{x}} & \sigma_p^2 - \frac{\sigma_{px}^2}{\sigma_x^2} & -\frac{\sigma_{px}\sigma_{\dot{p}x}}{\sigma_x^2} \\ \sigma_{\dot{p}\dot{x}} & -\frac{\sigma_{px}\sigma_{\dot{p}x}}{\sigma_x^2} & \sigma_{\dot{p}}^2 - \frac{\sigma_{\dot{p}x}^2}{\sigma_x^2} \end{pmatrix}. \tag{4.26}$$

By using another conditioning but now on the variable $\dot{x}$ and the notation analogous to above, we can express the density (4.24) as

$$\frac{\dot{x}}{\sigma_{\dot{x}}^2} e^{-\dot{x}^2/(2\sigma_{\dot{x}}^2)} f_{\mu_2(\dot{x}),\Sigma_2}(p,\dot{p}), \tag{4.27}$$

where

$$\mu_2(\dot{x}) = \mu_{1,p\dot{p}} + \Sigma_{1,p\dot{p}|\dot{x}}\Sigma_{1,\dot{x}}^{-1}(\dot{x} - \mu_{1,\dot{x}}) = \begin{pmatrix} \bar{p} + \frac{\sigma_{p\dot{x}}}{\sigma_{\dot{x}}^2}\dot{x} \\ \bar{\dot{p}} + \frac{\sigma_{\dot{p}\dot{x}}}{\sigma_{\dot{x}}^2}\dot{x} \end{pmatrix} = \begin{pmatrix} \frac{\sigma_{px}}{\sigma_x^2}x_m + \frac{\sigma_{p\dot{x}}}{\sigma_{\dot{x}}^2}\dot{x} \\ \frac{\sigma_{\dot{p}x}}{\sigma_x^2}x_m + \frac{\sigma_{\dot{p}\dot{x}}}{\sigma_{\dot{x}}^2}\dot{x} \end{pmatrix}, \tag{4.28}$$

$$\Sigma_2 = \Sigma_{1,p\dot{p}} - \Sigma_{1,p\dot{p}|\dot{x}}\Sigma_{1,\dot{x}}^{-1}\Sigma'_{1,p\dot{p}|\dot{x}} = \begin{pmatrix} \sigma_p^2 - \frac{\sigma_{px}^2}{\sigma_x^2} - \frac{\sigma_{p\dot{x}}^2}{\sigma_{\dot{x}}^2} & -\frac{\sigma_{px}\sigma_{\dot{p}x}}{\sigma_x^2} - \frac{\sigma_{p\dot{x}}\sigma_{\dot{p}\dot{x}}}{\sigma_{\dot{x}}^2} \\ -\frac{\sigma_{px}\sigma_{\dot{p}x}}{\sigma_x^2} - \frac{\sigma_{\dot{p}x}\sigma_{\dot{p}\dot{x}}}{\sigma_{\dot{x}}^2} & \sigma_{\dot{p}}^2 - \frac{\sigma_{\dot{p}x}^2}{\sigma_x^2} - \frac{\sigma_{\dot{p}\dot{x}}^2}{\sigma_{\dot{x}}^2} \end{pmatrix}. \tag{4.29}$$

The expression (4.27) for the density of $\dot{x}_1$, p_1 and $\dot{p}_1$ shows that, as expected, $\dot{x}_1$ follows the Rayleigh distribution with parameter $\sigma_{\dot{x}}^2$, and that conditionally on $\dot{x}_1$, the distribution of p_1 and $\dot{p}_1$ is bivariate normal with mean vector $\mu_2(\dot{x}_1)$ and covariance matrix Σ_2. Note that only the mean depends on $\dot{x}_1$.

These findings suggest that in practice, a sample of independent copies of the vector $(p_1, \dot{p}_1)'$ can be generated easily, and then substituted into (4.21) to get a sample of independent copies of $\dot{x}_{\text{cr}}$, whose distribution can then be examined using available exploratory tools. This is pursued further in Section 4.4.

We note that an explicit expression for the density of p_1 and $\dot{p}_1$ can also be obtained, after integrating out the variable $\dot{x}$ in (4.27). But its form is quite lengthy and will not be presented here. The same could also be said about the distribution of the critical response rate. In fact, as illustrated in Section 4.4 below, the distribution of the latter is close to a normal distribution. This perhaps should not be that surprising since the distribution of p_1 and $\dot{p}_1$ is (conditionally) normal. In particular, the distribution tails of $\dot{x}_{\text{cr}}$ are determined by the normal distribution tails of p_1 and $\dot{p}_1$.

The arguments presented above extend naturally to the split-time metric defined in (4.30), that is,

$$d_{\rm st} = 1 + \dot{x}_1 - \dot{x}_{\rm cr} = \begin{cases} 1 + \dot{x}_1 + \lambda_2(x_\nu - x_m), & \text{with excitation off,} \\ 1 + \dot{x}_1 + \lambda_2(x_\nu - x_m + p_1) - \dot{p}_1, & \text{with excitation on,} \end{cases} \tag{4.30}$$

in view of (4.16) and (4.21). Thus, with excitation off, the distribution of the metric is just a shifted Rayleigh distribution. That is, the distribution of the metric and its tail are completely determined by those of the response rate at the upcrossing. When excitation is on, a sample of independent copies of the metric $d_{\rm st}$ can be generated efficiently in the same way as for $\dot{x}_{\rm cr}$ discussed above and examined through available exploratory tools. The latter is pursued further in Section 4.4 below.

In fact, the asymptotic behavior of the distribution tail of $d_{\rm st}$ with turned on excitation can be derived easily. It follows from (4.27)–(4.29) that the density of $d_{\rm st}$ is

$$f_{d_{\rm st}}(y) = \int_0^\infty d\dot{x} \int_{\mathbb{R}} dp \; \frac{\dot{x}}{\sigma_{\dot{x}}^2} e^{-\dot{x}^2/(2\sigma_{\dot{x}}^2)} f_{\mu_2(\dot{x}),\Sigma_2}(p, \dot{x} + \lambda_2 p - y), \quad y \in \mathbb{R}. \tag{4.31}$$

Indeed, both integrals can be evaluated analytically. First, we rewrite the multivariate normal density as a univariate normal density in p (only) by completing the square, i.e. writing

$$f_{\mu_2(\dot{x}),\Sigma_2}(p, \dot{x} + \lambda_2 p - y) \propto e^{-\frac{1}{2}(Ap^2 - 2Bp + C)} = e^{-\frac{A}{2}(p - \frac{B}{A})^2} \; e^{-\frac{1}{2}\left(C - \frac{B^2}{A}\right)},$$

where $\propto$ denotes "proportional to," and B and C depend on $\dot{x}$ and y, while A is a constant with respect to both. This allows writing the density as

$$\begin{aligned} f_{d_{\rm st}}(y) &\propto \int_0^\infty d\dot{x} \; \dot{x} e^{-\frac{\dot{x}^2}{2\sigma_{\dot{x}}^2}} e^{-\frac{1}{2}\left(C - \frac{B^2}{A}\right)} \int_{\mathbb{R}} f_{\frac{B}{A}, \frac{1}{A}}(p) dp \\ &= \int_0^\infty d\dot{x} \; \dot{x} e^{-\frac{\dot{x}^2}{2\sigma_{\dot{x}}^2} - \frac{1}{2}\left(C - \frac{B^2}{A}\right)}, \quad y \in \mathbb{R}, \end{aligned} \tag{4.32}$$

where f_{μ,σ^2} denotes the univariate normal density as in the multivariate case. It can be seen that the exponent of the exponential function is quadratic in both $\dot{x}$ and y. A second application

of completing the square allows rewriting the exponential term with $\dot{x}$ only, i.e.

$$f_{d_{\text{st}}}(y) \propto e^{-\frac{1}{2}\left(C_0 - \frac{B_0^2}{A_0}\right)} \int_0^\infty d\dot{x}\ \dot{x} e^{-\frac{A_0}{2}\left(\dot{x}+\frac{B_0}{A_0}\right)^2}, \quad y \in \mathbb{R}, \tag{4.33}$$

where A_0 is a constant, B_0 is linear in y, and C_0 is quadratic in y.

Next, we evaluate the remaining integral by a simple change of variables. In particular,

$$\int_0^\infty d\dot{x}\ \dot{x} e^{-\frac{1}{2}A_0\left(\dot{x}+\frac{B_0}{A_0}\right)^2} = \frac{1}{A_0}\left[e^{-\frac{B_0^2}{2A_0}} + B_0\sqrt{\frac{\pi}{2A_0}}\left(\text{erf}\left(\frac{B_0}{\sqrt{2A_0}}\right) - 1\right)\right],$$

where $\text{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-u^2}du$. Hence,

$$\begin{aligned} f_{d_{\text{st}}}(y) &\propto e^{-\frac{1}{2}\left(C_0 - \frac{B_0^2}{A_0}\right)}\left[e^{-\frac{B_0^2}{2A_0}} + B_0\sqrt{\frac{\pi}{2A_0}}\left(\text{erf}\left(\frac{B_0}{\sqrt{2A_0}}\right) - 1\right)\right] \\ &= e^{-\frac{C_0}{2}} + \sqrt{\frac{\pi}{2A_0}}B_0 e^{-\frac{1}{2}\left(C_0 - \frac{B_0^2}{A_0}\right)}\left(\text{erf}\left(\frac{B_0}{2\sqrt{A_0}}\right) - 1\right), \quad y \in \mathbb{R}. \end{aligned}$$

Letting $B_0' := \frac{B_0}{\sqrt{2A}}$, we can rewrite this as

$$f_{d_{\text{st}}}(y) \propto e^{-\frac{C_0}{2}}\left[1 + \sqrt{\pi}B_0' e^{B_0'^2}\left(\text{erf}\left(B_0'\right) - 1\right)\right], \quad y \in \mathbb{R}. \tag{4.34}$$

Now, $\sqrt{\pi}xe^{x^2}(\text{erf}(x) - 1) \to -1$, as $x \to \infty$. Hence, the term in the square brackets in (4.34) behaves like a constant asymptotically. Since C_0 is quadratic in y, the right tail of $f_{d_{\text{st}}}(y)$ behaves like a (non-standard) normal distribution, so it is in the domain of attraction for the GPD with shape parameter $\xi = 0$ (see e.g. De Haan and Ferreira (2007), pp. 11-12).

4.3.2 DPWL oscillator with correlated excitation

We now turn to the DPWL oscillator, and suppose that the excitation is present in the regime $x_m < x < x_{m,1}$ but is switched off in the regime $x > x_{m,1}$. As in (4.20), we know that the critical response rate in the regime $x > x_{m,1}$ is

$$\dot{x}_{\text{cr},1} = -\lambda_{2,1}(x_{\nu,1} - x_{m,1}), \tag{4.35}$$

where as in (4.14),

$$\lambda_{1,1} = -\delta + \sqrt{k_1 w_0^2 + \delta^2} > 0, \quad \lambda_{2,1} = -\delta - \sqrt{k_1 w_0^2 + \delta^2} < 0. \tag{4.36}$$

Since the solution in the regime $x_m < x < x_{m,1}$ is still given by (4.13), the critical response rate for the DPWL oscillator should now satisfy: with $\dot{x} = \dot{x}_{\rm cr}$ and $t = t_{\rm cr}$,

$$\begin{cases} A(\dot{x})e^{\lambda_1 t} + B(\dot{x})e^{\lambda_2 t} + x_\nu + p_u(t) = x_{m,1}, \\ \lambda_1 A(\dot{x})e^{\lambda_1 t} + \lambda_2 B(\dot{x})e^{\lambda_2 t} + pu'(t) = -\lambda_{2,1}(x_{\nu,1} - x_{m,1}), \end{cases} \tag{4.37}$$

where $A = A(\dot{x})$ and $B = B(\dot{x})$ are given in (4.19).

We are interested in solving (4.37) numerically. In order to do so, we need to understand the structure of the process $p_u(t)$. By arguing as for the PWL system in Section 4.3.2, one can show that conditionally on $\dot{x} = \dot{x}_1$, the process $p_u(t)$ is Gaussian with mean

$$\mu_{p_u}(t) = \frac{\gamma_{xp}(t)}{\sigma_x^2} x_m + \frac{\gamma_{\dot{x}p}(t)}{\sigma_{\dot{x}}^2} \dot{x} \tag{4.38}$$

and covariance function

$$\gamma_{p_u}(t_1, t_2) = \gamma_p(t_1 - t_2) - \frac{\gamma_{xp}(t_1)\gamma_{xp}(t_2)}{\sigma_x^2} - \frac{\gamma_{\dot{x}p}(t_1)\gamma_{\dot{x}p}(t_2)}{\sigma_{\dot{x}}^2}, \tag{4.39}$$

where

$$\gamma_p(t) = \mathbb{E}p(0)p(t), \quad \gamma_{xp}(t) = \mathbb{E}x(0)p(t), \quad \gamma_{\dot{x}p}(t) = \mathbb{E}\dot{x}(0)p(t)$$

and $x(t)$ satisfies the linear equation (4.16), and $p(t)$ satisfies the linear equation (4.17). Thus, one can write

$$p_u(t) = \frac{\gamma_{xp}(t)}{\sigma_x^2} x_m + \frac{\gamma_{\dot{x}p}(t)}{\sigma_{\dot{x}}^2} \dot{x} + \kappa(t), \tag{4.40}$$

where $\kappa(t)$ is a Gaussian zero mean process with the same covariance as $p_u(t)$, that is,

$$\mathbb{E}\kappa(t_1)\kappa(t_2) = \gamma_p(t_1 - t_2) - \frac{\gamma_{xp}(t_1)\gamma_{xp}(t_2)}{\sigma_x^2} - \frac{\gamma_{\dot{x}p}(t_1)\gamma_{\dot{x}p}(t_2)}{\sigma_{\dot{x}}^2}, \tag{4.41}$$

This representation is known as the Slepian model for the particular solution $p(t)$ after the upcrossing of $x(t)$ of the level x_m. See e.g. Lindgren [62], Section 8.4. The next elementary lemma clarifies the structure of the process $\kappa(t)$.

Lemma 1. *With the above notation, we have the following representation:*

$$\kappa(t) = p(t) - \frac{\gamma_{xp}(t)}{\sigma_x^2} x(0) - \frac{\gamma_{\dot{x}p}(t)}{\sigma_{\dot{x}}^2} \dot{x}(0). \tag{4.42}$$

Proof. The proof is elementary by checking that the process on the right-hand side of (4.42) has the covariance function (4.39). □ □

By using (4.42), the representation (4.40) of the particular solution $p_u(t)$ can also be expressed as

$$p_u(t) = \frac{\gamma_{xp}(t)}{\sigma_x^2}(x_m - x(0)) + \frac{\gamma_{\dot{x}p}(t)}{\sigma_{\dot{x}}^2}(\dot{x} - \dot{x}(0)) + p(t). \tag{4.43}$$

This representation of $p_u(t)$ is most convenient when generating $p_u(t)$ in practice.

Several approximations of $p_u(t)$ can be tried when substituting (4.43) into (4.37). One approximation is

$$\begin{aligned} p_{u,\mathrm{app},1}(t) &= \frac{\gamma_{xp}(0) + \gamma'_{xp}(0)t}{\sigma_x^2}(x_m - x(0)) + \frac{\gamma_{\dot{x}p}(0) + \gamma'_{\dot{x}p}(0)t}{\sigma_{\dot{x}}^2}(\dot{x} - \dot{x}(0)) + p(0) + p'(0)t \\ &= \frac{\sigma_{xp} + \sigma_{x\dot{p}}t}{\sigma_x^2}(x_m - x(0)) + \frac{\sigma_{\dot{x}p} + \sigma_{\dot{x}\dot{p}}t}{\sigma_{\dot{x}}^2}(\dot{x} - \dot{x}(0)) + p(0) + p'(0)t. \end{aligned} \tag{4.44}$$

As illustrated in Section 4.4, this approximation is accurate till about $t = 1$. Another approximation is

$$p_{u,\mathrm{app},2}(t) = \frac{\gamma_{xp}(t)}{\sigma_x^2}(x_m - x(0)) + \frac{\gamma_{\dot{x}p}(t)}{\sigma_{\dot{x}}^2}(\dot{x} - \dot{x}(0)) + p(0) \tag{4.45}$$

(or without $p(0)$). This approximation does not appear accurate for larger t – keep in mind that the process $p(t)$ is stationary and its magnitude does not seem negligible. See Section 4.4.

In Section 4.4, we will report on the distributions of the critical response rate $\dot{x}_{\mathrm{cr}}$ when solving (4.37) numerically, and also the distribution of the resulting split-time metric. The findings are similar to those for the PWL oscillator.

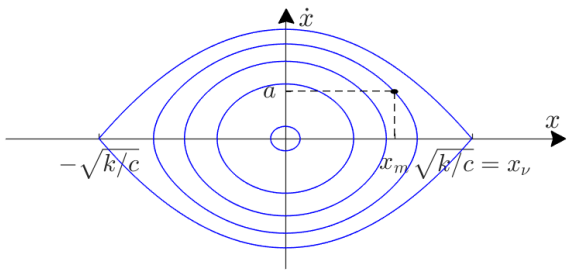

Figure 4.2: A phase portrait of the Duffing oscillator, restricted to the heteroclinical orbit connecting the unstable equilibria $\pm\sqrt{k/c}$.

4.3.3 Duffing oscillator with white noise excitation

We now turn to the Duffing oscillator with the restoring force $r(x)$ in (4.7) and the white noise excitation $y(t) = \sigma_f \dot{W}(t)$ characterized by (4.4). Figure 4.2 depicts a qualitative phase portrait of the (undamped, unforced) system, limited to the heteroclinical orbit connecting the two unstable equilibria $x_\nu = \pm\sqrt{k/c}$. The point of upcrossing $x(0) = x_m$ and the upcrossing response rate $\dot{x}(0) = a$ are also marked in the plot.

We define capsizing as crossing the heteroclinical orbit connecting the two unstable equilibrium without later coming back to it and again are interested in a critical response rate $\dot{x}_{\text{cr}}$. Note that the rate $\dot{x}_{\text{cr}}$ depends only on a future realization of white noise (and the upcrossing threshold x_m). Capsizing, on the other hand, will also depend on the rate at the upcrossing. In our calculations, we shall make the simplifying assumption that capsizing is "monotone" in the rate, that is, if a given rate leads to capsizing, then so would any larger rate (for a given future realization of white noise).

Since the critical response rate is defined as the rate leading to capsizing, note that

$$\mathbb{P}_{x_m}(\dot{x}_{\text{cr}} \leq a) = \mathbb{P}_{x_m}(\text{capsize}|\dot{x}(0) = a), \tag{4.46}$$

where $\mathbb{P}_{x_m}$ denotes a probability with respect to a future realization of white noise, conditioned on $x(0) = x_m$. The white noise realizations in the two events of (4.46) are the same: if a white noise realization is such that $\dot{x}_{\text{cr}} \leq a$, then by monotonicity, one should be capsizing with any larger velocity, including $\dot{x}(0) = a$; vice versa, if one is capsizing for a white noise realization with $\dot{x}(0) = a$, then by monotonicity, this also means that $\dot{x}_{\text{cr}} \leq a$ for such a realization.

In the rest of this section, we focus on computing the CDF of the critical response rate $\dot{x}_{\rm cr}$ using the right-hand side of (4.46). Strictly speaking, our calculations will be approximate and expected to be accurate only in certain regions. More specifically, to evaluate the capsizing probability of interest, we consider only short time intervals, which allows one to remove damping for its insignificant role and to assume a small variance of the excitation. A short time interval can be assumed if one is close to the heteroclinical orbit or if an initial displacement $x(0) = x_m$ is large. Two cases must be handled separately, according to whether the initial conditions a, x_m are within or beyond the heteroclinical orbit.

We first consider the case of starting within the orbit. Under the assumptions above, we shall find the mean and the variance of the maximum response which will lead to the probability of interest by conditioning on the maximum response being above the unstable equilibrium $x_\nu = \sqrt{k/c}$. Figure 4.3, left plot, illustrates our approach.

Under the assumptions above, the equation for the mean $\overline{x}(t)$ of the process is

$$\ddot{\overline{x}}(t) + k\overline{x}(t) - c\overline{x}(t)^3 = 0; \quad \dot{\overline{x}}(0) = a, \quad \overline{x}(0) = x_m. \tag{4.47}$$

By conservation of energy, the mean $\overline{x}_{\max}$ of the maximum response conditional on $\dot{\overline{x}}(t) = 0$ satisfies

$$k\overline{x}_{\max}^2 - \frac{1}{2}c\overline{x}_{\max}^4 = a^2 + kx_m^2 - \frac{1}{2}cx_m^4 =: 2E(a, x_m). \tag{4.48}$$

Solving this equation gives

$$\overline{x}_{\max} = \sqrt{\frac{-k + \sqrt{k^2 - 2(a^2 + kx_m^2 - \frac{1}{2}cx_m^4)c}}{-c}}, \tag{4.49}$$

which depends in a direct way on the initial conditions a, x_m. On the heteroclinical orbit, $E(a, x_m) = \frac{k^2}{2c}$, so (4.49) simplifies to $\overline{x}_{\max} = x_\nu$.

Similarly, we can calculate the variance. The period of the nonlinear oscillator can be computed as follows. Note that, over a quarter of a period $T_0(a, x_m)$, we should have $x(t)$ go

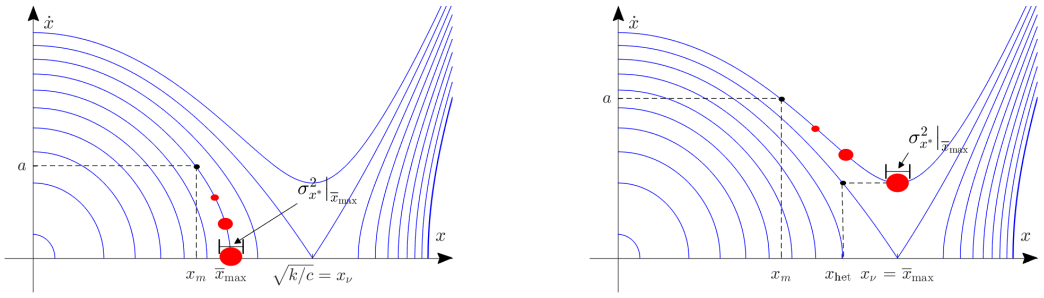

Figure 4.3: The path of $\overline{x}(t)$ to $\overline{x}_{\max}(a, x_m)$ and its variability characterized by $\sigma^2_{x^*}|_{\overline{x}_{\max}}$, for choices of a within (left) and beyond (right) the heteroclinical orbit connecting the two unstable equilibria.

from 0 to $\overline{x}_{\max}$. Then,

$$\frac{T_0(a, x_m)}{4} = \int_0^{\overline{x}_{\max}} \frac{dx}{\sqrt{2E(a, x_m) - kx^2 + \frac{1}{2}cx^4}}. \tag{4.50}$$

For each choice of a, x_m, to make calculations tractable, substitute the nonlinear oscillator with a linear one having the same natural period of oscillation:

$$\ddot{x}^*(t) + \omega_0^2(a, x_m)x^*(t) = \sigma_f \dot{W}(t), \quad \omega_0(a, x_m) = \frac{2\pi}{T_0(a, x_m)}, \tag{4.51}$$

where $x^*(t)$ is the fluctuation around the mean $\overline{x}(t)$ which is due to white noise. For this system, we have

$$\sigma^2_{x^*}(t) := \mathbb{E}x^*(t)^2 = \frac{\sigma_f t}{\omega_0^2(a, x_m)} - \frac{\sigma_f(\sin 2\omega_0(a, x_m)t)}{2\omega_0^3(a, x_m)} \tag{4.52}$$

(see e.g. Gitterman [42], Section 8.2; p. 85). The time it takes for the system to go from $(x(0), \dot{x}(0)) = (a, x_m)$ to $(\overline{x}_{\max}, 0)$ is

$$T_{\max}(a, x_m) = \int_{x_m}^{\overline{x}_{\max}} \frac{dx}{\sqrt{2E(a, x_m) - kx^2 + \frac{1}{2}cx^4}}. \tag{4.53}$$

This time has to be relatively small for our analysis to be valid. Thus,

$$\sigma^2_{x^*}|_{\overline{x}_{\max}} = \frac{\sigma_f T_{\max}(a, x_m)}{\omega_0^2(a, x_m)} - \frac{\sigma_f(\sin 2\omega_0(a, x_m)T_{\max}(a, x_m))}{2\omega_0^3(a, x_m)}. \tag{4.54}$$

In view of the developments above, we conclude that, under the heteroclinical orbit,

$$\mathbb{P}_{x_m}(\text{capsize}|\dot{x}(0)=a)=\int_{x_\nu}^{\infty}\varphi(u;\overline{x}_{\max},\sigma^2_{x^*}\big|_{\overline{x}_{\max}})du, \tag{4.55}$$

where $\varphi(u;\mu,\sigma^2)$ is the density of a normal distribution with mean μ and variance σ^2. Note that both $\overline{x}_{\max}$ and $\sigma^2_{x^*}\big|_{\overline{x}_{\max}}$ depend on a and x_m. After a change of variables, and by (4.46), we can write the CDF of the critical rate (under the heteroclinical orbit) as

$$F_{\dot{x}_{\text{cr}}}(a)=\overline{\Phi}((x_\nu-\overline{x}_{\max})/\sigma_{x^*}\big|_{\overline{x}_{\max}}), \tag{4.56}$$

where $\overline{\Phi}(z)=\mathbb{P}(N(0,1)>z)$ is the tail of the CDF of the standard normal distribution.

We now turn to the case of initial conditions being above the heteroclinical orbit. This approach is illustrated in Figure 4.3, right plot. Here, capsizing occurs for trajectories which remain outside the heteroclinical orbit. Since $\overline{x}(t)$ has no maximum in this case, we consider its variability when $\dot{x}(t)$ is minimized. It is straightforward to see from (4.48) that

$$\dot{x}(t)=\sqrt{2E(a,x_m)-kx(t)^2+\frac{1}{2}cx(t)^4}, \tag{4.57}$$

which is minimized when $x(t)=x_\nu$, hence

$$\dot{x}_{\min}=\sqrt{2E(a,x_m)-\frac{1}{2}\frac{k^2}{c}}. \tag{4.58}$$

We assume that the time it takes for $x(t)$ to go from 0 to x_ν is a quarter of a period, i.e. (4.50) and (4.53) still apply, with $\overline{x}_{\max}$ replaced by $\dot{x}_{\min}$. Thus, the variability, characterized by $\sigma^2_{x^*}\big|_{\dot{x}_{\min}}$, is computed as in (4.54). The probability of capsizing additionally depends on the (horizontal) distance from $(x(t),\dot{x}(t))=(x_\nu,\dot{x}_{\min})$ to $(x_{\text{het}},\dot{x}_{\min})$, where $x_{\text{het}}=x_{\text{het}}(a,x_m)$ is on the heteroclinical orbit. From (4.57), we find

$$x_{\text{het}}=\sqrt{\frac{k-\sqrt{2c\dot{x}^2_{\min}-k^2}}{c}}, \tag{4.59}$$

and

$$F_{\dot{x}_{\rm cr}}(a) = \overline{\Phi}((x_{\rm het} - x_\nu)/\sigma_{x^*}\big|_{\overline{x}_{\max}}). \tag{4.60}$$

Calculations involved in differentiating to compute the density of the critical rate have proven unwieldy. Simulation for the distributions (4.56) and (4.60) are discussed in Section 4.4 below.

An argument can be extended to generate observations of $\dot{x}$, $\dot{x}_{\rm cr}$ and hence those of the split-time metric. Since $\dot{x}_1$ depends on a past realization of white noise and $\dot{x}_{\rm cr}$ depends on the future, the joint density of the rate and critical rate is given by

$$f_{x_m}(\dot{x}_1 = b, \dot{x}_{\rm cr} = a) = f_{x_m}(\dot{x}_1 = b) f_{x_m}(\dot{x}_{\rm cr} = a), \tag{4.61}$$

where the density of $\dot{x}(0)$ is known explicitly from the Fokker-Planck-Kolmogorov equation, yielding

$$f_{x_m}(\dot{x}(0) = a) = Ce^{-\frac{2\delta}{\sigma_f^2}a^2}, \tag{4.62}$$

where C is a normalizing constant. See e.g. Chapter 3, (3.36), or Sobczyk [94], Theorem 1.6, p. 34 or p. 334. It follows that the density of the split-time metric $d_{\rm st}$ is

$$f_{x_m}(d_{\rm st} = z) = f_{x_m}(\dot{x}_1 - \dot{x}_{\rm cr} = z - 1) = \int_0^\infty f_{x_m}(\dot{x}_1 = u + z - 1) f_{x_m}(\dot{x}_{\rm cr} = u) du.$$

By equations (4.56) and (4.3.3), and using integration by parts, we can write

$$f_{x_m}(d_{\rm st} = z) = \int_0^\infty Ce^{-\frac{2\delta}{\sigma_f^2}(u+z-1)^2} dF_{\dot{x}_{\rm cr}}(u) = C_2 \int_0^\infty F_{\dot{x}_{\rm cr}}(u)(u+z-1)e^{-\frac{2\delta}{\sigma_f^2}(u+z-1)^2} du, \tag{4.63}$$

where C_2 is a normalizing constant.

4.4 Numerical results

We shall examine here the distributions of the critical response rate and the split-time metric for the PWL and DPWL systems and the Duffing oscillator through some available exploratory tools.

Parameter	PWL simulation	DPWL simulation
w_0	0.6	0.6
δ	$0.15w_0 = 0.09$	0.09
x_m	$30\pi/180 = 0.52$	0.52
$x_{m,1}$	-	$x_m + \frac{1}{2}(x_\nu - x_m)$
k	0.5	0.5
k_1	-	1.5
H_s	9	9
T_m	15	15

Table 4.1: Parameter values for simulations

4.4.1 PWL and DPWL oscillators

We generated $500,000$ independent copies of the critical response rates for the PWL oscillator, and $5,000$ for the DPWL oscillator, from the distributions and through the methods found in Sections 4.3.1 and 4.3.2. The parameter values for either set of the simulations are given in the middle columns of Table 4.1, and are identical with the exception of the linear segments of the restoring forces. See Figure 4.1.

Figure 4.4 depicts several exploratory plots for the critical response rate generated from the PWL oscillator. Clockwise from upper left these are: the histogram and fitted normal density, the standard normal quantile plot, and the shape parameter plots for the left and right distribution tails. The histogram appears normal. The quantile plot shows little deviation from normal quantiles, except perhaps with slightly lighter tails. The shape parameter plots for both the left and right tails estimate a shape parameter of about $\xi = -0.1$, although $\xi = 0$ is not outside the confidence bounds, especially for the left tail, which is consistent with the estimated shape parameter from the normal distribution. Thus, all four plots suggest agreement that the distribution of the critical response rate is close to normal.

Similar exploratory plots for the split-time metric generated from the PWL oscillator are depicted in Figure 4.5. These include, clockwise from upper left, a histogram of the split-time metric with a density estimated from kernel smoothing, the shape parameter plot for the right tail, the standard normal quantile plot, and a histogram of values above the median with a fitted half-normal density. The left tail is not considered as it corresponds to values of the metric far from capsizing. A small deviation from the normal distribution is seen in the quantile plot and histogram plots, toward slightly heavier tails. Similar to that of the critical response rate in

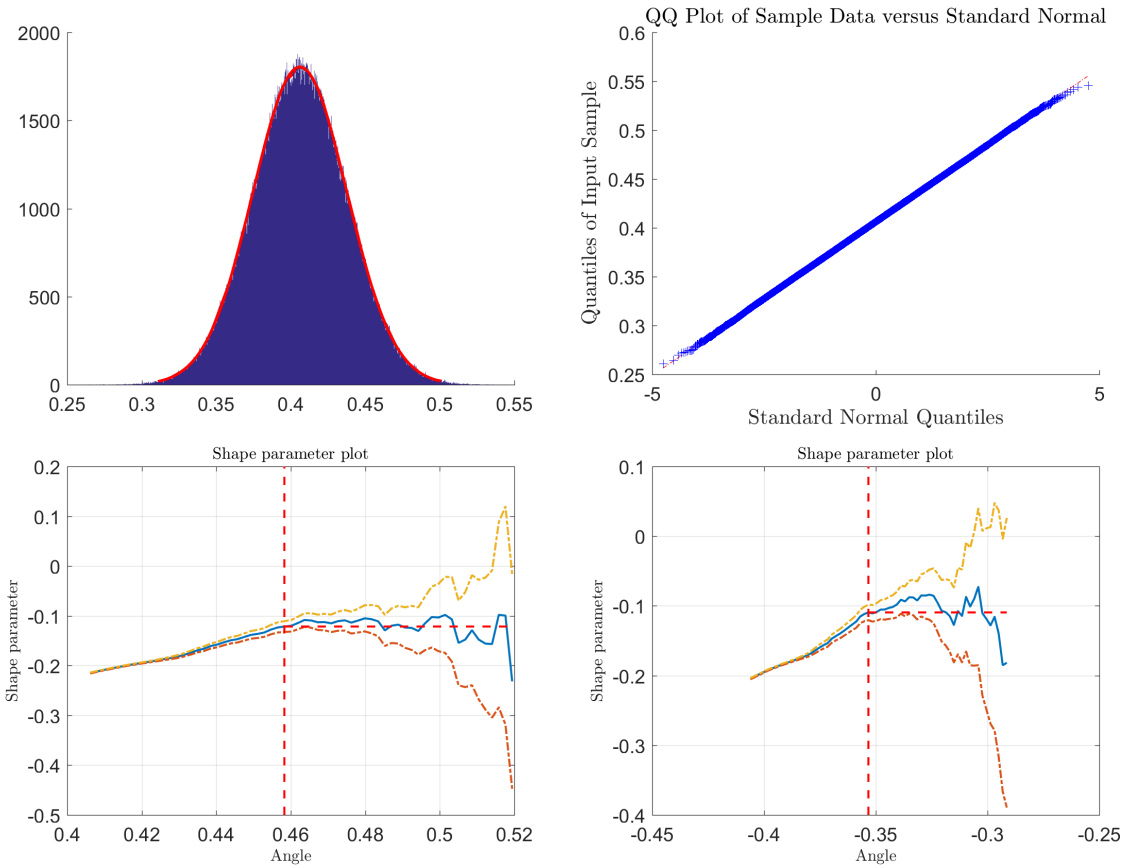

Figure 4.4: The PWL oscillator: Top left: histogram of the critical response rates with a fitted normal; Top right: the QQ-plot of the critical response rates; Bottom left: the shape parameter plot for the right tail; Bottom right: the shape parameter plot for the left tail.

Figure 4.4, the shape parameter plot estimates a shape parameter near $\xi = -0.1$, but $\xi = 0$ remains in the confidence bounds as well, suggesting the distribution of the split-time metric is also close to normal.

Considering the DPWL oscillator, four realizations of the particular solutions $p_u(t)$ and its two approximations discussed in Section 4.3.2 are depicted in Figure 4.6. These illustrate that the first approximation (linearizing p, in green) is reasonable, typically until approximately $t = 1$. The second approximation (linearizing r, in blue) is not quite as good, although it is still often accurate until around $t = 1$.

By using the parameter values above, we generated 5000 independent copies of the critical response rates for the DPWL oscillator. Figure 4.7 depicts the histogram, the normal quantile plot, and the shape parameter plots for the left and right tails, as in Figure 4.4. The results of the exploratory analysis are not too different from those for the PWL oscillator. In particular,

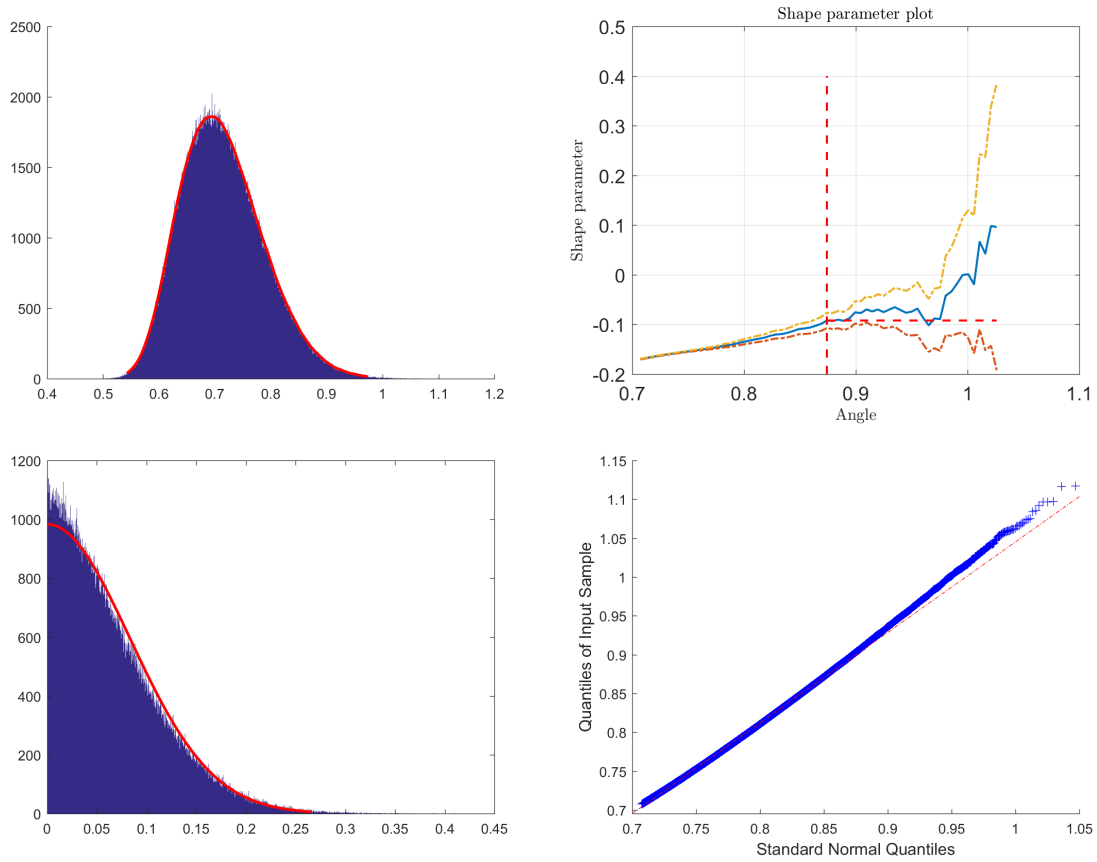

Figure 4.5: The split-time metric generated from the PWL oscillator: Top left: histogram of the split-time metric with a density estimated from kernel smoothing; Top right: the shape parameter plot for the right tail; Bottom left: histogram of values above the median, with fitted half-normal density; Bottom right: the QQ-plot of the metric with the right-tail of a standard normal distribution.

the distribution of the critical response rate is close to normal: the histogram appears normal, and the quantile plot may indicate a lighter tail on the left, and a heavier tail on the right, as compared to the normal distribution. However, the shape parameter plots for both the left and right tails estimate a shape parameter of about $\xi = -0.2$, and although $\xi = 0$ is not outside the confidence bounds, especially for the left tail, which is consistent with the estimated shape parameter from the normal distribution. Thus, all four plots suggest that the distribution of the critical response rate is close to normal.

4.4.2 Duffing oscillator

We compare the above oscillators to the Duffing oscillator with white noise excitation discussed in Section 4.3.3. We generated 10,000 independent copies of the critical response rate

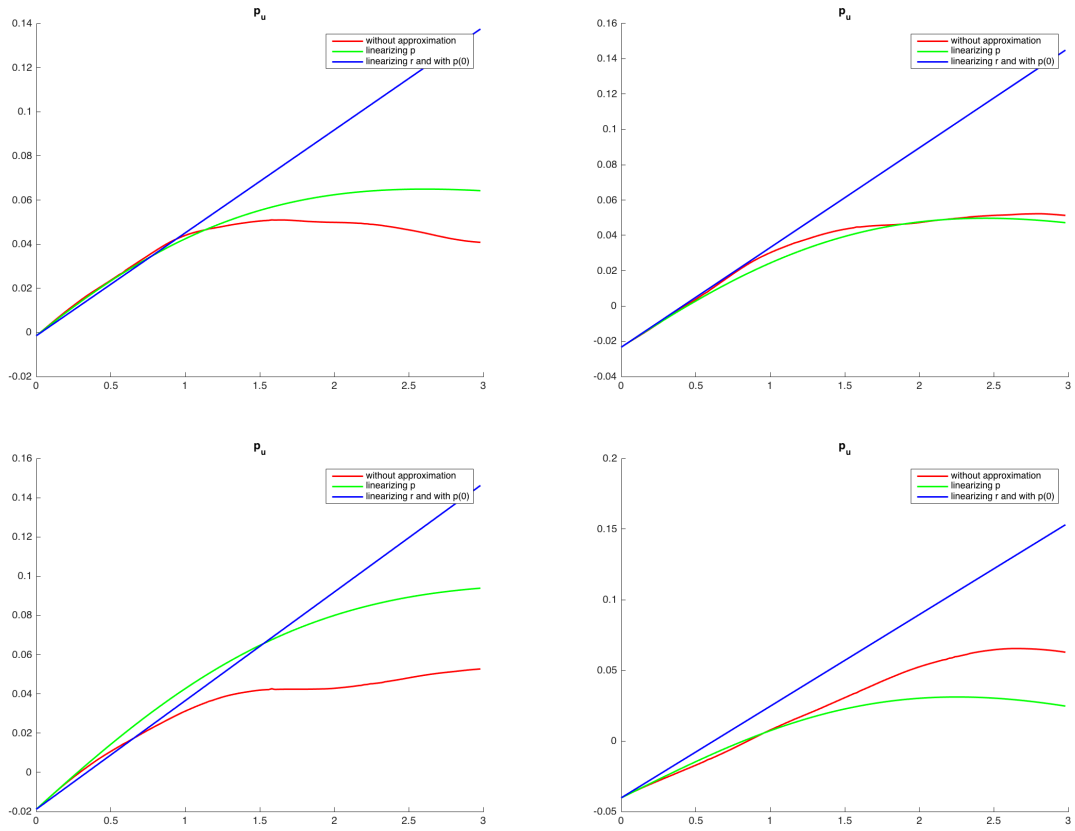

Figure 4.6: The particular solution (after the upcrossing of the process) and its approximations. The four plots are for four realizations.

for the Duffing oscillator with restoring force as in (4.7) with $k = 1$ and $c = 3$, and spectral density for the excitation as in (4.4) with $\sigma_f = 0.0730$, chosen so as to produce upcrossings of $x_m = 0.3$ at a similar frequency as upcrossings of the knuckle point in the simulation of the PWL oscillator.

The CDF of the critical rate was numerically calculated from (4.46) for values of a from 0 to 3 with a stepsize of 0.0005, with the density computed through differencing and depicted in Figure 4.8 (top right). The sample is generated from this density through rejection sampling, and its histogram with a density estimated from kernel smoothing is depicted in Figure 4.8 (top left). The bottom plots show the shape parameter plots for both the right and left tails; negative shape parameters are estimated in both cases.

An additional 10,000 independent copies of the sample of the split-time metric was generated directly from independent samples of the rate (taken from a half-normal distribution)

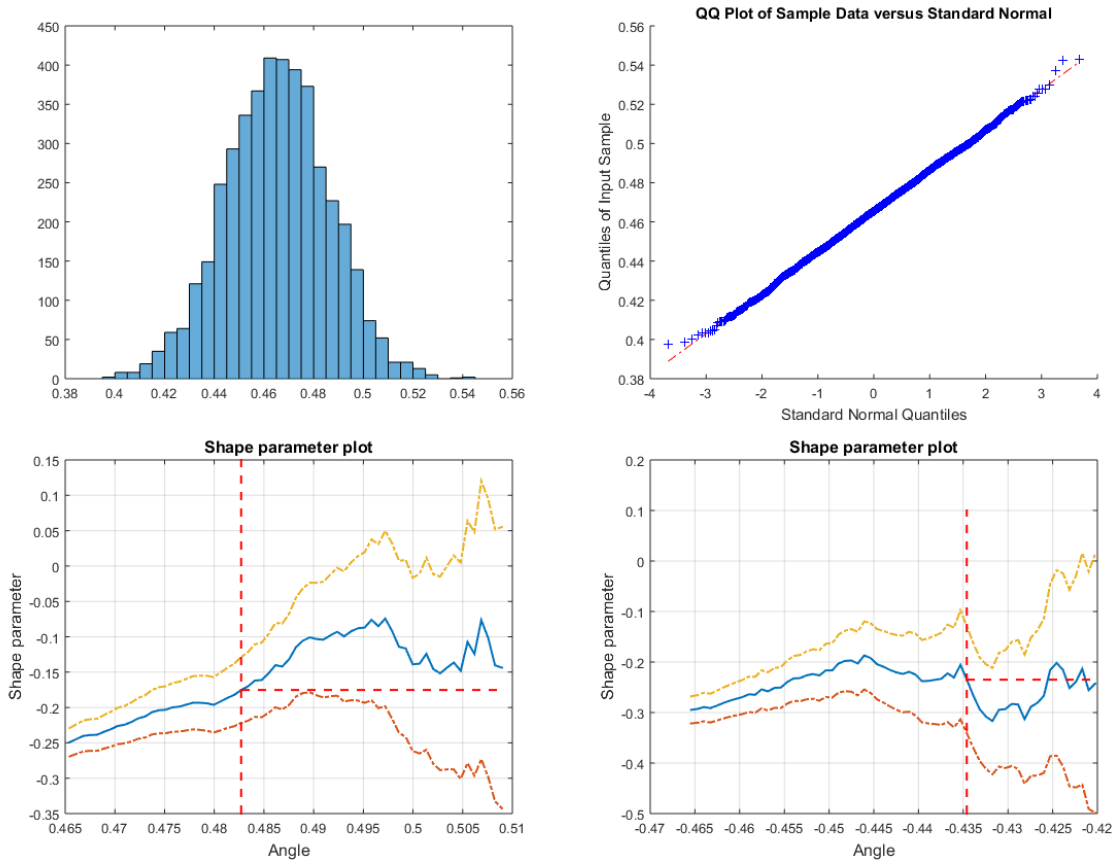

Figure 4.7: The DPWL oscillator: Top left: histogram of the critical response rates; Top right: the QQ-plot of the critical response rates; Bottom left: the shape parameter plot for the right tail; Bottom right: the shape parameter plot for the left tail.

and critical rate, and following equation (4.8). Exploratory plots are depicted in Figure 4.9, including a histogram with density estimated from kernel smoothing, histogram of the right tail with a fitted half-normal density, a shape parameter plot for the right tail, and a QQ-plot for the right tail only. As in the cases of the PWL and DPWL oscillator discussed in Section 4.4.1, a shape parameter value of $\xi = 0$ is within confidence bounds, consistent with a normal tail, although the QQ-plot also indicates some deviations.

4.5 Implications for extreme value analysis

We discuss here briefly several statistical implications of the findings in the previous sections, namely, concerning the uses of the exponential distribution (Section 4.5.1) and of the distribution with a Weibull tail (Section 4.5.2) in the POT approach.

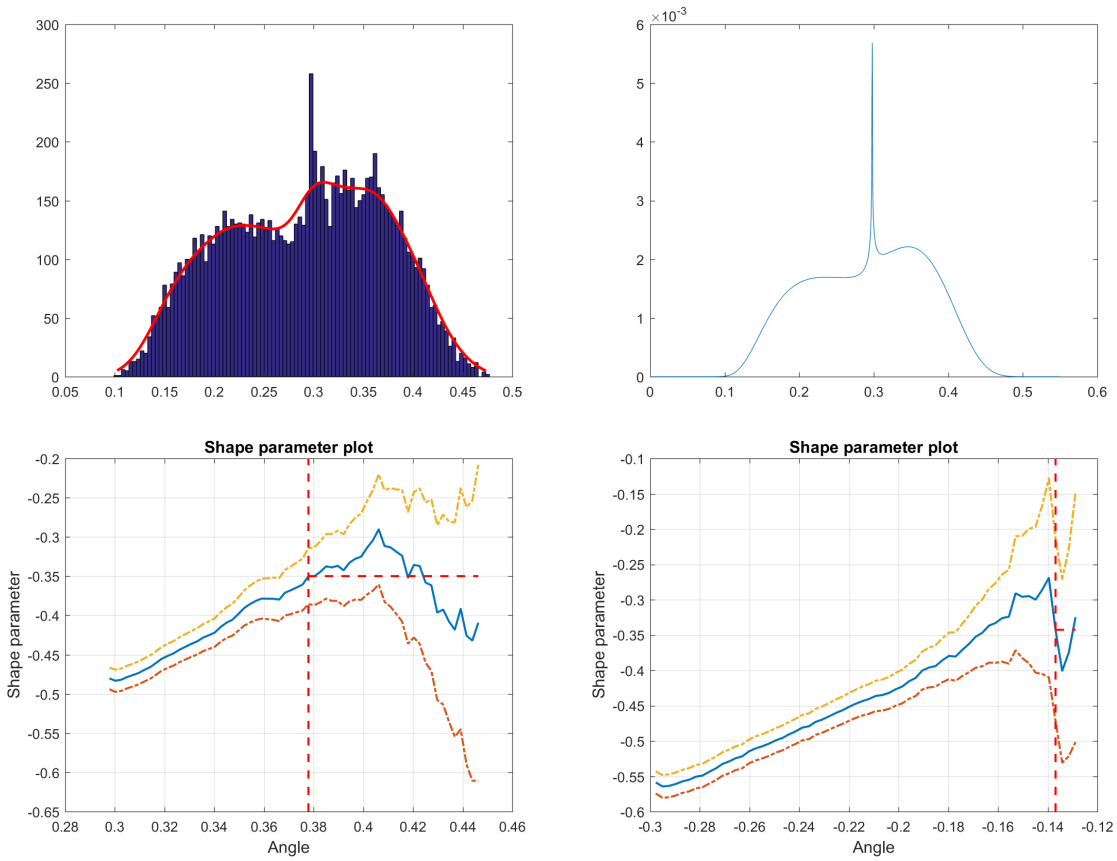

Figure 4.8: The Duffing oscillator: Top left: histogram of the critical response rates with a density estimated from kernel smoothing; Top right: density of the critical rate numerically calculated from (4.46); Bottom left: the shape parameter plot for the right tail; Bottom right: the shape parameter plot for the left tail.

4.5.1 POT approach with exponential distribution

The results of Sections 4.3 and 4.4 suggest (and in some cases show) that the distribution of the split-time metric for the considered models is in the domain of maxima attraction of the GPD with the shape parameter $\xi = 0$, that is, the exponential distribution. A natural question then is whether the POT approach described in Section 4.2.2 could be used with the exponential distribution for extrapolation above threshold, instead of the GPD. In fact, the POT approach has originated in Hydrology with using the exponential distribution for POT (e.g. see Ekanyake and Cruise [36], Madsen et al. [65] and references therein) though the setting in these early works is slightly different from that considered here (e.g. a threshold is selected based on the exceedance times following a Poisson process). We also note that fitting the exponential

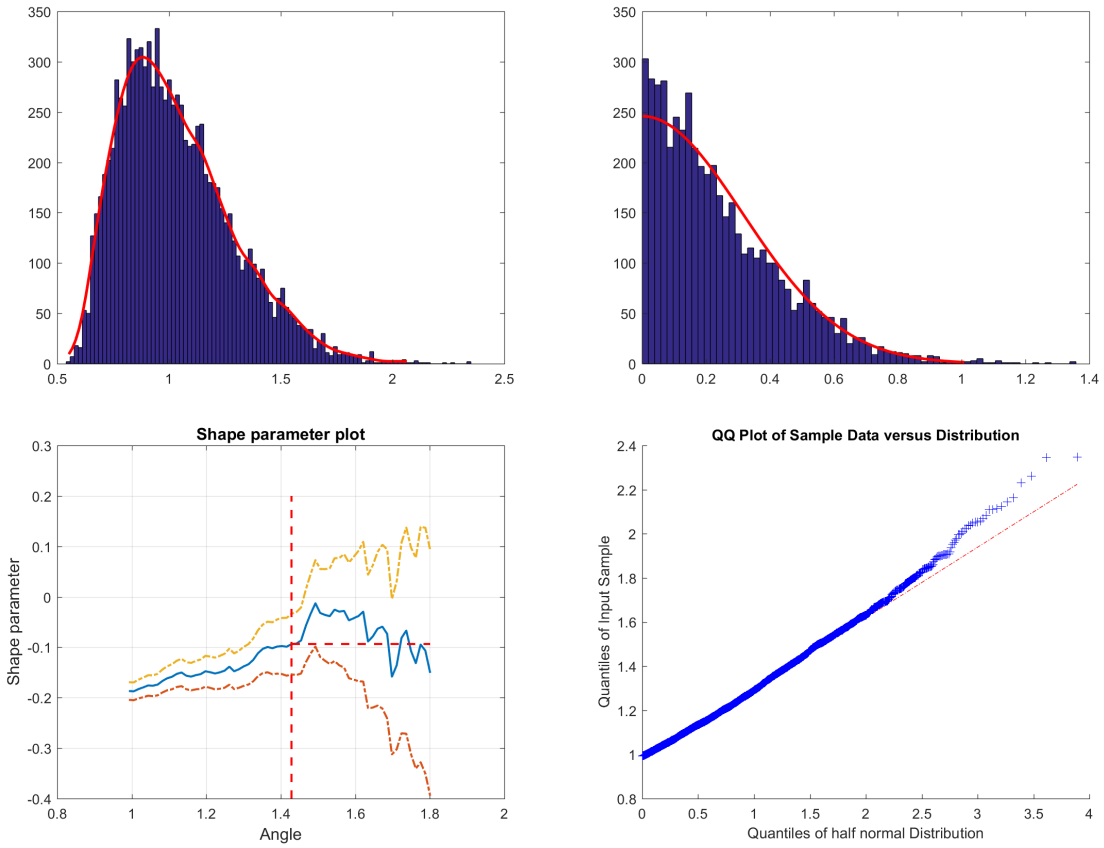

Figure 4.9: The split-time metric generated from the Duffing oscillator: Top left: histogram of the split-time metric with a numerically estimated density; Top right: histogram of the largest 50% of split-time metric values with a fitted half-normal density; Bottom left: the shape parameter plot for the right tail; Bottom right: the QQ-plot of the metric with the right-tail of a standard normal distribution.

distribution above threshold is essentially equivalent to fitting the Gumbel distribution (one of the three extreme value distributions) to block maxima (e.g. see Gomes and Guillou [45]).

In attempting to use the POT approach with the exponential distribution, the choice of a threshold (denoted u in (4.10)) is critical. The methods of threshold selection that we have been using are based on either the goodness-of-fit tests for the exponential distributions (Spinelli and Stephens [98]) or a prediction error criterion (Mager [66]). In the first approach, for a range of intermediate thresholds, we would test if the data above a given threshold is consistent with the exponential distribution (at some significance level α) and then choose the smallest threshold for which this consistency holds (the null hypothesis not rejected), including all larger thresholds. But we also found that for this method to work, the significance level should be quite large (say $\alpha = 30\%$ or 40%), that is, one should be quite sure of the fit of the exponential

distribution. Choosing large α will lead to larger selected threshold and fewer observations above the threshold, and thus also larger associated uncertainty (wider confidence intervals).

In the second approach, a threshold is chosen based on a prediction error criterion. More specifically, let $X_{n,n} \leq ... \leq X_{1,n}$ be the order statistics of a variable of interest, for example, the split-time metric d_{st}. A threshold $u = X_{k+1,n}$ or the index k is then selected as

$$\hat{k} = \arg\min_{k} \hat{\Gamma}(K). \tag{4.64}$$

Here, the minimum is searched over some range of values k, with Mager [66] suggesting to use $k \in [\min(40, 0.02n), 0.2n]$, though in simulations for synthetic distributions, we sometimes find the results to be sensitive to the choice of upper bound. The quantity $\hat{\Gamma}(k)$ is defined as

$$\begin{aligned}\hat{\Gamma}(k) = \hat{\sigma}^{-2} \sum_{i=1}^{k} \left(\frac{k+1}{i} - 1 \right)^{-1} \left(X_{i,n} - u + \hat{\sigma} \log \left(\frac{i}{k+1} \right) \right)^2 \\ + \frac{2}{k} \sum_{i=1}^{k} \left(\frac{k+1}{i} - 1 \right)^{-1} \log^2 \left(\frac{i}{k+1} \right) - 1,\end{aligned} \tag{4.65}$$

where $\hat{\sigma}$ is the scale parameter estimate of the exponential distribution based on $X_{i,n} - u$, $i = 1, ..., k$. See Mager [66], bottom of p. 64.

Figure 4.10 illustrates the POT approaches based on the GPD (left plot) and the exponential distribution (right plot), applied to extrapolating the distribution of the absolute value of the standard normal random variable with sample size $n = 1000$. A threshold for both methods is selected automatically by using the prediction error criterion mentioned above and found in Mager [66]. Each vertical line corresponds to a separate independent replication, with the circle indicating the estimate of the exceedance probability $\mathbb{P}(|N(0,1)| > 4.056) = 5 \times 10^{-5}$ and the line corresponding to the associated 95% confidence interval, all on a log vertical scale. The horizontal line shows the target probability 5×10^{-5}. As expected, about 95% of the vertical lines include the true probability in both GPD and exponential cases. But note that that confidence intervals for the exponential case are narrower, which might be an appealing feature for using the POT approach with the exponential distribution in practice – again this is assuming the knowledge that the data is from the distribution in the domain of attraction

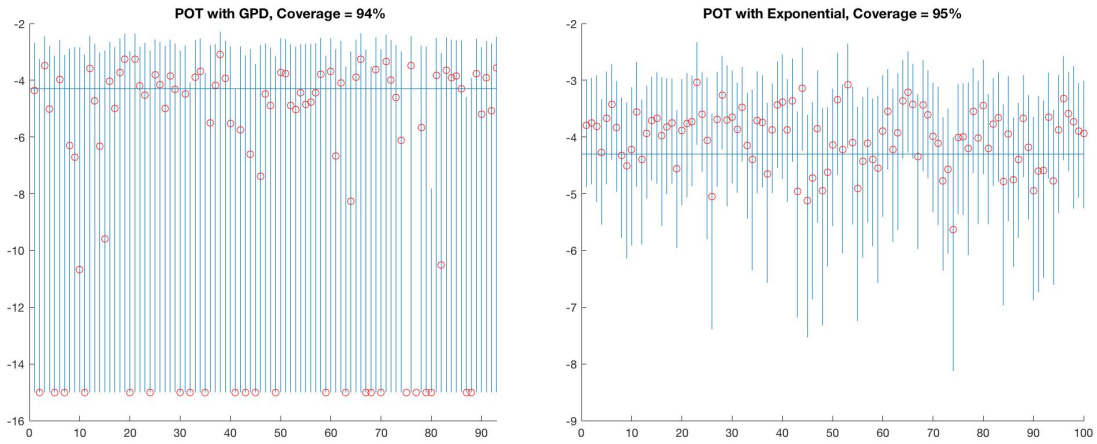

Figure 4.10: The POT approaches based on the GPD (left plot) and the exponential distribution (right plot).

of the exponential distribution, as suggested e.g. for the split-time metric through the models analyzed above.

4.5.2 POT approach for distribution with Weibull tail

Instead of using the POT approach with the exponential distribution, another interesting alternative is to use a POT approach for a distribution having a Weibull tail. More specifically, supposing X is a variable of interest (say positive), it is said to have a Weibull tail if

$$\mathbb{P}(X > x) = e^{-L(x)x^{1/\theta}}, \quad x > 0, \tag{4.66}$$

where θ is a parameter and $L(x)$ is a slowly varying function at infinity. Distributions with Weibull tails are of interest here for several reasons. First, the split-time metric in the models considered in Section 4.3 was either proved or strongly suggested to have a Weibull tail (with $\theta = 0.5$). Second, from the GPD perspective, a distribution with a Weibull tail falls in the domain of attraction of the GPD with zero shape parameter, that is, the exponential distribution. Distributions with Weibull tails then provide an interesting flexible family of distribution tails for extrapolation when in such a domain of attraction. Third, the parameter θ and its existence are informative on their own. Fourth, as noted below, there is an analogous POT approach for distributions with Weibull tails.

Indeed, numerous works in Extreme Value Analysis concern estimation of θ and related questions. See, for example, Gardes and Girard [40], Gardes and Girard [40], Diebolt et al. [33], Asimit et al. [5], Gardes et al. [41]. The POT approach for a distribution with a Weibull tail is based on the observation that, for some critical value of interest x_{cr},

$$\begin{aligned} \log \mathbb{P}(X > x_{\text{cr}}) &= \log \mathbb{P}(X > u) \frac{\log \mathbb{P}(X > x_{\text{cr}})}{\log \mathbb{P}(X > u)} \\ &= \log \mathbb{P}(X > u) \frac{L(x_{\text{cr}}) x_{\text{cr}}^{1/\theta}}{L(u) u^{1/\theta}} \approx (\log \mathbb{P}(X > u)) \left(\frac{x_{\text{cr}}}{u} \right)^{1/\theta}, \end{aligned} \tag{4.67}$$

where u is an intermediate threshold and we used (4.66). As with the usual POT approach involving GPD, the "non-rare" probability is estimated directly from the data as a suitable sample proportion. The rest of the procedure involves estimation of θ with a confidence interval and also setting a threshold u.

The issue of threshold selection is seemingly not quite resolved in the literature. See, in particular, Asimit et al. [5], Statement 1, and also Mercadier and Soulier [69], Section 4. The approach that we use adapts an ad hoc approach of Reiss and Thomas for GPD (see Neves and Fraga Alves [75]), wherein the usual bias-variance tradeoff in estimation is exploited. More specifically, an index k in $u = X_{k+1,n}$ (as in Section 4.5.1, $X_{i,n}, i = 1, ..., n$ denotes the order statistic) is chosen as

$$\hat{k} = \arg\min_k \frac{1}{k} \sum_{i=1}^{k} i^{\beta} \left| \hat{\theta}_k - \text{median} \left(\hat{\theta}_1, ..., \hat{\theta}_k \right) \right|, \tag{4.68}$$

where $\beta \in (0, \frac{1}{2}]$ is fixed and $\hat{\theta}_k$ is an estimate of θ based on $X_{i,n}, i = 1, ..., k$. Moreover, the minimum is often considered for $k \geq K$ for some fixed K (e.g. $K = 20$). Figure 4.11 presents a plot analogous to those in Figure 4.10 but based on the POT approach for a Weibull-tailed distribution. The results of the POT approach for a Weibull-tailed distribution are comparable to the POT approach based on the exponential distribution.

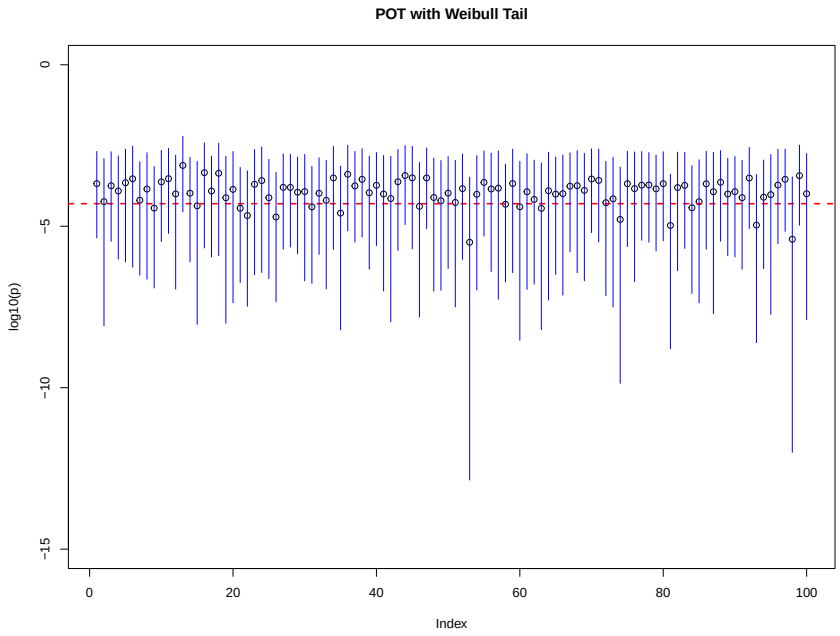

Figure 4.11: The POT approach based on the distribution with a Weibull tail.

4.6 Conclusions

In this chapter, we have examined the distributions of the critical response rate and split-time metric of the nonlinear random oscillator given in (4.1). We considered several forms of the restoring force, including piecewise linear and doubly piecewise linear stiffness functions, as well as the cubic one from the Duffing oscillator; and both correlated and white noise excitations. The distributions were examined analytically in some cases and also numerically, using techniques from extreme value theory such as the POT approach. Due to the nature of the observed tail of the distributions of both the critical response rate and the split-time metric, we considered modifying the POT approach to fit either an exponential distribution or a Weibull tail for peaks above threshold.

CHAPTER 5

Statistical inference for mean and variance of oscillatory processes

5.1 Introduction

Let $\{X_t\}_{t\in\mathbb{R}}$ be a stationary continuous-time stochastic process with mean $\mu(X) = \mathbb{E}X_t$ and variance $\sigma^2(X) = \mathrm{Var}(X_t)$. We are interested here in the processes representing random oscillatory systems, with a number of examples (linear random oscillators, etc.) discussed in Appendix B. The motivating real applications of our particular interest include processes related to ship motions (e.g. ship roll, pitch and yaw dynamics, bending moments, etc.), but potential applications would naturally arise in a range of other areas where random oscillatory systems play a fundamental role (see e.g. Gitterman [42], Neimark and Landa [74], Liang and Lee [58]). The basic problem addressed in this chapter is providing confidence intervals for the mean $\mu(X)$ and the variance $\sigma^2(X)$ (and consequently for the standard deviation $\sigma(X)$ as well) from the observed continuous-time sample X_t, $t \in [0, T]$, or the corresponding discrete sample $X_{\Delta k}$, $k = 1, \ldots, T/\Delta$. Though it should be kept in mind that the approach discussed here also applies to other quantities of possible interest, e.g. autocovariance at a given lag (see Remarks 5.3.1 and 5.3.2 below).

Focusing on the continuous-time sample X_t, $t \in [0, T]$, and on estimating the mean $\mu(X)$, consider the sample mean

$$\overline{X}_T = \frac{1}{T}\int_0^T X_s ds. \tag{5.1}$$

Its normalized asymptotic variance is

$$\lim_{T\to\infty} T\,\mathrm{Var}(\overline{X}_T) = \int_{-\infty}^{\infty} \Gamma_X(u)du =: \Pi(X), \tag{5.2}$$

where $\Gamma(u) = \mathbb{E}X_0X_u - \mu(X)^2$ is the autocovariance function. The quantity $\Pi(X)$ is known as the *long-run variance* of the process $X = \{X_t\}_{t\in\mathbb{R}}$. For example, its estimator would naturally

enter into the approximate normal confidence intervals for the mean $\mu(X)$. Similarly, define the sample variance

$$\widehat{\sigma}_T^2(X) = \frac{1}{T}\int_0^T (X_s - \overline{X}_T)^2 ds = \frac{1}{T}\int_0^T X_s^2 ds - (\overline{X}_T)^2 = \overline{X_T^2} - (\overline{X}_T)^2 \tag{5.3}$$

and set

$$\lim_{T\to\infty} T \operatorname{Var}\left(\begin{pmatrix} \overline{X}_T \\ \overline{X_T^2} \end{pmatrix}\right) = \begin{pmatrix} \int_{-\infty}^{\infty} \Gamma_X(u)du & \int_{-\infty}^{\infty} \Gamma_{X,X^2}(u)du \\ \int_{-\infty}^{\infty} \Gamma_{X,X^2}(u)du & \int_{-\infty}^{\infty} \Gamma_{X^2}(u)du \end{pmatrix}$$

$$=: \begin{pmatrix} \Pi(X) & \Pi(X,X^2) \\ \Pi(X,X^2) & \Pi(X^2) \end{pmatrix} =: \Pi^{(2)}(X,X^2), \tag{5.4}$$

where $\Gamma_{X,X^2}(u) = \mathbb{E}X(0)X(u)^2 - \mu(X)\mu(X^2)$. One then expects (see Section 5.3 for details) that

$$\lim_{T\to\infty} T \operatorname{Var}(\widehat{\sigma}_T^2(X)) = \Pi(X^2) - 4\mu(X)\Pi(X,X^2) + 4\mu(X)^2\Pi(X). \tag{5.5}$$

The matrix quantity $\Pi^{(2)}(X,X^2)$ in (5.4) is the *long-run variance* of the vector process $(X,X^2)' = \{(X_t, X_t^2)'\}_{t\in\mathbb{R}}$. Here and throughout the chapter, the prime indicates a transpose. The long-run variances are also equal to (possibly up to a multiplicative constant) the spectral densities of the processes at the origin.

Estimation of the long-run variances, such as $\Pi(X)$ and $\Pi^{(2)}(X,X^2)$ above, is a well-studied problem, especially for discrete-time processes (series) X. The often cited seminal reference is Andrews [3], though the pioneering work in the parallel problem of estimating the spectral density of the process goes back at least to Parzen [78] and others. In this chapter, we shall refer often to Lu and Park [64] who focused on continuous-time stochastic processes satisfying a stochastic differential equation and estimation issues from discrete samples. Other approaches not investigated here involve e.g. resampling (see e.g. Politis and Romano [82]) and self-normalization (see e.g. Shao [89].

One of the goals of this work is to revisit estimation of the long-run variances according to the proposed methods, and to assess their performance on the processes associated with oscillatory systems. The case of the inference of the mean is considered in Section 5.2, and that of the variance in Section 5.3. The numerical results are reported in Section 5.7. Aside from

some natural adaptations specific to oscillatory systems, this part of the work is rather of the informative nature, especially to those unfamiliar with the relevant literature.

Several other aspects of this work, however, are more novel. Estimation of the long-run variance referenced above assumes implicitly that the long-run variance is non-zero. In Section 5.4, we also study the case when the long-run variance is zero, which is quite plausible with oscillatory systems and is also the case in the considered data application (Section 5.7.2). As shown, perhaps somewhat surprisingly, inference about the mean in the case of zero long-run variance is in fact conceptually simpler than that in the case of non-zero long-run variance. In practice, as we argue below, the so-called unit root tests can be used to decide on whether given data are consistent with the hypotheses of non-zero long-run variance.

As another novel aspect of the general problem of estimation of long-run variance, we also study the case when a number of independent records of the same process are available. See Section 5.5. This is a quite common situation in several applications, for example, related to ship motions, where multiple records would correspond to different experimental runs in an actual model basin, or be obtained by using a computer generation. When considering multiple records, two natural estimators of the long-run variance can be considered: first, the average of the long-run variance estimators of the individual records, and second, the long-run variance estimator deduced from the sample variance of the sample means of individual records. In this regard, we introduce here an estimator that generally outperforms these two natural estimators of the long-run variance, and explain its improved performance.

Throughout this chapter, for the sake of clarity, we focus on the case when a sample of the process is available in the continuous time. Some issues behind using discrete samples of the process, which is the more realistic scenario in practice, are discussed in Section 5.6. Finally, as noted above, Section 5.7 contains numerical results assessing the performance of the methods described in this chapter, and Appendix B details a number of processes associated with oscillatory systems that are considered here.

5.2 Estimation of mean

As in Section 5.1, consider a continuous-time stationary process $X = \{X_t\}_{t\in\mathbb{R}}$ with mean $\mu(X) = \mathbb{E}X_t$ and finite variance $\sigma^2(X) = \mathrm{Var}(X_t)$. Suppose given a continuous-time sample X_t, $t \in [0, T]$, from which one would like to estimate the mean $\mu(X)$ by using the sample mean $\overline{X}_T$ in (5.1). The approximate normal confidence intervals for $\mu(X)$ would include the long-run variance $\Pi(X)$ in (5.2), which needs to be estimated from data. In defining the estimator of the long-run variance, the so-called kernel function plays a key role.

A kernel K is a function $K : \mathbb{R} \to \mathbb{R}$ such that it is symmetric, $K(0) = 1$ and $\int_{\mathbb{R}} K(x)^2 dx < \infty$. (In some instances, additional assumptions are made.) Let

$$K_r = \lim_{x\to 0} \frac{1 - K(x)}{|x|^r}, \quad r \in \mathbb{Z}^+, \tag{5.6}$$

and $\nu = \max\{r : K_r < \infty\}$. The quantity ν measures smoothness of the kernel $K(x)$ at $x = 0$. Examples of commonly used kernels are Quadratic Spectral (QS): $K(x) = \frac{25}{12\pi^2x^2}\Big(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\Big)$, Bartlett: $K(x) = (1 - |x|)1_{\{|x|\le 1\}}$, Truncated: $K(x) = 1_{\{|x|\le 1\}}$. (The notation 1_A stands for the indicator function of the set A, that is, $1_A(x) = 1$ if $x \in A$, and $= 0$ otherwise.) For QS kernel, $\nu = 2$ and $K_2 \approx 1.4212$. For Bartlett kernel, $\nu = 1$ and $K_1 = 1$.

Following the main approach found in the literature, the estimator of the long-run variance $\Pi(X)$ in (5.2) is then defined as

$$\widehat{\Pi}_T(X) = \int_{-T}^{T} K\Big(\frac{u}{S_T}\Big)\widehat{\Gamma}_T(u)du, \tag{5.7}$$

where S_T (which is smaller than T) is known as the *bandwidth*, and

$$\widehat{\Gamma}_T(u) = \frac{1}{T}\int_0^{T-u} (X_{s+u} - \overline{X}_T)(X_s - \overline{X}_T)ds, \quad \widehat{\Gamma}_T(-u) = \widehat{\Gamma}_T(u),\ 0 < u < T, \tag{5.8}$$

estimates the autocovariance function of the process X.

The optimal value of the bandwidth, balancing the asymptotic bias and variance of the estimator $\widehat{\Pi}_T(X)$, was derived by Lu and Park [64], and is given by

$$S_{\mathrm{opt},T} = \Big(\frac{\nu K_\nu^2 C_\nu(X)^2}{\int K(x)^2 dx} T\Big)^{1/(2\nu+1)}, \tag{5.9}$$

where K_ν and ν are associated with the kernel K and defined in (5.6) and following (5.6), respectively,

$$C_\nu(X) = \frac{\Lambda_\nu(X)}{\Pi(X)} \tag{5.10}$$

and

$$\Lambda_\nu(X) = \int_{\mathbb{R}} |u|^\nu \Gamma_X(u) du, \quad \nu \in \mathbb{Z}^+. \tag{5.11}$$

Remark 5.2.1. If S_X is the spectral density of the process X satisfying (by convention)

$$S_X(w) = \int_{-\infty}^{\infty} \Gamma_X(u)\cos(wu)du, \quad \Gamma_X(h) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(w)\cos(hw)dw, \tag{5.12}$$

then for even ν,

$$\Lambda_\nu(X) = (-1)^{\nu/2} \frac{d^\nu S_X(w)}{dw^\nu}\Big|_{w=0}.$$

As mentioned above, we also have $\Pi(X) = S_X(0)$.

Note that the constant $C_\nu(X)$ in (5.10) depends on the underlying process X. Several ways of estimating this constant have been suggested. The most popular method seems to be *model-driven*. For this method, a model would be fitted to data and the constant $C_\nu(X)$ would be calculated based on the fitted model. For example, Lu and Park [64] focus on continuous-time Ornstein-Uhlenbeck (mean reversion) type models. In the context of oscillatory systems, it seems natural to fit to the data a linear oscillator with white noise excitation (see Appendix B). Several ways and references for performing this task from discrete sample are discussed in Section 5.6.

Another method for estimating $C_\nu(X)$ in (5.10) is *data-driven*. In this method, the idea is to use kernel-based estimators of $\Pi(X)$ (as defined in (5.7) above) and $\Lambda_\nu(X)$ (as in (5.7) but replacing $\widehat{\Gamma}_T(u)$ by $|u|^\nu\widehat{\Gamma}_T(u)$) with some crude choice of the bandwidth S_T. In Section

5.7 below, we use the bandwidth $S_T = \sqrt{T}$ when estimating $\Pi(X)$ and $\Lambda_\nu(X)$ for the constant $C_\nu(X)$ (but see also Remark 5.6.2). The model-driven approach is often cited in the literature as being more stable than the data-driven approach, but our simulation study suggests that the data-driven method is often as good, if not superior.

Finally, for the considered estimator $\widehat{\Pi}_T(X)$ of the long-run variance, an approximate normal confidence interval for the mean is defined as

$$\Big(\overline{X}_T - z_\alpha\sqrt{\widehat{\mathrm{Var}}(\overline{X}_T)}, \overline{X}_T + z_\alpha\sqrt{\widehat{\mathrm{Var}}(\overline{X}_T)}\Big), \tag{5.13}$$

where

$$\widehat{\mathrm{Var}}(\overline{X}_T) = \frac{\widehat{\Pi}_T(X)}{T} \tag{5.14}$$

and z_α is the critical value for the standard normal distribution associated with a confidence level α.

5.3 Estimation of variance

We now turn to inference for the variance $\sigma^2(X)$ of a continuous-time stationary process $X = \{X_t\}_{t\in\mathbb{R}}$, through the sample variance $\widehat{\sigma}^2_T(X)$ defined in (5.3). As indicated around (5.3)–(5.5) in Section 5.1, the long-run variance $\Pi^{(2)}(X, X^2)$ of the bivariate process $(X, X^2)' = \{(X_t, X_t^2)'\}_{t\in\mathbb{R}}$ in (5.4) now plays a central role. It can be estimated similarly to the long-run variance $\Pi(X)$, as outlined next.

As in Section 5.2, define the kernel-based estimator of the long-run variance $\Pi^{(2)}(X, X^2)$ as

$$\widehat{\Pi}^{(2)}_T(X, X^2) = \int_{-T}^{T} K\Big(\frac{u}{S_T}\Big)\widehat{\Gamma}^{(2)}_T(u)du, \tag{5.15}$$

where S_T is the *bandwidth*, and

$$\widehat{\Gamma}^{(2)}_T(u) = \frac{1}{T}\int_0^{T-u} \begin{pmatrix} X_{s+u} - \overline{X}_T \\ X^2_{s+u} - \overline{X^2_T} \end{pmatrix}\begin{pmatrix} X_s - \overline{X}_T \\ X^2_s - \overline{X^2_T} \end{pmatrix}' ds, \quad \widehat{\Gamma}^{(2)}_T(-u) = \widehat{\Gamma}^{(2)}_T(u),\ 0 < u < T, \tag{5.16}$$

estimates the autocovariance matrix function of the bivariate process $(X, X^2)'$. This also naturally yields the estimators $\widehat{\Pi}_T(X)$, $\widehat{\Pi}_T(X^2)$ and $\widehat{\Pi}_T(X, X^2)$ of the components $\Pi(X)$, $\Pi(X^2)$ and $\Pi(X, X^2)$ of the long-run variance matrix $\Pi^{(2)}(X, X^2)$ in (5.4). The optimal value of the bandwidth, balancing the asymptotic bias and variance of the estimator $\widehat{\Pi}_T^{(2)}(X, X^2)$, was derived by Lu and Park [64], and is given by

$$S_{\text{opt},T} = \Big(\frac{\nu K_\nu^2 C_\nu(X, X^2)^2}{\int K(x)^2 dx} T\Big)^{1/(2\nu+1)}, \tag{5.17}$$

where K_ν and ν are associated with the kernel K and defined in (5.6) and following (5.6), respectively,

$$C_\nu(X, X^2) = \Big(\frac{\Lambda_\nu^2(X) + \Lambda_\nu^2(X^2)}{\Pi^2(X) + \Pi^2(X^2)}\Big)^{1/2} \tag{5.18}$$

and $\Lambda_\nu(\cdot)$ is defined in (5.11). (In fact, the optimal bandwidth derived by Lu and Park [64] allows for different weights in balancing the asymptotic bias and variance of the components of the matrix estimator $\widehat{\Pi}_T^{(2)}(X, X^2)$; the bandwidth given above corresponds to equal "diagonal" weights in the weighing scheme.) In practice, the constant $C_\nu(X, X^2)$ is estimated through either the model- or data-driven approaches in the same way as discussed in Section 5.2 following Remark 5.2.1.

Finally, with the estimator $\widehat{\Pi}_T^{(2)}(X, X^2) =: (\widehat{\Pi}_T(X)\ \widehat{\Pi}_T(X, X^2); \widehat{\Pi}_T(X, X^2)\ \widehat{\Pi}_T(X^2))$ of the long-run variance matrix given by (5.15)–(5.16), an approximate normal confidence interval for the variance is defined as

$$\Big(\widehat{\sigma}_T^2(X) - z_\alpha \sqrt{\widehat{\mathrm{Var}}(\widehat{\sigma}_T^2(X))}, \widehat{\sigma}_T^2(X) + z_\alpha \sqrt{\widehat{\mathrm{Var}}(\widehat{\sigma}_T^2(X))}\Big), \tag{5.19}$$

where

$$\widehat{\mathrm{Var}}(\widehat{\sigma}_T^2(X)) = \frac{\widehat{\Pi}_T(X^2) - 4\overline{X}_T \widehat{\Pi}_T(X, X^2) + 4(\overline{X}_T)^2 \widehat{\Pi}_T(X)}{T} \tag{5.20}$$

and z_α is the critical value for the standard normal distribution associated with confidence level α. The numerator in (5.20) is motivated by the form of the asymptotic variance of $\widehat{\sigma}_T^2(X)$ in (5.5). The latter form itself is a consequence of the delta method applied to the right-hand side of (5.3) and using (5.4).

Remark 5.3.1. A similar application of the delta method to the sample standard deviation $\widehat{\sigma}_T(X) = (\widehat{\sigma}_T^2(X))^{1/2}$ yields an approximate normal confidence interval for the standard deviation $\sigma(X)$ of the process, defined as

$$\Big(\widehat{\sigma}_T(X) - z_\alpha\sqrt{\widehat{\mathrm{Var}}(\widehat{\sigma}_T(X))}, \widehat{\sigma}_T(X) + z_\alpha\sqrt{\widehat{\mathrm{Var}}(\widehat{\sigma}_T(X))}\Big), \tag{5.21}$$

where z_α is as in (5.19) and

$$\widehat{\mathrm{Var}}(\widehat{\sigma}_T(X)) = \frac{\widehat{\Pi}_T(X^2) - 2(\overline{X}_T)^2\widehat{\Pi}_T(X) - 4(\overline{X}_T)^2\widehat{\Pi}_T(X, X^2)}{T(4(\overline{X^2}_T - (\overline{X}_T)^2))}. \tag{5.22}$$

Remark 5.3.2. The inference approach outlined above can also be applied to estimating quantities of interest other than the mean, the variance or the standard deviation. For example, the autocovariance of a process X at lag $h > 0$, defined as

$$\gamma_X(h) = \mathbb{E}X_0X_h - \mu(X)^2 = \mathbb{E}(X_0 - \mu(X))(X_h - \mu(X)),$$

is naturally estimated through

$$\widehat{\gamma}_T(h) = (\overline{XX}_{\cdot+h})_T - (\overline{X}_T)^2,$$

where

$$(\overline{XX}_{\cdot+h})_T = \frac{1}{T}\int_0^{T-h}(X_{s+h} - \overline{X}_T)(X_s - \overline{X}_T)^2 ds.$$

In this case, the long-run variance matrix of the bivariate process $(X, \overline{XX}_{\cdot+h})' = \{(X_t, X_tX_{t+h})',\ t \in \mathbb{R}\}$ is relevant and could be estimated similarly to (5.15).

5.4 Degenerate case

The discussion in Sections 5.2 and 5.3 assumes implicitly that the long-run variance $\Pi(X) \neq 0$. What happens in the case

$$\Pi(X) = 0 \quad (\text{equivalently}, S_X(0) = 0), \tag{5.23}$$

where S_X is the spectral density of the process X defined in Remark 5.2.1? This case is known as *degenerate* (e.g. Lee [55], pp. 1-22). Having zero long-run variance is quite plausible for processes associated with oscillatory systems, since their autocovariance function exhibits naturally an oscillating pattern and thus can integrate to 0.

In the degenerate case, since $S_X(0) = 0$ and $S_X(w)$ is even and often smooth at $w = 0$, it is expected that

$$\int_{-\infty}^{\infty} \frac{S_X(w)}{|w|^2} dw < \infty. \tag{5.24}$$

Under this assumption, one can write

$$\int_0^t (X_s - \mu(X))ds = V_t - V_0, \tag{5.25}$$

where $\{V_t\}_{t\in\mathbb{R}}$ is a *stationary* process with zero mean. Indeed, by writing X_s in the spectral domain as $X_s - \mu(X) = \int_{-\infty}^{\infty} e^{isw}\sqrt{S_X(w)}Z(dw)$ with a complex-valued random measure $Z(dw)$ having orthogonal increments and $\mathbb{E}|Z(dw)| = dw^2/(2\pi)$, note that

$$\int_0^t (X_s - \mu(X))ds = \int_{-\infty}^{\infty} \Big(\int_0^t e^{isw} ds \Big) \sqrt{S_X(w)} Z(dw)$$

$$= \int_{-\infty}^{\infty} e^{itw} \frac{\sqrt{S_X(w)}}{iw} Z(dw) - \int_{-\infty}^{\infty} \frac{\sqrt{S_X(w)}}{iw} Z(dw),$$

that is, the stationary process V_t can be taken as $\int_{-\infty}^{\infty} e^{itw}(\sqrt{S_X(w)}/iw)Z(dw)$ and has the spectral density $S_X(w)/|w|^2$. It follows from (5.25) that

$$T(\overline{X}_T - \mu(X)) = \int_0^T (X_s - \mu(X))ds \stackrel{d}{=} V_T - V_0. \tag{5.26}$$

Note that, compared to (5.2), the convergence rate of $\mathrm{Var}(\overline{X}_T)$ is then T^2, and not T. Note also that $V_T - V_0$ does not need to be Gaussian.

Thus, in view of (5.25), the integrated process $\int_0^t (X_s - \mu(X))ds$ is nearly stationary itself, especially as t increases (since dependence between V_t and V_0 is then expected negligible). In particular, we expect that, for large T,

$$T(\overline{X}_T - \mu(X)) \approx V_0' - V_0, \tag{5.27}$$

where V_0' is an independent copy of V_0, and an approximate (not necessarily normal) confidence interval can be constructed as

$$\Big(\overline{X}_T - \frac{t_\alpha(\widehat{V_0' - V_0})}{T}, \overline{X}_T + \frac{t_\alpha(\widehat{V_0' - V_0})}{T}\Big), \tag{5.28}$$

where $t_\alpha(V_0' - V_0)$ denotes the $(1+\alpha)/2$ quantile of $V_0' - V_0$ and the hat its estimator. Note that, in practice, the value $(-V_0)$ can be estimated as the sample mean of $\int_0^t (X_s - \overline{X}_T)ds$, $0 < t < T$, and then removed from $V_t - V_0$ to obtain a realization of V_t. The latter realization can then be used to estimate the corresponding quantile of $V_0' - V_0$: in practice, we use resampling from a realization of V_t to obtain a sample of the values of $V_0' - V_0$, from which we then select a desired quantile.

We also note that from a practical perspective, a decision needs to be made on whether the underlying process is in the degenerate case or, equivalently, whether the integrated process $\int_0^t (X_s - \mu(X))ds$ can be viewed stationary. As discussed in Section 5.7 below, this can be achieved by using one of the available so-called unit root tests.

Finally, the discussion above concerns the case of estimating the mean. Turning to estimation of the variance, recall that it involves estimation of the mean of the square process X^2. If X can be expected to fall in the degenerate case, we do not expect this to be the case with the square process X^2. For example, if X is Gaussian, then $\Gamma_{X^2}(h) = 2(\Gamma_X(h))^2$ and $\Pi(X^2) = \int_{-\infty}^{\infty} \Gamma_{X^2}(h) \neq 0$. If X falls in the degenerate case but X^2 does not, then

$$\lim_{T\to\infty} T \operatorname{Var}(\widehat{\sigma}_T^2(X)) = \Pi(X^2),$$

since the rate of convergence of $(\overline{X}_T)^2$ to $\mu(X)^2$ is faster than $\sqrt{T}$. Hence, only the long-run variance of the process X^2 needs to be estimated.

5.5 The case of multiple records

Consider a number of records $X_t^{(r)}$, $t \in [0, T_r]$, $r = 1, \ldots, R$, that are independent across r. Suppose that

$$T_r = C_r T \quad \text{with} \quad C_r \in (0,1),\ \sum_r C_r = 1. \tag{5.29}$$

Thus, $\sum_r T_r = T$. For example, the case of the records of equal length corresponds to

$$C_r = 1/R \quad \text{and} \quad T_r = T/R.$$

Considering inference for the mean first, it can be estimated through

$$\overline{X}_T = \sum_r C_r \overline{X^{(r)}_{T_r}}, \tag{5.30}$$

that is, as the appropriately weighted average of the sample means of the R records. Note that

$$T \operatorname{Var}(\overline{X}_T) = T \sum_r C_r^2 \operatorname{Var}(\overline{X^{(r)}_{T_r}}) = \sum_r C_r T_r \operatorname{Var}(\overline{X^{(r)}_{T_r}}) \to \sum_r C_r \Pi(X) = \Pi(X). \tag{5.31}$$

For confidence intervals, the long-run variance $\Pi(X)$ thus again needs to be estimated, but this time from R records. We focus in this section on the non-degenerate case when the long-run variance is non-zero.

We note first that several natural estimators of the long-run variance can be introduced in the case of multiple records. First, there is the *weighted average* of the estimators of the long-run variance across the records, defined as

$$\widehat{\Pi}_{\text{ave},T}(X) = \sum_r C_r \widehat{\Pi}^{(r)}_{T_r}(X) = \sum_r C_r \int_{-T_r}^{T_r} K\Big(\frac{u}{S_{T_r}}\Big) \widehat{\Gamma}^{(r)}_{T_r}(u) du, \tag{5.32}$$

where $\widehat{\Pi}^{(r)}_{T_r}(X)$ is the estimator of the long-run variance for the rth record. Second, a *direct estimator* of the long-run variance can be defined as a properly normalized sample variance of the sample means across the records. Indeed, consider the case of equal length records and recall that $\Pi(X)$ approximately equals $T_r \operatorname{Var}(\overline{X}_{T_r})$. But since one now has R sample means $\overline{X}_{T_r}$, $r = 1, \ldots, R$, one can naturally estimate $\operatorname{Var}(\overline{X}_{T_r})$ through the sample variance of $\overline{X}_{T_r}$, $r = 1, \ldots, R$, and then normalize it by T_r to get the estimator of the long-run variance. In the case of equal length records, this direct estimator is

$$\widehat{\Pi}_{\text{dir},T}(X) = \frac{T_r}{R} \sum_r \Big(\overline{X}_{T_r} - \overline{X}_T\Big)^2 = T \sum_r \frac{1}{R^2} \Big(\overline{X}_{T_r} - \overline{X}_T\Big)^2. \tag{5.33}$$

This motivates the following definition in the general case,

$$\widehat{\Pi}_{\mathrm{dir},T}(X) = T\sum_r C_r^2\Big(\overline{X}_{T_r} - \overline{X}_T\Big)^2. \tag{5.34}$$

Note that this definition is also consistent with the expressions in (5.31) where $\mathrm{Var}(\overline{X_{T_r}^{(r)}})$ is replaced by a "naive" estimator $(\overline{X}_{T_r} - \overline{X}_T)^2$. Note also that the estimator $\widehat{\Pi}_{\mathrm{dir},T}(X)$ is unique to the case of multiple records.

Which of the estimators, $\widehat{\Pi}_{\mathrm{ave},T}(X)$ or $\widehat{\Pi}_{\mathrm{dir},T}(X)$, should be preferred? The simulation results in Section 5.7 suggest that $\widehat{\Pi}_{\mathrm{ave},T}(X)$ is superior to $\widehat{\Pi}_{\mathrm{dir},T}(X)$ in terms of the variance, but that it can also be inferior in terms of the bias. In fact, another estimator can be introduced which enjoys both advantages of the two natural estimators. To motivate the definition of the new estimator, we shall rewrite the estimator $\widehat{\Pi}_{\mathrm{dir},T}(X)$ in a different form as follows.

Observe that

$$\widehat{\Pi}_{\mathrm{dir},T}(X) = T\sum_r C_r^2\Big(\frac{1}{T_r}\int_0^{T_r} X_s^{(r)}ds - \overline{X}_T\Big)^2 = T\sum_r \frac{C_r^2}{T_r^2}\Big(\int_0^{T_r}(X_s^{(r)} - \overline{X}_T)ds\Big)^2$$

$$= \sum_r \frac{C_r}{T_r}\int_0^{T_r}\int_0^{T_r}(X_s^{(r)} - \overline{X}_T)(X_t^{(r)} - \overline{X}_T)dsdt = \sum_r C_r\int_{-T_r}^{T_r}\widehat{\Gamma}_{0,T_r}^{(r)}(u)du, \tag{5.35}$$

where

$$\widehat{\Gamma}_{0,T_r}^{(r)}(u) = \frac{1}{T_r}\int_0^{T_r-u}(X_{s+u}^{(r)} - \overline{X}_T)(X_s^{(r)} - \overline{X}_T)ds, \quad \widehat{\Gamma}_{0,T_r}^{(r)}(-u) = \widehat{\Gamma}_{0,T_r}^{(r)}(u),\ 0 < u < T_r. \tag{5.36}$$

There are two key differences between the *average* estimator (5.32) and the *direct* estimator (5.35): first, the *average* estimator uses a kernel function to down weigh the effects of the estimator of the autocovariance function at large lags (and thus to reduce the variance), and second, the *direct* estimator employs the mean across the records in estimating the autocovariance function (and thus reducing the bias). In view of these differences, it natural to introduce the following estimator of the long-run variance,

$$\widehat{\Pi}_T(X) = \sum_r C_r\int_{-T_r}^{T_r} K\Big(\frac{u}{S_{T_r}}\Big)\widehat{\Gamma}_{0,T_r}^{(r)}(u)du =: \sum_r C_r\widehat{\Pi}_{0,T_r}^{(r)}(X), \tag{5.37}$$

that is, defined as the average of the estimators of the long-run variance for the R records which use the mean across all records in the estimation of the autocovariance function. As shown in Section 5.7, this estimator generally outperforms the *average* and the *direct* estimators of the long-run variance.

Finally, we note that the superiority of the estimator in (5.37) is more apparent when X is replaced by the square process X^2, which is relevant when estimating the variance $\sigma^2(X)$ of the process (Section 5.3 above).

5.6 Discretization and other issues

In Sections 5.2–5.5, we assumed that a continuous-time sample (or multiple samples) of the analyzed process is available. In practice, however, a discrete(-time) sample of the process is given, namely, $X_{k\Delta}$, $k = 1, \ldots, T/\Delta$, where $\Delta > 0$ is the discretization step and we assume for simplicity that

$$\frac{T}{\Delta} = n \tag{5.38}$$

is an integer. We examine here a number of issues behind using a discrete sample rather than a continuous-time one, when estimating the process mean and the process variance. Our goal is not to provide any formal proofs (as e.g. in Lu and Park [64] concerning estimation of the long-run variance through discrete sample) but rather to guide a practitioner through a number of issues that arise from the practical perspective.

5.6.1 Discretizing proposed estimators

The various estimators of long-run variances introduced above (see (5.7), (5.15), (5.32), (5.34) and (5.37)) involve integrals in continuous time which can be discretized when working with discrete samples. For example, the discrete version of the estimator (5.7) of the long-run

variance is defined as

$$\begin{aligned}\widehat{\Pi}_{T,\Delta}(X) &= \sum_{j=-(T-\Delta)/\Delta}^{(T-\Delta)/\Delta} K\Big(\frac{j\Delta}{S_{T,\Delta}}\Big)\widehat{\Gamma}_{T,\Delta}(j\Delta)\Delta \\ &= \Delta \sum_{j=-(n-1)}^{n-1} K\Big(\frac{j}{m}\Big)\widehat{\Gamma}_n(j) =: \Delta \cdot \Omega_n, \end{aligned} \tag{5.39}$$

where

$$m = \frac{S_{T,\Delta}}{\Delta} \tag{5.40}$$

is referred to as the bandwidth for the discrete sample with some discrete version $S_{T,\Delta}$ of the continuous-time bandwidth S_T,

$$\begin{aligned}\widehat{\Gamma}_{T,\Delta}(j\Delta) &= \frac{1}{T}\sum_{k=1}^{(T-j\Delta)/\Delta} (X_{(k+j)\Delta} - \overline{X}_{T,\Delta})(X_{k\Delta} - \overline{X}_{T,\Delta})\Delta \\ &= \frac{1}{n}\sum_{k=1}^{n-j}(X_{(k+j)\Delta} - \overline{X}_{T,\Delta})(X_{k\Delta} - \overline{X}_{T,\Delta}) \\ &= : \widehat{\Gamma}_n(j), \quad j = 0, 1, \ldots, n-1, \end{aligned} \tag{5.41}$$

and $\widehat{\Gamma}_{T,\Delta}(-j\Delta) = \widehat{\Gamma}_{T,\Delta}(j\Delta)$, $\widehat{\Gamma}_n(-j) = \widehat{\Gamma}_n(j)$, and

$$\overline{X}_{T,\Delta} = \frac{1}{T}\sum_{k=1}^{T/\Delta} X_{k\Delta}\Delta = \frac{1}{n}\sum_{k=1}^{n} X_{k\Delta}. \tag{5.42}$$

The expressions following the first equality signs in (5.39), (5.41) and (5.42) are written as to emphasize that the integrals in the corresponding continuous-time estimators are discretized. The quantity Ω_n in (5.39) is the usual estimator of the long-run variance

$$\Pi(X_\Delta) = \sum_{k=-\infty}^{\infty} \Gamma_{X_\Delta}(k) \quad \text{with } \Gamma_{X_\Delta}(k) = \mathbb{E}X_0X_{k\Delta} - \mu(X)^2 = \Gamma_X(k\Delta),$$

of the discrete time series $X_\Delta = \{X_{k\Delta}\}_{k\in\mathbb{Z}}$. Note that it is meaningful to multiply Ω_n by Δ in (5.39) when estimating the long-run variance $\Pi(X)$ of the continuous time process $X = \{X_T\}_{t\in\mathbb{R}}$ since

$$\Delta \cdot \Pi(X_\Delta) = \sum_{k=-\infty}^{\infty} \Gamma_X(k\Delta)\Delta \approx \int_{\mathbb{R}} \Gamma_X(u)du = \Pi(X). \tag{5.43}$$

The discrete versions of other introduced estimators, namely, (5.15), (5.32), (5.34) and (5.37), are defined in an analogous fashion by discretizing all the integrals involved.

5.6.2 Model-driven bandwidth selection

The discrete version $\widehat{\Pi}_{T,\Delta}$ of the estimator of the long-run variance defined in (5.39) uses the bandwidth m in (5.40), which requires a discrete version of $S_{T,\Delta}$ of the continuous-time bandwidth S_T defined in (5.9). We focus here on the optimal continuous-time bandwidth $S_{\mathrm{opt},T}$ in (5.9). In Section 5.2, we discussed two ways of computing $S_{\mathrm{opt},T}$ and, more specifically, the constant $C_\nu(X)$ entering $S_{\mathrm{opt},T}$: the data-driven and the model-driven approaches. The data-driven approach for discrete sample is discussed in Section 5.6.3 below. For the model-driven approach, it was suggested to use a linear oscillator with white noise excitation (see Appendix B) as an underlying model. The question then is how to fit such continuous-time model given a discrete sample.

The problem of fitting a linear oscillator with white noise excitation from discrete sample has been studied quite extensively in the literature (in fact, not just for a linear oscillator but for the more general class of the so-called continuous AR models). See e.g. Soderstrom et al. [95], Fan et al. [38], Kirshner et al. [52], Lin and Lototsky [59, 60], Pham [80]. As discussed in these references, there are delicate issues in how a discrete version of a linear oscillator is formulated and fitted to the data (assuming the linear oscillator is indeed the underlying model). For example, a scheme that works is to use the following discretization of the derivatives at $t = k\Delta$,

$$\begin{aligned} \ddot{X}_t &: \frac{X_{t+3\Delta} - 2X_{t+2\Delta} + X_{t+\Delta}}{\Delta^2} =: (\Delta^2 X)_t, \\ \dot{X}_t &: \frac{X_{t+\Delta} - X_t}{\Delta} =: (\Delta X)_t, \end{aligned} \tag{5.44}$$

and to estimate the coefficients of the model by regressing $(\Delta^2 X)_t$ on $-(\Delta X)_t$ and $-X_t$. See Soderstrom et al. [95], Example 3.3 on page 662. An estimator of the coefficient σ controlling the strength of the white noise excitation (see Appendix B) can also be given. Other "valid" discretization schemes are available as well, but in our simulations, we use the discretization scheme (5.44) only.

5.6.3 Connections to discrete time analysis

We noted following the relation (5.42) that the quantity Ω_n in (5.39) is a commonly used estimator of the long-run variance $\Pi(X_\Delta)$ of the discrete-time series $X_\Delta = \{X_{k\Delta}\}_{k\in\mathbb{Z}}$. It is naturally multiplied by Δ in (5.39) in view of (5.43). From the perspective of confidence intervals (5.13)–(5.14), note that the standard error of the sample mean used in the intervals is

$$\sqrt{\frac{\widehat{\Pi}_{T,\Delta}(X)}{T}} = \sqrt{\frac{\Delta\cdot\Omega_n}{T}} = \sqrt{\frac{\Omega_n}{n}}, \tag{5.45}$$

which is exactly the same if the whole analysis is carried out for the discrete sample itself, without any reference to the continuous-time process.

Despite the latter conclusion, however, a more delicate issue concerns the choice of the bandwidth m when used with the discrete sample $X_{k\Delta}$, $k = 1,\dots,n$. The bandwidth m was defined in (5.40) by relating it to a discrete version $S_{T,\Delta}$ of the continuous-time bandwidth S_T, but it can also be defined with the reference to the underlying discrete-time series $X_\Delta = \{X_{k\Delta}\}_{k\in\mathbb{Z}}$ alone. For a discrete-time series X_Δ, the optimal bandwidth is defined as (Lu and Park [64])

$$m_{\text{opt}} = \Big(\frac{\nu K_\nu^2 C_\nu(X_\Delta)^2}{\int K(x)^2 dx}n\Big)^{1/(2\nu+1)}, \tag{5.46}$$

where ν and K_ν are associated with a kernel function K as in Section 5.2,

$$C_\nu(X_\Delta) = \frac{\Lambda_\nu(X_\Delta)}{\Pi(X_\Delta)} \tag{5.47}$$

and

$$\Lambda_\nu(X_\Delta) = \sum_{k=-\infty}^{\infty}|k|^\nu\Gamma_{X_\Delta}(k), \quad \Pi(X_\Delta) = \sum_{k=-\infty}^{\infty}\Gamma_{X_\Delta}(k) \tag{5.48}$$

with $\Gamma_{X_\Delta}(h) = \mathbb{E}X_0X_{h\Delta} - \mu(X)^2$. It is not immediate to see but the two optimal bandwidths (5.46) and (5.9) are, in fact, connected in a natural way, as we explain next.

5.6.3.1 The case of the data-driven approach

To understand the relationship between the optimal bandwidths (5.46) and (5.9), consider first the case of the data-driven approach. In this approach and with discrete sample, the constant $C_\nu(X_\Delta)$ in (5.47) is estimated as

$$\widehat{C}_\nu(X_\Delta) = \frac{\widehat{\Lambda}_\nu(X_\Delta)}{\widehat{\Pi}(X_\Delta)}, \tag{5.49}$$

where

$$\widehat{\Pi}(X_\Delta) = \sum_{j=-(n-1)}^{n-1} K\Big(\frac{j}{m_0}\Big)\Gamma_n(j), \quad \widehat{\Lambda}_\nu(X_\Delta) = \sum_{j=-(n-1)}^{n-1} K\Big(\frac{j}{m_0}\Big)|j|^\nu\Gamma_n(j) \tag{5.50}$$

and m_0 is some preliminary crude estimate of the bandwidth. But note that

$$\Delta \cdot \widehat{\Pi}(X_\Delta) = \sum_{j=-(T-\Delta)/\Delta}^{(T-\Delta)/\Delta} K\Big(\frac{j\Delta}{m_0\Delta}\Big)\Gamma_T(j\Delta)\Delta =: \widehat{\Pi}_{T,\Delta}(X), \tag{5.51}$$

$$\Delta^{\nu+1} \cdot \widehat{\Lambda}_\nu(X_\Delta) = \sum_{j=-(T-\Delta)/\Delta}^{(T-\Delta)/\Delta} K\Big(\frac{j\Delta}{m_0\Delta}\Big)|j\Delta|^\nu\Gamma_T(j\Delta)\Delta =: \widehat{\Lambda}_{\nu,T,\Delta}(X). \tag{5.52}$$

The quantities $\widehat{\Pi}_{T,\Delta}(X)$ and $\widehat{\Lambda}_{T,\Delta}(X)$ are the discrete-time estimators of the constants $\Pi(X)$ and $\Lambda_\nu(X)$ when using the bandwidth $S_{0,T} = m_0\Delta$, and would similarly appear in the data-driven discrete-time estimator of the constant $C_\nu(X)$ in (5.10), that is,

$$\widehat{C}_{\nu,\Delta}(X) = \frac{\widehat{\Lambda}_{\nu,T,\Delta}(X)}{\widehat{\Pi}_{T,\Delta}(X)}. \tag{5.53}$$

The relations (5.51) and (5.52) show that

$$\widehat{C}_{\nu,\Delta}(X) = \Delta^{\nu+\frac{1}{2}}\widehat{C}_\nu(X_\Delta) \tag{5.54}$$

as long us the underlying bandwidths satisfy

$$S_{0,T} = m_0\Delta. \tag{5.55}$$

The relation (5.54) then implies that (the discrete estimator of) the optimal bandwidth $S_{\mathrm{opt},T}$ in (5.9) and (the estimator of) the optimal bandwidth m_{opt} in (5.46) are related as

$$S_{\mathrm{opt},T} = m_{\mathrm{opt}}\Delta. \tag{5.56}$$

Thus, in view of (5.45) and in the case of the data-driven approach, when the optimal bandwidths are used with the underlying bandwidths satisfying (5.55), inference about the sample mean is exactly the same when applying the discrete time series analysis to the series X_Δ and when the analysis is based on the assumption of the underlying continuous-time process X.

5.6.3.2 The case of the model-driven approach

Turning to the relationship between the optimal bandwidths (5.46) and (5.9) in the model-driven approach, the situation is more delicate and we shall make just a few points through a concrete example.

For discrete-time series, commonly chosen underlying models for the bandwidth calculation are the AR series (Andrews [3], Lu and Park [64]). For example, for the AR(2) series $Y = \{Y_n\}_{n\in\mathbb{Z}}$ satisfying $Y_n - \phi_1 Y_{n-1} - \phi_2 Y_{n-2} = \sigma_Z Z_n$ with white noise series $\{Z_n\}$ and coefficients ϕ_1, ϕ_2, σ_Z, and when $\nu = 2$, one has

$$\Pi(Y) = \frac{\sigma_Z^2}{2\pi}\frac{1}{(1-\phi_1-\phi_2)^2}, \quad \Lambda_2(Y) = \frac{\sigma_Z^2}{2\pi}\frac{8\phi_2 + 2\phi_1 - 2\phi_1\phi_2}{(1-\phi_1-\phi_2)^4}$$

(e.g. Andrews [3]). Then,

$$C_2(Y) = \frac{8\phi_2 + 2\phi_1 - 2\phi_1\phi_2}{(1-\phi_1-\phi_2)^2} \tag{5.57}$$

and in the model-driven approach, this constant would be estimated as

$$\widehat{C}_2(Y) = \frac{8\widehat{\phi}_2 + 2\widehat{\phi}_1 - 2\widehat{\phi}_1\widehat{\phi}_2}{(1-\widehat{\phi}_1-\widehat{\phi}_2)^2}, \tag{5.58}$$

where $\widehat{\phi}_1, \widehat{\phi}_2$ are the AR(2) parameters estimated from a discrete sample. Since the AR(2) series can exhibit oscillatory behavior, this series also seems natural to consider for discrete-time series $Y = X_\Delta$.

On the other hand, there is an analogous constant $C_2(X)$ for the model-driven approach in the continuous time supposing a linear oscillator with white noise excitation, defined by (B.1) and (B.2) in Appendix B and characterized by the parameters δ, w_0^2 and σ. For this model (see Appendix B),

$$\Pi(X) = \frac{\sigma^2}{w_0^4}, \quad \Lambda_2(X) = \frac{4(2\delta^2 - w_0^2)\sigma^2}{w_0^8},$$

so that

$$C_2(X) = \frac{4(2\delta^2 - w_0^2)}{w_0^4}. \tag{5.59}$$

What is the relationship between the constant $\widehat{C}_2(Y)$ in (5.58) when $Y = X_\Delta$ and the constant $C_2(X)$, assuming that a linear oscillator with the white noise excitation is the underlying continuous-time model?

To answer this question, let $\widehat{\phi}_1, \widehat{\phi}_2$ be the AR(2) coefficients estimated from the sample $X_{k\Delta}$, $k = 1, \ldots, n$. These are the regression coefficients when regressing $X_{(k+2)\Delta}$ on $X_{(k+1)\Delta}$ and $X_{k\Delta}$, respectively. On the other hand, let $-2\widehat{\delta}_1$ and $-\widehat{w}_{1,0}^2$ be the regression coefficients when regressing $(X_{(k+2)\Delta} - 2X_{(k+1)\Delta} + X_{k\Delta})/\Delta^2$ (which can be thought as a discrete version of $\ddot{X}_{k\Delta}$) on $(X_{(k+1)\Delta} - X_{k\Delta})/\Delta$ (which can be thought as a discrete version of $\dot{X}_{k\Delta}$) and $X_{k\Delta}$. By comaring the two regressions, note that

$$\widehat{\phi}_1 = 2 - 2\widehat{\delta}_1\Delta, \quad \widehat{\phi}_2 = -1 + 2\widehat{\delta}_1\Delta - \widehat{w}_{1,0}^2\Delta^2. \tag{5.60}$$

Then, substituting these expressions into (5.58), we obtain that

$$\widehat{C}_2(X_\Delta) = \frac{4(2\widehat{\delta}_1^2 - \widehat{w}_{1,0}^2 - \widehat{\delta}_1\widehat{w}_{1,0}^2\Delta)}{\widehat{w}_{1,0}^4} \tag{5.61}$$

It is known (see Soderstrom et al. [95], Example 3.3 on page 662) that

$$\widehat{\delta}_1 \xrightarrow{p} \frac{2}{3}\delta, \quad \widehat{w}_{1,0}^2 \xrightarrow{p} w_0^2, \tag{5.62}$$

as $\Delta \to 0$, where $\xrightarrow{p}$ denotes the convergence in probability. Thus,

$$\widehat{C}_2(X_\Delta) \xrightarrow{p} \frac{4(2(2/3)^2\delta^2 - w_0^2)}{w_0^4} \neq \frac{4(2\delta^2 - w_0^2)}{w_0^4} = C_2(X). \tag{5.63}$$

That is, the two constants in the limit of $\Delta \to 0$ are slightly different, and the two analyses, the discrete-time series analysis for the series X_Δ and the continuous-time process analysis for X, would give slightly different results in the model-driven approach.

Remark 5.6.1. In the regression above, we used the discretization scheme

$$\begin{aligned} \ddot{X}_t &: \frac{X_{t+2\Delta} - 2X_{t+\Delta} + X_t}{\Delta^2}, \\ \dot{X}_t &: \frac{X_{t+\Delta} - X_t}{\Delta}, \end{aligned} \tag{5.64}$$

for the underlying oscillator but it did not lead to the consistent estimators as noted in (5.62). A discretization scheme leading to consistent estimators was given in (5.60). Since this scheme involves $X_{t+3\Delta}$, $X_{t+2\Delta}$, $X_{t+\Delta}$ and X_t, it may appear to correspond to fitting an AR(3) series to the discrete sample $X_{k\Delta}$, $k = 1, \ldots, n$. But note that the regression of $(\Delta^2 X)_t$ on $(-\Delta X)_t$ and $(-X_t)$ for $t = k\Delta$ has the coefficient equal exactly to 2 at $X_{t+2\Delta}$ since $X_{t+2\Delta}$ appears with a factor of (-2) in $(\Delta^2 X)_t$ and not in $(-\Delta X)_t$, nor in $(-X_t)$. Thus, using the consistent discretization scheme would not be equivalent to fitting the AR(3) series to the discrete sample $X_{k\Delta}$, $k = 1, \ldots, n$.

5.6.4 Range of discretization step Δ

As the discretization step Δ approaches 0, the discrete version $\widehat{\Pi}_{T,\Delta}$ of the estimator of the long-run variance converges to $\widehat{\Pi}_T(X)$, which itself converges to the long-run variance under suitable assumptions (e.g. Lu and Park [64]). Thus, the estimator $\widehat{\Pi}_{T,\Delta}$ should stabilize as Δ becomes small. On the other hand, as Δ increases, we should see deviations of $\widehat{\Pi}_{T,\Delta}$ from $\widehat{\Pi}_T$ and hence also from the long-run variance itself. For what values of Δ do these deviations occur?

From numerical simulations across a range of oscillatory processes, we find that the estimator $\widehat{\Pi}_{T,\Delta}$ is quite stable up to about

$$\Delta_{\max} = \frac{T_m}{4}, \tag{5.65}$$

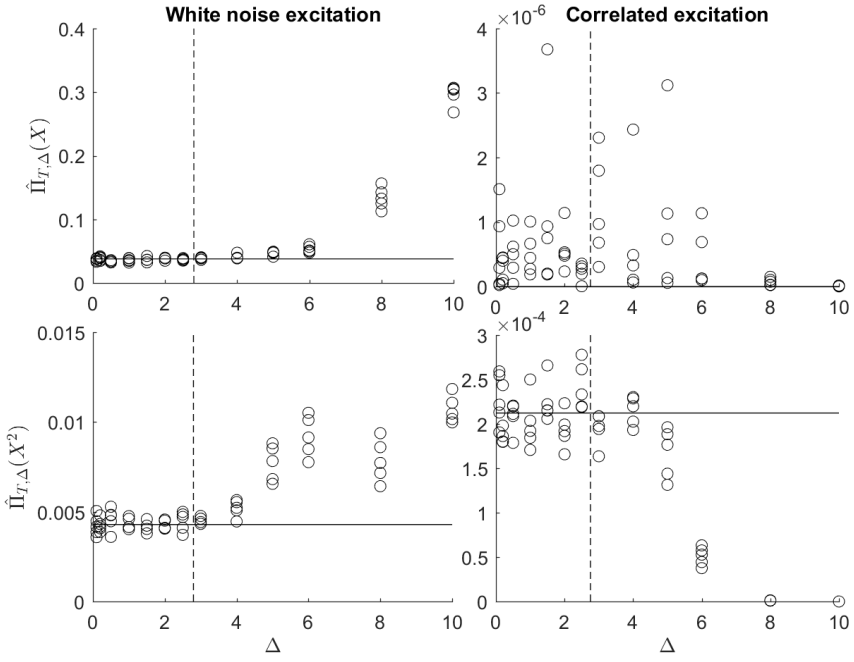

Figure 5.1: Estimated long-run variance $\widehat{\Pi}_{T,\Delta}$ against different choices of Δ. Five estimates are plotted for each Δ, using the data-driven approach with the QS kernel (see Section 5.2). Estimates are made for long-run variance of the original series $\widehat{\Pi}_{T,\Delta}(X)$ (top row), the squared series $\widehat{\Pi}_{T,\Delta}(X^2)$ (bottom row), under both white noise (left column) and correlated excitations (right column). Model parameters are as in Section 5.7.1, and the models are defined in Appendix B. Solid (horizontal) line gives true long-run variance; dashed (vertical) line marks $\frac{T_m}{4}$.

where T_m is the modal period of the oscillation associated with the frequency at which the spectrum is largest. This is illustrated in Figure 5.1 where the estimates $\widehat{\Pi}_{T,\Delta}(X)$ (top row) and $\widehat{\Pi}_{T,\Delta}(X^2)$ (bottom row) are plotted for several realizations of a linear oscillator with white noise excitation (left panel) and correlated excitation (right panel). The model parameters are the same as used in Section 5.7.1 below, and the models are defined in Appendix B. The value of $\Delta_{\max}$ in (5.65) is indicated by a vertical dashed line, and the true value of the long-run variance by a horizontal. (Note that the long-run variance $\Pi(X)$ is zero in the case of correlated excitation, in which case the corresponding plot only serves to show similar variability of the estimates up to $\Delta_{\max}$.)

Note also that $\Delta_{\max}$ in (5.65) is natural in the sense that $\int_{u_0}^{u_0+T_m} \Gamma_X(u)du$ (that is, the integral of $\Gamma_X(u)$ over its one approximate period of oscillation, as part of the long-run variance $\Pi(X)$) is expected to be approximated well enough by the integral discretized at step Δ, as long as $\Delta \leq \Delta_{\max}$. The latter is not meant as a rigorous statement. Note that if $\Delta > \Delta_{\max}$, e.g.

$\Delta = T_m/2$ or T_m, then the discretization of the above integral could "pick up" only e.g. time points where $\Gamma_X(u)$ crosses zero (so that the discrete approximation will no longer be expected to be good).

5.6.5 Choice of time scale

Another practical issue, although not directly related to discretization, is the choice of a time scale. That is, whereas for discrete samples, the index scale is always the set of integers, a time scale for continuous-time process is subject to the practitioner's choice. For example, half an hour of data can be associated with $T = 1,800$ seconds (the time scale of seconds), as well as $T = 1/2$ hour (the time scale of hours). The practitioner should be aware of several implications of the choice of a time scale on the analysis presented in this chapter.

We shall add a subscript, 1 or 2, to the quantities below to refer to the time scale 1 or 2, respectively (e.g. $T_1 = 1,800$ seconds and $T_2 = 1/2$ hours). The key observation here is that the value of the long-run variance $\Pi(X)$ in (5.2) actually depends on the chosen time scale. Indeed, note that

$$\Pi_2(X) = \int_{\mathbb{R}} \Gamma_{2,X}(u_2)du_2 = \frac{T_2}{T_1}\int_{\mathbb{R}} \Gamma_{2,X}\Big(\frac{T_2}{T_1}u_1\Big)du_1 = \frac{T_2}{T_1}\int_{\mathbb{R}} \Gamma_{1,X}(u_1)du_1 =: \frac{T_2}{T_1}\Pi_1(X), \quad (5.66)$$

since the relation between the two time scales is $u_1 = (T_1/T_2)u_2$. But observe also that the time scale does not affect the variance of the sample mean, which would be used in the confidence intervals (see (5.13) and (5.14)), since

$$\frac{\Pi_2(X)}{T_2} = \frac{\Pi_1(X)}{T_1}.$$

The choice of a time scale affects similarly the estimators of the long-run variance. Indeed, note that, by arguing similarly as above,

$$\widehat{\Pi}_{2,T_2}(X) = \int_{-T_2}^{T_2} K\Big(\frac{u_2}{S_{T_2}}\Big)\Gamma_{T_2}(u_2)du_2 = \frac{T_2}{T_1}\int_{-T_1}^{T_1} K\Big(\frac{u_2}{S_{T_2}(T_1/T_2)}\Big)\Gamma_{T_1}(u_1)du_1 = \frac{T_2}{T_1}\widehat{\Pi}_{1,T_1}(X),$$

since $S_{T_2}(T_1/T_2) = S_{T_1}$ for optimal bandwidths $S_{T_1} = S_{\text{opt},1,T_1}$ and $S_{T_2} = S_{\text{opt},2,T_2}$. The latter relation follows by observing similarly that

$$C_{2,\nu}(X) = \frac{\Lambda_{2,\nu}(X)}{\Pi_2(X)} = \frac{(T_2/T_1)^{\nu+1}\Lambda_{1,\nu}(X)}{(T_2/T_1)\Pi_1(X)} = \Big(\frac{T_2}{T_1}\Big)^{\nu} C_{1,\nu}(X)$$

and hence indeed

$$S_{\text{opt},2,T_2} = \Big(\frac{\nu K_\nu^2 C_{2,\nu}^2(X)}{\int K(x)^2 dx} T_2\Big)^{1/(2\nu+1)} = \Big(\frac{\nu K_\nu^2 C_{1,\nu}^2(X)}{\int K(x)^2 dx}\Big(\frac{T_2}{T_1}\Big)^{2\nu+1} T_1\Big)^{1/(2\nu+1)} = \frac{T_2}{T_1} S_{\text{opt},1,T_1}.$$

Remark 5.6.2. The choice of a time scale does, however, affect one aspect of our analysis. The data-driven approach for estimating $\Pi(X)$ and $\Lambda_\nu(X)$ in the constant $C_\nu(X)$ in (5.10) (see the discussion following Remark 5.2.1) involves a preliminary choice of the bandwidth S_T, which was suggested as $\sqrt{T}$. Note that the latter choice depends on the time scale. Moreover, the choice of $\sqrt{T}$ was motivated by the fact that $\sqrt{T}$ should be smaller than T and that $\sqrt{T}/T \to 0$, as $T \to \infty$. A downside is that $\sqrt{T}$ is a meaningless choice in the case when $T < 1$ (e.g. $T = 1/2$ hour). As another possibility which does not have this problem and adapts naturally to the chosen time scale, a preliminary choice of S_T can be determined from the "decorrelation" method. This is an ad hoc method sometimes used in practice with S_T chosen as a time point where the sample autocorrelation function falls below a certain level (see Section 5.7.1 for a more detailed description).

5.7 Numerical results

In this section, we examine the methods proposed in Sections 5.2-5.6 through a simulation study (Section 5.7.1) and applications to real data (Section 5.7.2).

5.7.1 Simulation study

The simulation results presented in this section concern several synthetic processes often used in modeling oscillatory phenomena: a linear oscillator defined by (B.1) and (B.2), and a nonlinear (piecewise linear) oscillator defined by (B.1) and (B.3), in both cases with either a white noise excitation having a spectral density (B.4) or the correlated excitation having

a spectral density (B.5). In the case of a linear oscillator, the parameters used are $\omega_0 = 0.6$, $\delta = 0.09$, and $\sigma = 0.07$ (the white noise excitation), $\sigma = 0.7$, $H_s = 9$, $T_1 = 11.595$ (the correlated excitation). In the case of a piecewise linear oscillator, we take $\omega_0 = 0.6, \delta = 0.09, k_1 = 0.1, x_m = \pi/6$, and the same parameters for the white noise and correlated excitations as in the linear case.

Tables 5.1–5.4 present simulation results for the linear and nonlinear oscillators with white noise and correlated excitations: Table 5.1 concerns the linear oscillator with white noise excitation, Table 5.2 concerns the linear oscillator with correlated excitation, Table 5.3 is for the piecewise linear oscillator with white noise excitation, and Table 5.4 is for the piecewise linear oscillator with correlated excitations. The first column in the tables indicates the length of the record (that is, $T = 50$ or 100 hours), with the associated simulated process sampled at $\Delta = 1/2$ second. The considered record lengths are typical to ship rolling applications. The second column indicates the kernel used in the estimation: QS for the Quadratic Spectral and B for the Bartlett kernel (see Section 5.2). The third column refers to the method used for estimation: "data" for data-driven approach, "model" for model-driven approach, "fixed" for estimation with a fixed bandwidth $S_T = \sqrt{T}$, and "decor." for decorrelation method. The latter is an ad hoc method sometimes used in practice where the bandwidth S_T is chosen as the cutoff point where an envelope of the autocorrelation function of the process crosses the level 0.05 (that is, 5% of the sample autocovariance) for the first time.

The next three columns in the tables present results when estimating the long-run variances $\Pi(X)$, $\Pi(X^2)$, and $\Pi(X, X^2)$. There are two entries in each box associated with a particular estimation scheme: the top entry gives the bias in estimation and the bottom entry gives the standard deviation in estimation, computed from 100 replications. We also note that the true value of $\Pi(X)$, $\Pi(X^2)$, and $\Pi(X, X^2)$ can be computed through explicit formulae in the case of a linear oscillator (see Appendix B), but that this is not the case for a piecewise linear oscillator, in which case we use an estimate from a record of 10,000 hours.

The fifth column of the tables presents the empirical coverage proporitons of the proposed 95% confidence intervals for the mean $\mu(X)$ (the top entry in each box associated with a particular estimation scheme) and the standard deviation $\sigma(X)$ (the bottom entry in each box). As with the long-run variances above, the true standard deviation $\sigma(X)$ can be computed for

Hours	Kernel	Method	$\Pi(X)$	$\Pi(X^2)$	$\Pi(X, X^2)$	Coverage prop.	Interval length
50	QS	data	7.27e-02	1.29e+00	2.38e-01	0.96	1.29e-02
			4.97e-01	1.28e+01	1.88e+00	0.91	1.44e-02
50	QS	model	7.74e-02	-3.41e+00	6.66e-02	0.95	1.29e-02
			1.48e-01	5.45e+00	6.01e-01	0.93	1.42e-02
50	QS	fixed	-7.09e-02	-2.49e+00	-1.84e-01	0.95	1.28e-02
			5.13e-01	1.20e+01	2.03e+00	0.96	1.42e-02
50	QS	decor.	2.90e-01	-1.10e+01	3.48e-02	0.98	1.31e-02
			1.02e-01	4.56e+00	3.84e-01	0.92	1.39e-02
50	B	data	4.96e-01	-1.26e+01	1.49e-01	0.97	1.32e-02
			5.93e-01	1.17e+01	6.00e-01	0.95	1.38e-02
50	B	model	1.89e-01	-5.62e+00	-5.67e-03	0.96	1.30e-02
			1.99e-01	5.85e+00	7.69e-01	0.95	1.41e-02
50	B	fixed	6.59e-02	-2.31e+00	-9.29e-02	0.98	1.29e-02
			4.08e-01	1.06e+01	1.59e+00	0.95	1.42e-02
50	B	decor.	1.00e+00	-2.50e+01	-6.63e-03	0.98	1.36e-02
			1.18e-01	3.86e+00	3.70e-01	0.94	1.33e-02
100	QS	data	7.21e-02	4.34e-01	-1.69e-01	0.93	9.11e-03
			4.40e-01	9.59e+00	1.67e+00	0.93	1.02e-02
100	QS	model	7.10e-02	-2.00e+00	-1.63e-02	0.95	9.12e-03
			1.14e-01	4.04e+00	4.23e-01	0.93	1.01e-02
100	QS	fixed	1.71e-02	-7.73e-02	2.27e-02	0.98	9.08e-03
			3.78e-01	9.48e+00	1.69e+00	0.95	1.01e-02
100	QS	decor.	2.91e-01	-1.09e+01	-5.00e-03	0.96	9.24e-03
			8.37e-02	3.02e+00	2.69e-01	0.93	9.83e-03
100	B	data	3.56e-01	-8.54e+00	2.57e-02	0.96	9.28e-03
			3.44e-01	8.46e+00	4.74e-01	0.98	9.89e-03
100	B	model	1.11e-01	-3.11e+00	5.07e-02	1.00	9.14e-03
			2.07e-01	4.87e+00	5.78e-01	0.96	1.01e-02
100	B	fixed	-2.52e-02	-2.40e+00	6.85e-02	0.97	9.06e-03
			3.58e-01	7.88e+00	1.13e+00	0.96	1.01e-02
100	B	decor.	9.99e-01	-2.60e+01	5.46e-03	0.99	9.64e-03
			7.41e-02	2.33e+00	2.57e-01	0.94	9.38e-03

Table 5.1: Simulation results for the linear oscillator with white noise excitation. See Section 5.7.1 for discussion.

the linear oscillator, but not for the nonlinear oscillator, in which case we use an estimate from a record of 10,000 hours. Finally, the last column of the tables gives the average half length of the corresponding confidence intervals (for the mean on the top, and for the standard deviation on the bottom). We also note that in Table 5.2, the last two rows report on the empirical coverages of the confidence intervals using quantiles, following Section 5.4.

Several conclusions can be drawn from Tables 5.1–5.4. First, the decorrelation method seems to be the worst in general, both in terms of estimating long-run variances and coverages of confidence intervals. Second, the model-driven approach seems to perform best in general in terms of estimating long-run variances, in both linear and nonlinear cases. The performances of the data-driven and fixed approaches are difficult to discern, as are the confidence intervals

Hours	Kernel	Method	$\Pi(X)$	$\Pi(X^2)$	$\Pi(X, X^2)$	Coverage prop.	Interval length
50	QS	data	1.47e-07	1.07e-04	-2.25e-08	1.00	1.57e-06
			1.55e-07	1.72e-05	1.97e-07	0.95	5.38e-04
50	QS	model	1.58e-07	9.83e-05	-4.84e-09	1.00	1.63e-06
			1.42e-07	6.52e-06	8.54e-08	0.96	5.27e-04
50	QS	fixed	1.27e-07	1.10e-04	-1.59e-09	1.00	1.44e-06
			1.26e-07	1.62e-05	1.88e-07	0.94	5.42e-04
50	QS	decor.	1.23e-07	8.73e-05	-1.44e-09	1.00	1.43e-06
			1.24e-07	5.57e-06	4.63e-08	0.95	5.12e-04
50	B	data	5.53e-05	1.04e-04	2.20e-08	1.00	3.43e-05
			2.26e-06	1.40e-05	1.71e-07	0.96	5.35e-04
50	B	model	1.14e-04	1.02e-04	-4.13e-10	1.00	4.94e-05
			1.02e-06	1.10e-05	9.69e-08	0.93	5.31e-04
50	B	fixed	5.72e-05	1.03e-04	1.47e-08	1.00	3.49e-05
			6.44e-07	1.38e-05	1.10e-07	0.95	5.33e-04
50	B	decor.	1.01e-03	6.35e-05	-1.14e-08	1.00	1.47e-04
			3.08e-05	4.52e-06	2.87e-07	0.95	4.80e-04
100	QS	data	7.61e-08	1.09e-04	1.06e-08	1.00	8.29e-07
			6.20e-08	1.19e-05	1.14e-07	0.88	3.82e-04
100	QS	model	6.92e-08	1.00e-04	-1.76e-09	1.00	7.86e-07
			5.23e-08	4.98e-06	2.84e-08	0.89	3.74e-04
100	QS	fixed	6.79e-08	1.07e-04	-1.96e-08	1.00	7.59e-07
			6.99e-08	1.32e-05	1.05e-07	0.90	3.81e-04
100	QS	decor.	6.15e-08	8.64e-05	4.90e-10	1.00	7.12e-07
			6.22e-08	4.60e-06	1.98e-08	0.88	3.62e-04
100	B	data	4.34e-05	1.06e-04	-9.07e-09	1.00	2.15e-05
			1.35e-06	1.01e-05	8.62e-08	0.96	3.80e-04
100	B	model	9.05e-05	1.04e-04	2.29e-10	1.00	3.11e-05
			5.86e-07	8.02e-06	7.00e-08	0.98	3.78e-04
100	B	fixed	4.05e-05	1.05e-04	-5.15e-10	1.00	2.08e-05
			2.90e-07	9.86e-06	1.29e-07	0.93	3.78e-04
100	B	decor.	1.01e-03	6.31e-05	9.26e-09	1.00	1.04e-04
			2.15e-05	4.00e-06	2.23e-07	0.91	3.39e-04
50	-	quantile	-	-	-	0.943	1.70e-08
100	-	quantile	-	-	-	0.955	8.47e-09

Table 5.2: Simulation results for the linear oscillator with correlated excitation. See Section 5.7.1 for discussion.

among all 3 methods: model-driven, data-driven and fixed. Third, as expected, the confidence intervals for the mean $\mu(X)$ have 100% coverage in Table 5.2 – the appropriate shorter confidence intervals based on the quantile method have coverage close to 95%. Fourth, regarding the use of different kernels, estimation using the QS kernel seems generally superior to that for the Bartlett kernel, at least when the model-driven approach is used.

Tables 5.5–5.6 present simulation results in estimation of long-run variance when multiple records are given. Table 5.5 is for a linear oscillator, and Table 5.6 is for a piecewise linear oscillator. The oscillator parameters are the same as those used in Tables 5.1–5.4. For both cases, $R = 10$ records of length $T_r = 5$ hours each are considered. Three different methods of

Hours	Kernel	Method	$\Pi(X)$	$\Pi(X^2)$	$\Pi(X, X^2)$	Coverage prop.	Interval length
50	QS	data	2.84e-05	2.15e-03	-1.22e-04	0.92	9.29e-04
			2.76e-03	3.41e-03	1.73e-03	0.95	2.09e-03
50	QS	model	5.01e-04	-5.76e-04	-4.41e-06	0.97	9.35e-04
			8.56e-04	1.73e-03	8.94e-04	0.96	2.02e-03
50	QS	fixed	3.43e-06	2.52e-03	4.08e-04	0.99	9.29e-04
			2.61e-03	3.52e-03	1.99e-03	0.94	2.10e-03
50	QS	decor.	5.39e-04	-1.64e-04	6.31e-05	0.91	9.36e-04
			8.91e-04	1.76e-03	7.66e-04	0.94	2.03e-03
50	B	data	4.17e-03	-1.49e-03	-1.53e-04	0.97	9.76e-04
			2.73e-03	3.04e-03	8.73e-04	0.94	1.99e-03
50	B	model	1.84e-03	1.22e-04	-8.53e-05	0.94	9.50e-04
			1.42e-03	2.47e-03	1.22e-03	0.95	2.04e-03
50	B	fixed	7.85e-04	2.07e-03	2.21e-05	0.93	9.38e-04
			2.29e-03	3.00e-03	1.62e-03	0.97	2.08e-03
50	B	decor.	6.52e-03	-4.16e-03	-1.44e-05	0.98	1.00e-03
			8.30e-04	1.61e-03	7.75e-04	0.95	1.92e-03
100	QS	data	-1.22e-04	2.79e-03	3.64e-04	0.93	6.56e-04
			2.54e-03	2.71e-03	1.79e-03	0.96	1.49e-03
100	QS	model	4.45e-04	1.57e-04	-5.03e-05	1.00	6.61e-04
			7.17e-04	1.39e-03	5.33e-04	0.98	1.44e-03
100	QS	fixed	1.85e-04	2.14e-03	-9.14e-05	0.88	6.59e-04
			2.01e-03	3.03e-03	1.67e-03	0.94	1.48e-03
100	QS	decor.	5.40e-04	-4.02e-04	-6.01e-05	0.97	6.62e-04
			6.51e-04	1.43e-03	5.87e-04	0.94	1.43e-03
100	B	data	2.78e-03	-3.98e-04	6.14e-05	0.95	6.79e-04
			1.79e-03	2.12e-03	6.34e-04	0.96	1.43e-03
100	B	model	9.84e-04	1.57e-03	-4.11e-05	0.97	6.65e-04
			1.19e-03	1.65e-03	7.90e-04	0.94	1.47e-03
100	B	fixed	5.89e-04	2.63e-03	1.95e-04	0.96	6.62e-04
			1.78e-03	2.19e-03	1.49e-03	0.96	1.48e-03
100	B	decor.	6.71e-03	-4.33e-03	-4.11e-05	0.95	7.10e-04
			5.82e-04	1.23e-03	4.92e-04	0.94	1.35e-03

Table 5.3: Simulation results for the piecewise linear oscillator with white noise excitation. See Section 5.7.1 for discussion.

estimating $\Pi(X)$ (corresponding to X_t in the first column) and $\Pi(X^2)$ (X_t^2 in the first column) are examined: the proposed new ("mean-all"), the average ("separate") and the direct methods (see Section 5.5). The Kernel and Method columns are the same as in Tables 5.1–5.4, except that the decorrelation method is excluded. The entries in each box associated with a particular estimation scheme now indicate the bias (top entry), the standard deviation (middle entry), and the mean-squared error (bottom entry). The entries with the smallest mean-squared errors are indicated in bold.

It can be seen from Tables 5.5–5.6 that the proposed estimator (mean-all) has superior performance in the largest number of cases and always performs better when the model-based

Hours	Kernel	Method	$\Pi(X)$	$\Pi(X^2)$	$\Pi(X, X^2)$	Coverage prop.	Interval length
50	QS	data	-1.43e-03	2.81e-03	1.07e-05	0.94	5.35e-05
			1.59e-05	4.17e-03	1.63e-04	0.93	2.38e-03
50	QS	model	-1.43e-03	-1.64e-03	-1.49e-05	0.98	5.51e-05
			1.27e-05	2.22e-03	1.17e-04	0.93	2.28e-03
50	QS	fixed	-1.43e-03	3.61e-03	1.50e-05	0.91	5.36e-05
			1.64e-05	4.42e-03	1.93e-04	0.90	2.40e-03
50	QS	decor.	-1.43e-03	-8.01e-04	-4.33e-06	0.96	5.49e-05
			1.42e-05	2.15e-03	1.26e-04	0.91	2.30e-03
50	B	data	2.03e-03	-1.19e-03	1.49e-07	1.00	2.75e-04
			8.72e-04	2.69e-03	1.13e-04	0.98	2.28e-03
50	B	model	-3.22e-04	1.14e-03	-3.63e-07	1.00	1.63e-04
			2.26e-05	2.91e-03	1.37e-04	0.98	2.34e-03
50	B	fixed	-7.23e-04	1.83e-03	2.60e-05	1.00	1.34e-04
			2.16e-05	3.58e-03	1.39e-04	0.99	2.36e-03
50	B	decor.	5.29e-03	-6.35e-03	-5.88e-06	1.00	3.82e-04
			2.84e-04	1.99e-03	1.09e-04	0.91	2.16e-03
100	QS	data	-1.43e-03	2.61e-03	1.18e-05	0.94	3.79e-05
			1.16e-05	3.35e-03	1.44e-04	0.95	1.68e-03
100	QS	model	-1.43e-03	-6.35e-04	2.70e-06	0.97	3.89e-05
			9.01e-06	1.43e-03	7.49e-05	0.97	1.63e-03
100	QS	fixed	-1.43e-03	2.19e-03	2.59e-05	0.95	3.79e-05
			1.24e-05	3.74e-03	1.19e-04	0.93	1.67e-03
100	QS	decor.	-1.43e-03	-7.20e-04	2.40e-05	0.96	3.89e-05
			9.19e-06	1.42e-03	8.64e-05	0.91	1.63e-03
100	B	data	1.37e-03	-8.13e-04	-1.46e-06	1.00	1.76e-04
			7.78e-04	1.89e-03	7.93e-05	0.96	1.62e-03
100	B	model	-5.51e-04	1.62e-03	-9.09e-06	1.00	1.04e-04
			1.44e-05	2.35e-03	9.87e-05	0.94	1.66e-03
100	B	fixed	-9.26e-04	2.44e-03	-3.13e-06	1.00	8.27e-05
			1.46e-05	3.07e-03	1.22e-04	0.90	1.68e-03
100	B	decor.	5.27e-03	-6.09e-03	-1.43e-06	1.00	2.70e-04
			1.78e-04	1.26e-03	8.96e-05	0.94	1.53e-03

Table 5.4: Simulation results for the piecewise linear oscillator with correlated excitation. See Section 5.7.1 for discussion.

approach and the QS kernel are used. This approach/kernel was suggested above as superior for single records.

5.7.2 Data applications

We illustrate here the proposed methodology on the data generated by a high-fidelity ship motion simulation code (more specifically, Large Amplitude Motion Program or LAMP of Lin and Yue [61]). The data in question concerns loads at a particular point of a ship. The time plot of the data is depicted in Figure 5.2 (left). The duration of the record is $T = 819$ seconds (with the first 20 seconds discarded), and we consider a sampling rate ranging from $\Delta = 0.02$

Series	Kernel	Method	Separate	Mean-all	Direct
X_t	QS	data	-0.6492	0.0002	1.4624
			0.0410	0.0614	11.1956
			0.4624	**0.0614**	13.3342
X_t	QS	model	-0.7979	-0.2130	2.0293
			0.0087	0.0125	8.9627
			0.6453	**0.0578**	13.0807
X_t	QS	fixed	-0.0885	0.0168	1.6874
			0.0746	0.1122	7.5065
			0.0825	0.1125	10.3537
X_t	B	data	-0.7729	-1.2907	1.9251
			0.3506	0.2105	9.5462
			0.9480	1.8765	13.2522
X_t	B	model	-0.0540	-0.4557	1.5672
			0.0291	0.0191	9.3266
			0.0320	0.2267	11.7828
X_t	B	fixed	0.0735	-0.1111	1.7914
			0.0733	0.0494	8.7393
			0.0787	**0.0618**	11.9486
X_t^2	QS	data	1.8409	0.7258	3.4816
			45.1346	44.7539	8442.4808
			48.5234	**45.2807**	8454.6025
X_t^2	QS	model	11.4759	11.2818	9.0762
			18.6724	18.7036	6326.1594
			150.3685	**145.9823**	6408.5360
X_t^2	QS	fixed	3.4290	1.8819	-6.6153
			66.8617	67.4763	8543.4466
			78.6199	**71.0178**	8587.2089
X_t^2	B	data	33.3687	32.0303	-10.4373
			58.4451	56.7885	6685.3033
			1171.9130	**1082.7290**	6794.2412
X_t^2	B	model	13.8039	13.4828	10.6049
			27.7812	27.8504	5690.6623
			218.3299	**209.6351**	5803.1269
X_t^2	B	fixed	4.5421	3.2672	-12.0358
			44.0204	44.0750	6730.3875
			64.6511	**54.7495**	6875.2490

Table 5.5: Simulation results for multiple records of the linear oscillator with white noise excitation. See Section 5.7.1 for discussion.

to $\Delta = 0.16$ seconds. We are interested in providing confidence intervals for the mean and the standard deviation of the underlying process.

As a first step, we need to decide whether the data points to the degenerate case. As discussed in Section 5.4, the degenerate case is associated with the fact that the process $\int_0^t (X_s - \mu(X))ds$ is stationary (and vice-versa, the non-degenerate case with the process being non-stationary). The time plot of the sample analogue of the process, namely, the discrete version of $\int_0^t (X_s - \overline{X}_T)ds$, is given in Figure 5.2 (right). To see whether the non-degenerate case can be assumed, we postulate it as a null hypothesis to be tested. We use the well-known augmented

Series	Kernel	Method	Separate	Mean-all	Direct
X_t	QS	data	-8.9125e-03	-4.8590e-04	-1.7689e-04
			7.4359e-06	3.3759e-06	3.4772e-04
			8.6869e-05	**3.6120e-06**	3.4775e-04
X_t	QS	model	-1.0305e-02	-1.1530e-03	7.1772e-04
			3.6025e-07	4.5355e-07	2.5875e-04
			1.0655e-04	**1.7830e-06**	2.5927e-04
X_t	QS	fixed	-1.6811e-03	-5.9613e-05	-7.3207e-04
			1.3867e-06	2.1663e-06	3.2782e-04
			4.2126e-06	**2.1699e-06**	3.2835e-04
X_t	B	data	-6.0523e-03	-1.3383e-02	2.6429e-04
			3.1945e-05	2.9260e-05	3.5582e-04
			6.8575e-05	2.0835e-04	3.5589e-04
X_t	B	model	-1.7149e-04	-3.5594e-03	-3.1205e-03
			1.1444e-06	7.8580e-07	4.0284e-04
			1.1738e-06	1.3455e-05	4.1258e-04
X_t	B	fixed	-4.9586e-05	-2.2208e-03	6.2363e-04
			2.2776e-06	1.4943e-06	2.8804e-04
			2.2800e-06	6.4264e-06	2.8843e-04
X_t^2	QS	data	-1.4019e-03	-1.6416e-03	-4.0445e-04
			4.4542e-06	4.5681e-06	4.6088e-04
			6.4195e-06	7.2628e-06	4.6104e-04
X_t^2	QS	model	3.4570e-03	3.3914e-03	-3.9850e-03
			2.1546e-06	2.1500e-06	2.6057e-04
			1.4106e-05	**1.3652e-05**	2.7645e-04
X_t^2	QS	fixed	-1.6560e-03	-2.0158e-03	-5.6502e-03
			6.1992e-06	6.4010e-06	3.9172e-04
			8.9414e-06	1.0464e-05	4.2365e-04
X_t^2	B	data	8.1631e-03	7.7285e-03	-7.1866e-03
			6.3211e-06	6.3417e-06	4.3724e-04
			7.2957e-05	**6.6071e-05**	4.8889e-04
X_t^2	B	model	3.4164e-03	3.3269e-03	1.6838e-03
			3.1475e-06	3.2081e-06	3.4893e-04
			1.4819e-05	**1.4276e-05**	3.5176e-04
X_t^2	B	fixed	-4.1757e-04	-6.8928e-04	-3.2341e-03
			4.0635e-06	4.1038e-06	3.3397e-04
			4.2379e-06	4.5789e-06	3.4443e-04

Table 5.6: Simulation results for multiple records of the piecewise linear oscillator with correlated excitation. See Section 5.7.1 for discussion.

Dickey-Fuller (ADF) test for the null hypothesis – see e.g. Chapter 3 in Pfaff [79]. The test statistic value is -24.3202, and the critical value is -2.58 at $\alpha = 1\%$ (-1.95 at $\alpha = 5\%$). Based on these values, we reject the null hypothesis, i.e. conclude that the data is consistent with the degenerate case.

Proceeding with the methods proposed in the degenerate case, the sample mean of the process with $\Delta = 0.02$ (the smallest available) is 313739.6, and the corresponding 95% confidence intervals for the mean for several choices of Δ are shown in Figure 5.3 (left).

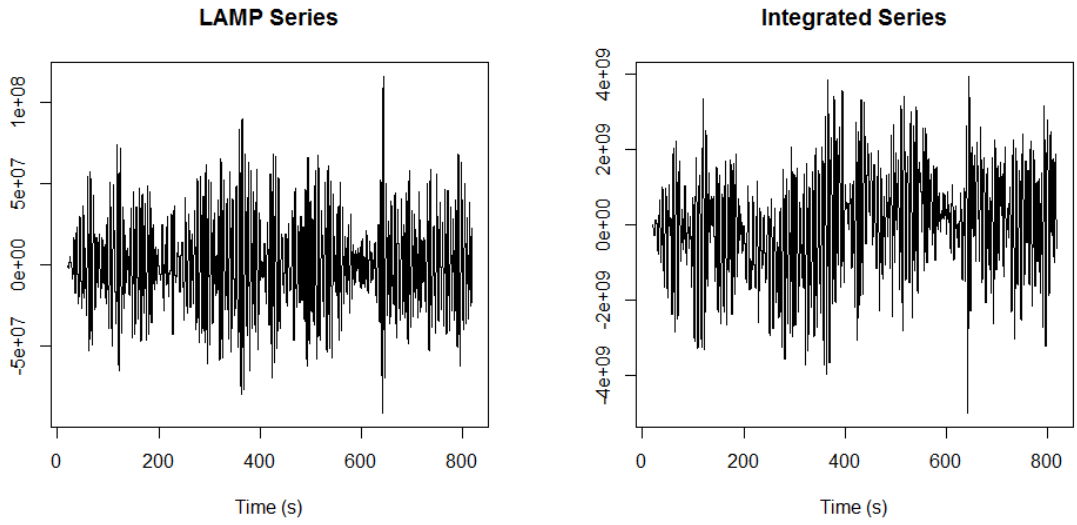

Figure 5.2: Time series data simulated from LAMP. Left: the series itself with $\Delta = 0.02$. Right: the discrete version of $\int_0^t (X_s - \overline{X}_T)ds$.

To construct confidence intervals for the standard deviation for the LAMP series, we need only look at the squared series to estimate its long-run variance (see Section 5.4). The squared series and its integrated series are plotted in Figure 5.4. The ADF test gives a test statistic value of -2.3433; as expected, we consider the squared series to be non-degenerate and proceed with the methods proposed in Section 5.3, using the QS kernel and a fixed bandwidth of $\sqrt{T}$. The sample standard deviation of the process with $\Delta = 0.02$ is 28119779, and the corresponding 95% confidence intervals are shown in Figure 5.3 (right).

5.8 Conclusions

The focus of this chapter has been on inference (i.e. confidence intervals) for the mean and variance of random oscillatory processes by making use of estimation of the long-run variance of the system. We consider processes with both positive long-run variance, as well as the "degenerate case" with zero long-run variance; such analysis requires different estimation techniques. Additionally, we examine the case of multiple independent records of the same stochastic process, from which we develop a more promising method of estimation. We present numerical results of the proposed methods through a simulation study, demonstrating the performance of these estimators.

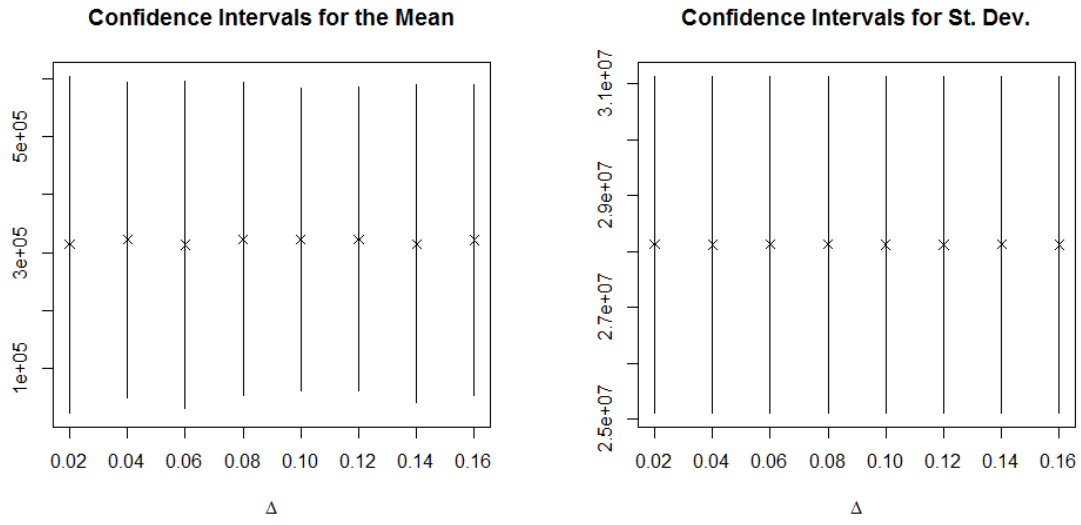

Figure 5.3: Confidence intervals for the mean from the LAMP series for different sampling rates. Left: confidence intervals for the mean. Right: confidence intervals for the standard deviation.

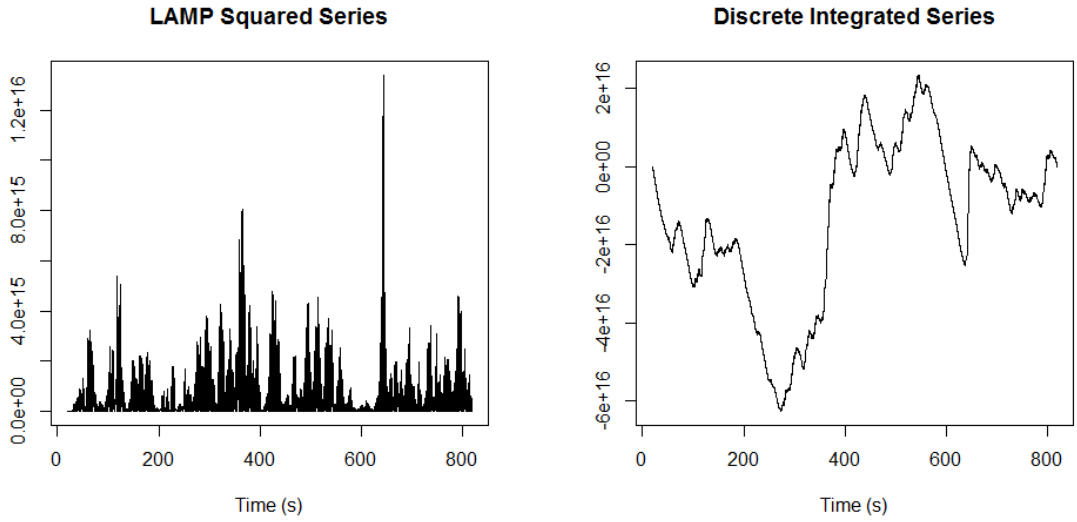

Figure 5.4: Time series data simulated from LAMP and squared. Left: the squared series itself with $\Delta = 0.02$. Right: the discrete version of $\int_0^t (X_s^2 - \overline{X_T^2})ds$.

APPENDIX A: CORRELATED EXCITATION AND ITS SPECTRAL DENSITY

We describe the parameters in use for the single-degree-of-freedom random oscillator given by (3.1). The Bretschneider spectral density given by

$$s_w(\omega) = \frac{A}{\omega^5} e^{-\frac{B}{\omega^4}}, \quad \omega > 0, \tag{A.1}$$

is taken for the wave elevation. Because excitation of waves acts on a ship through the slope, the spectral density of the excitation $y(t)$ is taken as

$$s_y(\omega) = \omega_0^4 \left(\frac{\omega^2}{g}\right)^2 s_w(\omega), \quad \omega > 0, \tag{A.2}$$

where ω is wave frequency. The parameters include $A = 173 H_s^2 T_1^{-4}$ and $B = 691 T_1^{-4}$, where H_s is significant wave height, i.e. twice the amplitude in meters of the highest one-third of waves, and T_1 is the period corresponding to mean frequency of waves; $g = 9.807$ is gravitational acceleration and ω_0 is as in (3.1). The range of ω is trucated to $0.65\omega_{\max} \leq \omega \leq 0.65\omega_{\max} + 2\omega_{\max}$, where $\omega_{\max}$ is the peak frequency. Throughout this chapter, unless stated otherwise, the values taken are $A = 0.775, B = 0.038, H_s = 9, T_1 = 11.595$ and $\omega_{\max} = 0.419$. By convention, the autocovariance function $\gamma(t)$ is related to the spectral density through the Fourier transform as

$$\gamma(t) = \int_0^\infty s(\omega) \cos(\omega t) d\omega, \quad t \in \mathbb{R}. \tag{A.3}$$

We also mention that the spectral density of the solution of the linear system (3.1) with $r(x) = \omega_0^2 x$ is given by

$$s_l(\omega) = \frac{s_g(\omega)}{(\omega_0^2 - \omega^2)^2 + (2\delta\omega)^2} = \frac{\omega_0^4 \left(\frac{\omega^2}{g}\right)^2 \frac{A}{\omega^5} e^{-\frac{B}{\omega^{-4}}}}{(\omega_0^2 - \omega^2)^2 + (2\delta\omega)^2}. \tag{A.4}$$

Hence, the variance $\sigma_{\dot{x}}^2$ can be computed as

$$\int_0^\infty \omega^2 \left(\frac{\omega_0^4 \left(\frac{\omega^2}{g}\right)^2 \frac{A}{\omega^5} e^{-\frac{B}{\omega^{-4}}}}{(\omega_0^2 - \omega^2)^2 + (2\delta\omega)^2} \right) d\omega. \tag{A.5}$$

In (A.4) and (A.5), the damping parameter δ affects the oscillations of the system and the natural frequency parameter ω_0 represents the frequency of oscillations for an undamped ($\delta = 0$) system.

APPENDIX B: OSCILLATORY SYSTEMS OF INTEREST

We describe here several oscillatory processes that are used and referred to in Chapter 5. In the simulations of Section 5.7.1, we consider the *oscillator* X_t satisfying the general equation

$$\ddot{X}_t + 2\delta\dot{X}_t + r(X_t) = Z_t. \tag{B.1}$$

Here, $\delta > 0$ is the damping parameter, $r(x)$ is the restoring force and Z_t is an external excitation. A *linear oscillator* corresponds to the linear restoring function

$$r(x) = w_0^2 x, \tag{B.2}$$

where w_0 is the natural frequency parameter. A *nonlinear oscillator* is associated with a nonlinear restoring force. In Section 5.7.1, we use a *piecewise linear oscillator* with the restoring force

$$r(x) = \begin{cases} -k_1 w_0^2(x + x_m) - w_0^2 x_m, & x < -x_m, \\ w_0^2 x, & -x_m < x < x_m, \\ -k_1 w_0^2(x - x_m) + w_0^2 x_m, & x > x_m, \end{cases} \tag{B.3}$$

that is, where the restoring force has a negative slope ($-k_1 w_0^2$) in the nonlinear regime $|x| > x_m$. Though other nonlinear oscillators (e.g. the Duffing oscillator) could be considered as well.

Two forms of the external excitation are considered. First, there is a white noise excitation $Z_t = \sigma\dot{W}_t$, where W_t is a standard Brownian motion and $\sigma > 0$ is the parameter controlling the strength of the excitation. Its spectral density can be thought as

$$S_Z(w) = \sigma^2. \tag{B.4}$$

Second, motivated by ship rolling applications, we also consider the stationary Gaussian excitation having the spectral density

$$S_Z(w) = w_0^4 \Big(\frac{w^2}{g}\Big)^2 \frac{A}{w^5} e^{-\frac{B}{w^4}}, \quad w > 0, \tag{B.5}$$

where w_0 is as in (B.2) and (B.3) and $g = 9.807$ is the gravitational acceleration. The parameters A and B are taken as $A = 173H_s^2T_1^{-4}$ and $B = 691T_1^{-4}$, where H_s is significant wave height, i.e. twice the amplitude in meters of the highest one-third of waves, and T_1 is the period corresponding to the mean frequency of waves.

We also note that the spectral density of the linear oscillator with excitation Z is given by

$$S_X(w) = \frac{S_Z(w)}{(w_0^2 - w^2)^2 + (2\delta w)^2}. \tag{B.6}$$

Since $S_Z(0) = 0$ for the correlated excitation with the spectral density (B.5), we have $S_X(0) = 0$ for the linear oscillator with the correlated excitation. Thus, in view of (5.23), this oscillator falls into the degenerate case.

A convenient fact about a linear oscillator with white noise excitation is that its auto-covariance function can be computed explicitly as follows. In view of (5.12) and (B.6), we have

$$\Gamma_X(h) = \frac{1}{2\pi}\int_{\mathbb{R}} \cos(hw) S_X(w) dw = \frac{\sigma^2}{\pi}\int_0^\infty \frac{\cos(hw)}{(w_0^2 - w^2)^2 + (2\delta w)^2} dw$$

$$= \frac{\sigma^2}{\pi}\int_0^\infty \frac{\cos(hw)}{w^4 + 2w_0^2 w^2(2(\delta/w_0)^2 - 1) + w_0^4} dw = \frac{\sigma^2}{\pi}\int_0^\infty \frac{\cos(hw)}{w^4 + 2w_0^2 w^2\cos(2t) + w_0^4} dw,$$

where $\cos(2t) = 2(\delta/w_0)^2 - 1$ with $0 < t < \pi/2$, assuming that

$$\delta < w_0.$$

Then, by using Formula 3.733.1 in Gradshteyn and Ryzhik [46], p. 428,

$$\Gamma_X(h) = \frac{\sigma^2}{2w_0^3 \sin(2t)} e^{-hw_0\cos(t)} \sin(t + hw_0 \sin(t))$$

$$= \frac{\sigma^2}{4w_0\delta\sqrt{w_0^2 - \delta^2}} e^{-h\delta} \sin(t + h\sqrt{w_0^2 - \delta^2}) \tag{B.7}$$

by using the facts that $w_0\cos(t) = \delta$ and $w_0 \sin(t) = \sqrt{w_0^2 - \delta^2}$. This also allows one to compute explicitly the long-run variance of the process as: by using Formula 3.893.1 in Gradshteyn and

Ryzhik [46], p. 486,

$$\Pi(X) = \int_0^\infty \Gamma_X(h)dh = \frac{\sigma^2}{w_0^3 \sin(2t)} \int_0^\infty e^{-hw_0 \cos(t)} \sin(t + hw_0 \sin(t))dh$$

$$= \frac{\sigma^2}{w_0^3 \sin(2t)} \cdot \frac{1}{w_0^2 \cos^2(t) + w_0^2 \sin^2(t)} \Big(w_0 \sin(t)\cos(t) + w_0 \cos(t)\sin(t) \Big) = \frac{\sigma^2}{w_0^4}. \tag{B.8}$$

Note that this is also the same as $S_X(0)$. One can also compute $\Lambda_2(X)$ in (5.10)–(5.11) explicitly (see Remark 5.2.1) as

$$\Lambda_2(X) = (-1)\frac{d^2 S_X(\omega)}{d\omega^2}\Big|_{w=0} = \frac{4\sigma^2(2\delta^2 - \omega_0^2)}{\omega_0^8}. \tag{B.9}$$

In the Gaussian case which we assume for a linear oscillator with white or uncorrelated excitation, long-run variances associated with the squares of the process can be calculated by using the well-known relations

$$\Gamma_{X^2}(h) = 2(\Gamma_X(h))^2, \quad \Gamma_{X,X^2}(h) = 0. \tag{B.10}$$

For nonlinear oscillators (with any excitation), no explicit formulae are known for either the spectral density or the autocovariance function.

BIBLIOGRAPHY

[1] I. B. Aban, M. M. Meerschaert, and A. K. Panorska. Parameter estimation for the truncated Pareto distribution. *Journal of the American Statistical Association*, 101(473):270–277, 2006.

[2] P. A. Anastopoulos, K. J. Spyrou, C. C. Bassler, and V. Belenky. Towards an improved critical wave groups method for the probabilistic assessment of large ship motions in irregular sea. *Probabilistic Engineering Mechanics*, 44:18–27, 2016.

[3] D. W. K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59(3):817–858, 1991.

[4] A. A. Andronov, A. A. Vitt, and S. E. Khaikin. *Theory of Oscillators*. Pergamon Press, 1966.

[5] A. V. Asimit, D. Li, and L. Peng. Pitfalls in using Weibull tailed distributions. *Journal of Statistical Planning and Inference*, 140(7):2018–2024, 2010.

[6] J. Beirlant, I. Fraga Alves, and I. Gomes. Tail fitting for truncated and non-truncated Pareto-type distributions. Preprint, 2015.

[7] J. Beirlant, Y. Goegebeur, J. Teugels, and J. Segers. *Statistics of Extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2004. Theory and applications, with contributions from Daniel De Waal and Chris Ferro.

[8] V. Belenky. A capsizing probability computation method. *Journal of Ship Research*, 37(3):200–207, 1993.

[9] V. Belenky. Piecewise linear approach to nonlinear ship dynamics. In D. Vassalos, M. Hamamoto, A. Papanikolaou, and D. Molyneux, editors, *Contemporary Ideas on Ship Stability*, pages 149–160. Elsevier, 2000.

[10] V. Belenky and B. Campbell. Evaluation of the exceedance rate of a stationary stochastic process by statistical extrapolation using the envelope peaks over threshold (EPOT) method. Hydromechanics Department Report NSWCCD-50-TR-2011/032, Naval Surface Warfare Center, Carderock Division, West Bethesda, MD, 2011.

[11] V. Belenky and N. B. Sevastianov. *Stability and Safety of Ships: Risk of Capsizing*. Society of Naval Architects and Marine Engineers, 2nd edition, 2007.

[12] V. Belenky and K. Weems. Probabilistic properties of parametric roll. In T. I. Fossen and H. Nijmeijer, editors, *Parametric Resonance in Dynamical Systems*, pages 129–146. Springer, 2011.

[13] V. Belenky, K. M. Weems, C .C. Bassler, M. J. Dipper, B. Campbell, and K. Spyrou. Approaches to rare events in stochastic dynamics of ships. *Probabilistic Engineering Mechanics*, 28:30–38, 2012.

[14] V. Belenky, K. M. Weems, and W. M. Lin. Numerical procedure for evaluation of capsizing probability with split time method. In *27th Symposium on Naval Hydrodynamics*, 2008. Seoul, Korea.

[15] V. Belenky, K. M. Weems, and W. M. Lin. Numerical procedure for evaluation of capsizing probability with split time method. In *27th Symposium on Naval Hydrodynamics*, 2008. Seoul, Korea.

[16] V. Belenky, K. M. Weems, and W. M. Lin. Split-time method for estimation of probability of capsizing caused by pure loss of stability. *Ocean Engineering*, 122:333–343, 2016.

[17] V. Belenky, K. M. Weems, K. Spyrou, V. Pipiras, and T. Sapsis. Modeling broaching-to and capsizing with extreme value theory. In *Proceedings of the 16th International Ship Stability Workshop*, 2017. Belgrade, Serbia.

[18] H. Benford. *Naval Architecture for Non-Naval Architects*. Society of Naval Architects and Marine Engineers, 1991.

[19] R. C. Bishop, W. Belknap, C. Turner, B. Simon, and J. H. Kim. Parametric investigation on the influence of GM, roll damping, and above-water form on the roll response of model 5613. Report NSWCCD-50-TR-2005/027, David Taylor Model Basin –- NSWCCD, 2005.

[20] B. Campbell, V. Belenky, and V Pipiras. Properties of the tail of envelope peaks and their use for the prediction of the probability of exceedance for ship motions in irregular waves. In P. Spanos and G. Deodatis, editors, *CSM7 Computational Stochastic Mechanics*, Santorini, Greece, 2014.

[21] B. Campbell, V. Belenky, and V. Pipiras. Application of the envelope peaks over threshold (EPOT) method for probabilistic assessment of dynamic stability. *Ocean Engineering*, 120:298–304, 2016.

[22] W. Chai, A. Naess, and B. J. Leira. Stochastic nonlinear ship rolling in random beam seas by the path integration method. *Probabilistic Engineering Mechanics*, 44:43–52, 2016.

[23] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics. Springer-Verlag London Ltd., London, 2001.

[24] A. C. Davison and R. L. Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society. Series B. Methodological*, 52(3), 1990. With discussion and a reply by the authors.

[25] M. de Carvalho, K. F. Turkman, and A. Rua. Dynamic threshold modelling and the US business cycle. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, 62(4):535–550, 2013.

[26] L. de Haan and J. de Ronde. Sea and wind: multivariate extremes at work. *Extremes*, 1(1):7–45, 1998.

[27] L. De Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Science & Business Media, 2007.

[28] L. de Haan and A. K. Sinha. Estimating the probability of a rare event. *The Annals of Statistics*, 27(2):732–759, 1999.

[29] P. de Zea Bermudez and S. Kotz. Parameter estimation of the generalized Pareto distribution. I. *Journal of Statistical Planning and Inference*, 140(6):1353–1373, 2010.

[30] P. de Zea Bermudez and S. Kotz. Parameter estimation of the generalized Pareto distribution. II. *Journal of Statistical Planning and Inference*, 140(6):1374–1388, 2010.

[31] A. Degtyarev and I. Gankevich. Hydrodynamic pressure under sea surface on basis of autoregressive model of irregular waves. *Physics of Particles and Nuclei Letters*, 12(3):389–391, 2015.

[32] A. Degtyarev and A. M. Reed. Synoptic and short-term modeling of ocean waves. *International Shipbuilding Progress*, 60(1-4):523–553, 2013.

[33] J. Diebolt, L. Gardes, S. Girard, and A. Guillou. Bias-reduced estimators of the Weibull tail-coefficient. *Test*, 17(2):311–331, 2008.

[34] D. Draghicescu and R. Ignaccolo. Modeling threshold exceedance probabilities of spatially correlated time series. *Electronic Journal of Statistics*, 3:149–164, 2009.

[35] H. Drees and L. de Haan. Estimating failure probabilities. Preprint, 2014.

[36] S. T. Ekanayake and J. F. Cruise. Comparisons of Weibull-and exponential-based partial duration stochastic flood models. *Stochastic Hydrology and Hydraulics*, 7(4):283–297, 1993.

[37] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extreme Events*. Springer-Verlag, New York, 1997.

[38] H. Fan, T. Soderstrom, M. Mossberg, B. Carlsson, and Y. Zou. Estimation of continuous-time AR process parameters from discrete-time data. *IEEE Transactions on Signal Processing*, 47(5):1232–1244, 1999.

[39] A. Ferreira and L. de Haan. The generalized Pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737, 2014.

[40] L. Gardes and S. Girard. Estimating extreme quantiles of Weibull tail distributions. *Communications in Statistics—Theory and Methods*, 34(5):1065–1080, 2005.

[41] L. Gardes, S. Girard, and A. Guillou. Weibull tail-distributions revisited: a new look at some tail estimators. *Journal of Statistical Planning and Inference*, 141(1):429–444, 2011.

[42] M. Gitterman. *The Noisy Oscillator: The First Hundred Years, From Einstein Until Now*. World Scientific, 2005.

[43] D. Glotzer and V. Pipiras. Statistical perspectives on some problems arising in Naval Engineering. In *The 30th American Towing Tank Conference*, West Bethesda, MD, USA, 2017.

[44] D. Glotzer, V. Pipiras, V. Belenky, B. Campbell, and T. Smith. Confidence intervals for exceedance probabilities with application to extreme ship motions. *REVSTAT Statistical Journal*, 15(4):537–563, 2017.

[45] M. I. Gomes and A. Guillou. Extreme value theory and statistics of univariate extremes: a review. *International Statistical Review*, 83(2):263–292, 2015.

[46] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Academic press, 2014.

[47] A. Guha, A. Somayajula, and J. Falzarano. Analysis of causeway ferry dynamics for safe operation of improved navy lighterage system. In *13th International Ship Stability Workshop*, Brest, France, 2013.

[48] C. Hayashi. *Nonlinear Oscillations in Physical Systems*. Princeton University Press, 2014.

[49] J. E. Heffernan and J. A. Tawn. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 66(3):497–546, 2004. With discussions and reply by the authors.

[50] Y. Hong, W. Q. Meeker, and L. A. Escobar. The relationship between confidence intervals for failure probabilities and life time quantiles. *Reliability, IEEE Transactions on*, 57(2):260–266, June 2008.

[51] J. R. M. Hosking and J. R. Wallis. Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*, 29(3):339–349, 1987.

[52] H. Kirshner, M. Unser, and J. P. Ward. On the unique identification of continuous-time autoregressive models from sampled data. *IEEE Transactions on Signal Processing*, 62(6):1361–1376, 2014.

[53] I. A. Kougioumtzoglou and P. D. Spanos. Stochastic response analysis of the softening duffing oscillator and ship capsizing probability determination via a numerical path integral approach. *Probabilistic Engineering Mechanics*, 35:67–74, jan 2014.

[54] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. Springer-Verlag, New York-Berlin, 1983.

[55] J. Lee. Long-run variance estimation for linear processes under possible degeneracy. *Journal of Economic Theory and Econometrics*, 21(1), 2010.

[56] E. V. Lewis, editor. *Principles of Naval Architecture*. SNAME, 1989.

[57] E. V. Lewis and Society of Naval Architects and Marine Engineers (U.S.). *Principles of Naval Architecture: Volume III Motions in Waves and Controllability*. Principles of Naval Architecture. Society of Naval Architects and Marine Engineers, 1990.

[58] Z. Liang and G. C. Lee. *Random Vibration: Mechanical, Structural, and Earthquake Engineering Applications*. CRC Press/Taylor & Francis Group, 2015.

[59] N. Lin and S. V. Lototsky. Undamped harmonic oscillator driven by additive Gaussian white noise: a statistical analysis. *Communications on Stochastic Analysis*, 5(1):233–250, 2011.

[60] N. Lin and S. V. Lototsky. Second-order continuous-time non-stationary Gaussian autoregression. *Statistical Inference for Stochastic Processes*, 17(1):19–49, 2014.

[61] W. M. Lin and D. K. P. Yue. Numerical solutions for large amplitude ship motions in the time-domain. In *Proceedings of the 18th Symposium on Naval Hydrodynamics*, pages 41–66, Ann Arbor, 1991.

[62] G. Lindgren. *Stationary Stochastic Processes*. Chapman & Hall/CRC Texts in Statistical Science Series. CRC Press, Boca Raton, FL, 2013. Theory and applications.

[63] M. S. Longuet-Higgins. On the motions of ships in confused seas. *Philosophical Transactions of the Royal Society of London, Ser. A, Mathematical and Physical Sciences*, 249(966):321–387, 1962.

[64] Y. Lu and J. Y. Park. Estimation of longrun variance of continuous time stochastic process using discrete sample. Preprint, 2014.

[65] D. Madsen, H. Rosbjerg and P. Harremoës. Application of the partial duration series approach in the analysis of extreme rainfalls. In *Proceedings of the Yokohama Symposium*. The International Association of Hydrological Sciences, 1993.

[66] J. Mager. Automatic threshold selection of the peaks over threshold method. 2015. Master's Thesis, the Technical University of Munich.

[67] G. Malara, P. D. Spanos, and F. Arena. Maximum roll angle estimation of a ship in confused sea via quasi-deterministic approach. *Probabilistic Engineering Mechanics*, 35:75–81, 2014.

[68] K. A. McTaggart. Ship capsize risk in a seaway using fitted distributions to roll maxima. *Journal of Offshore Mechanics and Arctic Engineering*, 122(2):141–146, 2000.

[69] C. Mercadier and P. Soulier. Optimal rates of convergence in the Weibull model based on kernel-type estimators. *Statistics & Probability Letters*, 82(3):548–556, 2012.

[70] M. A. Mohamad, W. Cousins, and T. P. Sapsis. A probabilistic decomposition-synthesis method for the quantification of rare events due to internal instabilities. *Journal of Computational Physics*, 322:288–308, 2016.

[71] M. A. Mohamad and T. P. Sapsis. Probabilistic description of extreme events in intermittently unstable systems excited by correlated stochastic processes. *SIAM/ASA Journal of Uncertainty Quantification*, 3:709–736, 2015.

[72] M. A. Mohamad and T. P. Sapsis. Probabilistic response and rare events in Mathieu's equation under correlated parametric excitation. *Ocean Engineering*, 120:289–297, 2016.

[73] A. H. Nayfeh and D. T. Mook. *Nonlinear Oscillations*. John Wiley & Sons, 2008.

[74] J. I. Neimark and P. S. Landa. *Stochastic and Chaotic Oscillations*, volume 77. Springer Science & Business Media, 2012.

[75] C. Neves and M. I. Fraga Alves. Reiss and Thomas' automatic selection of the number of extremes. *Computational Statistics & Data Analysis*, 47(4):689–704, 2004.

[76] M. A. S. Neves, V. Belenky, J. O. de Kat, K. Spyrou, and N. Umeda. *Contemporary Ideas on Ship Stability and Capsizing in Waves*. Fluid Mechanics And Its Applications. Springer, 2011.

[77] D. Paroka and N. Umeda. Capsizing probability prediction of the large passenger ship in irregular beam wind and waves: Comparison of analytical and numerical methods. *Journal of Ship Research*, 50(4):371–377, 2006.

[78] E. Parzen. On consistent estimates of the spectrum of a stationary time series. *Annals of Mathematical Statistics*, 28:329–348, 1957.

[79] B. Pfaff. *Analysis of Integrated and Cointegrated Time Series with R.* Springer Science & Business Media, 2008.

[80] D. T. Pham. Estimation of continuous-time autoregressive model from finely sampled data. *IEEE Transactions on Signal Processing*, 48(9):2576–2584, 2000.

[81] J. Pickands. Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131, 1975.

[82] D. N. Politis and J. P. Romano. The stationary bootstrap. *Journal of the American Statistical Association*, 89(428):1303–1313, 1994.

[83] A. M. Reed. Discussion to paper "Numerical procedure for evaluation of capsizing probability with split time method". In *27th Symposium on Naval Hydrodynamics*, 2008. Seoul, Korea.

[84] A. M. Reed, R. Beck, and W. Belknap. Advances in predictive capability of ship dynamics in waves. In *Proceedings of the 30th Symposium of Naval Hydrodynamics*, Hobart, November 2014.

[85] A. M. Reed and A. Zuzick. Direct assessment will require accreditation – what this means. In *12th International Conference on the Stability of Ships and Ocean Vehicles*, Glasgow, UK, 2015.

[86] R. D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values With Applications To Insurance, Finance, Hydrology and Other Fields.* Birkhäuser Verlag, Basel, third edition, 2007.

[87] D. Rosbjerg, H. Madsen, and P. F. Rasmussen. Prediction in partial duration series with generalized Pareto-distributed exceedances. *Water Resources Research*, 28(11):3001–3010, 1992.

[88] C. Scarrott and A. MacDonald. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT*, 10(1):33–60, 2012.

[89] X. Shao. Self-normalization for time series: a review of recent developments. *Journal of the American Statistical Society*, 110(512):1797–1817, 2015.

[90] R. L. Smith. Estimating tails of probability distributions. *The Annals of Statistics*, 15:1174–1207, 1987.

[91] R. L. Smith and T. S. Shively. Point process approach to modeling trends in tropospheric ozone based on exceedances of a high threshold. *Atmospheric Environment*, 29(23):3489–3499, 1995.

[92] R. L. Smith, J. A. Tawn, and S. G. Coles. Markov chain models for threshold exceedances. *Biometrika*, 84(2):249–268, 1997.

[93] T. Smith and A. Zuzick. Validation of statistical extrapolation methods for large motion prediction. In *12th International Conference on the Stability of Ships and Ocean Vehicles*, Glasgow, UK, 2015.

[94] K. Sobczyk. *Stochastic Differential Equations.* Kluwer Academic Publishers, Dordrecht, The Netherlands, 1991.

[95] T. Soderstrom, H. Fan, B. Carlsson, and S. Bigi. Least squares parameter estimation of continuous-time ARX models from discrete-time data. *IEEE Transactions on Automatic Control*, 42(5):659–673, 1997.

[96] J. Sólnes. *Stochastic Processes and Random Vibrations: Theory and Practice*. Wiley, 1997.

[97] P. D. Spanos. ARMA algorithms for ocean wave modeling. *Journal of Energy Resources Technology*, 105:300–309, 1983.

[98] J. Spinelli and M. A. Stephens. Tests for exponentiality when origin and scale parameters are unknown. *Technometrics*, 29(4):471–476, 1987.

[99] M. St. Denis and W. J. Pierson. On the motions of ships in confused seas. In *SNAME Transactions*, 1953.

[100] N. Tajvidi. Confidence intervals and accuracy estimation for heavy-tailed generalized Pareto distributions. *Extremes*, 6(2):111–123 (2004), 2003.

[101] N. H. Timm. *Applied Multivariate Analysis*. Springer, 2002.

[102] P. Todorovic and E. Zelenhasic. A stochastic model for flood analysis. *Water Resources Research*, 6(6):1641 – 1648, 1970.

[103] K. Weems, A. M. Reed, A. Degtyarev, and I. Gankevich. Implementation of an autoregressive wave model in a numerical simulation code. In *31st Symposium on Naval Hydrodynamics*, Monterey, CA, USA, 2016.

[104] K. Weems and D. Wundrow. Hybrid models for fast time-domain simulation of stability failures in irregular waves with volume-based calculations for Froude-Krylov and hydrostatic forces. In *Proceedings of the 13th International Ship Stability Workshop*, Brest, September 2013.