

PREDICTABILITY OF EXTREME EVENTS IN TIME SERIES

Dissertation

zur Erlangung des Doktorgrades
des Fachbereichs Mathematik und Naturwissenschaften
(Fachgruppe Physik)
Doctor rerum naturalium (Dr. rer. nat.)
der Bergischen Universität Wuppertal

WUB-DIS 2008-05

vorgelegt von

Sarah Hallerberg

geboren am 28.12.1979 in Bad Oeynhausen

JUNI 2008

erstellt am

MAX-PLANCK-INSTITUT FÜR PHYSIK KOMPLEXER SYSTEME
DRESDEN

Diese Dissertation kann wie folgt zitiert werden:

urn:nbn:de:hbz:468-20080438

[<http://nbn-resolving.de/urn/resolver.pl?urn=urn%3Anbn%3Ade%3Ahbz%3A468-20080438>]

Abstract

In this thesis we assess the prediction of extreme events observing precursory structures, which were identified using a maximum likelihood approach. The main goal of this thesis is to investigate the dependence of the quality of a prediction on the magnitude of the events under study. Until now, this dependence was only sporadically reported for different phenomena without being understood as a general feature of predictions. We propose the *magnitude dependence* as a property of a prediction, indicating, whether larger events can be better, harder or equally well predicted than smaller events. Furthermore we specify a condition which can characterize the magnitude dependence of a distinguished measure for the quality of a prediction, the Receiver Operator characteristic curve (ROC). This test condition allows to relate the magnitude dependence of a prediction task to the joint PDF of events and precursory variables. If we are able to describe the numerical estimate of this joint PDF by an analytic expression, we can not only characterize the magnitude dependence of events observed so far, but infer the magnitude dependence of events, larger than the observed events. Having the test condition specified, we study the magnitude dependence for the prediction of increments and threshold crossings in sequences of random variables and short- and long-range correlated stochastic processes. In dependence of the distribution of the process under study we obtain different magnitude dependences for the prediction of increments in Gaussian, exponentially, symmetrized exponentially, power-law and symmetrized power-law distributed processes. For threshold crossings we obtain the same magnitude dependence for all distributions studied. Furthermore we study the dependence on the event magnitude for the prediction of increments and threshold crossings in velocity increments, measured in a free jet flow and in wind-speed measurements. Additionally we introduce a method of post-processing the output of ensemble weather forecast models in order to identify precursory behavior, which could indicate failures of weather forecasts. We then study not only the success of this method, but also the magnitude dependence.

keywords: extreme events, statistical inference, prediction via precursors, ROC curves, likelihood ratio, magnitude dependence

Zusammenfassung

In dieser Arbeit untersuchen wir die Vorhersagen von Extremereignissen in Zeitreihen, welche mittels Beobachtung von Vorläuferstrukturen gemacht werden. Geeignete Vorläuferstrukturen wählen wir mit Hilfe von bedingten Wahrscheinlichkeitsverteilungen (Maximum Likelihood Ansatz) aus. Das Hauptanliegen dieser Arbeit ist es, die Abhängigkeit der Vorhersagen von der Ereignisgröße zu untersuchen. Bisher wurde sporadisch für einzelne Phänomene beschrieben, daß größere Ereignisse besser vorhersagbar sind, als kleinere, ohne dies jedoch als eine allgemeine Eigenschaft von Vorhersagen zu interpretieren. Wir führen somit die *Ereignisgrößenabhängigkeit* als Eigenschaft einer Vorhersage ein. Die Ereignisgrößenabhängigkeit kann positiv, negativ oder null sein und somit angeben, ob größere und somit extremere Ereignisse, besser, schlechter oder gleichgut vorherzusagen sind, wie kleinere. Des Weiteren geben wir eine Bedingung an, welche die Ereignisgrößenabhängigkeit im Bezug auf ein bestimmtes Maß für die Qualität von Vorhersagen, die *Receiver Operator Characteristic*, angibt. Besagte Bedingung erlaubt es, die Ereignisgrößenabhängigkeit auf die Verbundverteilung von Ereignis und Vorläuferstruktur zurückzuführen. Sofern wir in der Lage sind, diese Verteilung analytisch zu beschreiben, können wir aus bisherigen Beobachtungen die Größenabhängigkeit von Ereignissen einschätzen, die größer sind, als alle bisher beobachteten Ereignisse. Dabei setzen wir voraus, daß extreme und nicht extreme Ereignisse der gleichen Wahrscheinlichkeitsverteilung folgen. Mit Hilfe der Testbedingung untersuchen wir nun die Ereignisgrößenabhängigkeit der Vorhersagen von Inkrementen und Schwellwertüberschreitungen in Folgen von Zufallszahlen, sowie in kurz- und langreichweitig korrelierten stochastischen Prozessen. In Abhängigkeit von der Verteilung des untersuchten Prozesses erhalten wir unterschiedliche Ergebnisse für die Ereignisgrößenabhängigkeit der Vorhersagen von Inkrementen. Die Vorhersagen von Schwellwertüberschreitungen zeigten jedoch für alle untersuchten Verteilungen qualitativ die gleiche Ereignisgrößenabhängigkeit. Des Weiteren haben wir Vorhersagen von Inkrementen und Schwellwertüberschreitungen in Zeitreihen aus Geschwindigkeitsinkrementen, die in einem Freistrahlexperiment und in bodennahem Wind gemessen wurden, untersucht. Zusätzlich haben wir auch eine Methode zur Vorhersage von abweichenden Wettervorhersagen entwickelt. Diese Methode konstruiert aus Ensemblewettervorhersagen Indikatoren für eine große Abweichung zwischen der vorhergesagten und der beobachteten Temperatur. Außerdem wurde auch an den Vorhersagen von abweichenden Temperaturvorhersagen die Ereignisgrößenabhängigkeit untersucht.

Stichwörter: Extremereignisse, Vorhersagen mittels Vorläuferstrukturen, Receiver Operator Charakteristik, Maximum-Likelihood-Ansatz, Ereignisgrößenabhängigkeit

Acknowledgments

First of all I would like to thank Holger Kantz for suggesting the idea to study the magnitude dependence, for innumerable discussions, advice, support and encouragement during his supervision of my thesis.

Additionally I have benefited from various discussions about the contents of this thesis and other topics. Therefore I would like to thank Eduardo Goldani Altmann, Anja Riegert, Alexander Croy, Detlef Holstein, Sara Pinto Garcia and Friedrich Lenz.

I am especially grateful to Jochen Bröcker for his suggestions concerning the precursors for large failures of weather forecast and for pointing out the importance of the Neyman- Pearson Lemma.

I am also grateful to the group of Joachim Peinke at the University of Oldenburg, especially to Julia Gotschall, for providing us with their excellent data from the free-jet experiment and their wind speed measurements.

The weather predictions and the corresponding verifications were provided from the European Center for Medium Range Weather forecasts and they were made accessible to me due to a collaboration with Leonard A. Smith and his group in the Center for the Analysis of Time Series at the London School of Economics. I would therefore like to use the opportunity to thank also the members of this group for their support.

Furthermore I would like to thank Astrid S. de Wijn, Anja Riegert and Sara Pinto Garcia for their moral support, encouragement and friendship.

Many thanks also to all the colleagues and staff of the Max Planck Institute for the Physics of Complex Systems for providing a friendly working environment and for their help in many practical issues.

I am also extremely grateful to my parents for their help and support in many practical issues, especially for taking care of moving my possessions and refurbishing my flat these days, so that I could concentrate on completing this thesis.

Finally I would like to thank the Max Planck Society for funding this thesis and the DAAD for funding the collaboration with the Center for the Analysis of Time Series in London.

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland, noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die vorliegende Dissertation wurde am Max-Planck-Institut für Physik komplexer Systeme unter der wissenschaftlichen Betreuung von Prof. Dr. Holger Kantz angefertigt.

Es haben keine früheren Promotionsverfahren von mir stattgefunden.

Hiermit erkenne ich die Promotionsordnung des Fachbereiches Mathematik und Naturwissenschaften der Bergischen Universität Wuppertal an.

Contents

1	Introduction	1
2	Theoretical Background	3
2.1	Extreme Events in time series	3
2.1.1	What are extreme events?	3
2.1.2	Extreme Events in Time series	4
2.1.3	Assigning a magnitude to the events under study	7
2.2	Predictions via Precursory Structures	8
2.2.1	Describing stochastic processes with conditional probabilities via Markov Chains	10
2.2.2	Applying finite conditioning to general (non Markovian) processes	11
2.2.3	From Hypothesis Testing to the Bayes Classifier	11
2.3	Measuring the Quality of a Prediction	18
2.3.1	Predictability, Kullback Leibler distance and Mutual Information	19
2.3.2	Forecast Verification	20
2.3.3	Brier Score and Ignorance	20
2.3.4	The Receiver Operating Characteristic and the Likelihood Ratio	21
2.3.5	Consequences of the Neyman-Pearson Lemma	24
3	Predicting increments in AR(1) processes	27
3.1	AR(1) process	27
3.2	Determining the precursor value	31
3.2.1	The precursor according to strategy I	32
3.2.2	The precursor according to strategy II	33
3.3	Testing the Performance of the Precursor	33
3.4	Analytical investigation of the Precursor Performance	37
3.4.1	The theoretical precursor given by strategy II	37
3.4.2	Finite (realistic) precursors	38
3.4.3	The dependence on the coupling strength	43
3.5	Long-range correlated ARMA processes	46
3.6	Summary	47
4	Investigating the Influence of the Event Size	49
4.1	Magnitude Dependence of a Prediction	49
4.2	The Test-condition for Likelihood Ratios	50
4.3	The magnitude dependence of the Kullback Leibler distance	54
4.4	Frequently asked Questions	55
4.5	Summary	56

5	Predictions of Increments in i.i.d. random numbers	57
5.1	Numerical evaluation of the test condition	57
5.2	Gaussian distributed random numbers	58
5.3	Exponentially distributed random variables	61
5.4	Symmetrized exponentially distributed random variables	62
5.5	Power-Law distributed random variables	66
5.6	Symmetrized Power-Law distributed Random Variables	70
5.7	Is there a “universal” ROC-Curve for $\eta = 0$?	71
5.8	Symmetry of the distribution	73
5.9	Conclusions	73
6	Predicting threshold crossings	75
6.1	Non-Gaussian AR(1) processes	75
6.2	Results for the Gaussian AR(1) process	76
6.3	AR(1) Process with Symmetrized Exponential Distribution	80
6.4	Power-Law Distributed random variables	82
6.5	A more realistic prediction procedure	84
6.6	ROC curves for $\eta = 0$	86
6.7	Summary	86
7	Predictions of Events in Fluid Flows	89
7.1	Distributions of velocity increments in the presence of dissipation range intermit- tency	89
7.2	Predictions of Increments in a Free Jet Flow	90
7.3	Predictions of Threshold Crossings in a Free Jet Flow	94
7.4	Prediction of Increments in Wind Speeds	96
7.5	Summary	96
8	Predicting errors in weather forecasts	99
8.1	The ECMWF’s medium range temperature forecast	99
8.2	Definition of the Events	101
8.3	Identification of the precursor	102
8.3.1	Testing the Magnitude dependence	103
8.3.2	AUC and ROC	106
8.4	Brier Score and Ignorance	108
8.5	Summary	109
9	Conclusions and open Questions	111
9.1	List of specific results	112
9.2	Specific open Questions	114
9.3	Open Questions on Extreme Events and Outlook	114
A	Increments and threshold crossings in ARMA processes	117
B	Evaluating the CPDFs of precursors and events	119

List of Abbreviations and Notations

η	relative event magnitude, measured in units of the standard deviation
d	absolute event magnitude
$\{x_n\}$	time series
$t_n = t_0 + n\Delta$	discrete time steps
Δ	sampling intervall
$\{\chi_n\}$	tracer time series of forecast verifications observed at t_m , corresponding to forecasts issued at time t_n , so that $m > n$;
$\kappa = (m - n)\Delta$	lead time, time difference between the time t_n , when a forecast was issued and the observation of the corresponding event or non-event;
τ	time delay used to construct the precursory variable and the precursor, i.e. dimension of prec. variable and precursor
$\mathbf{s}_n = (x_{n-\tau+1}, x_{n-\tau+2}, \dots, x_{n-1}, x_n)$	precursory variable
$\mathbf{u} = (u_{\tau-1}, u_{\tau-2}, \dots, u_0)$	precursor
$V(\mathbf{u}, \delta)$	the <i>alarm volume (tolerance volume)</i> around a precursor
$I(u, \delta)$	the <i>alarm interval</i> for a one-dimensional precursor
δ	the parameter which determines the size of the alarm volume
$\mathbb{P}(\cdot)$	cumulative probability distribution function
$p(\cdot)$	probability distribution function
$\rho(\cdot)$	probability density function
PDF	probability density function and/or probability distribution function, depending on the context;
CPDF	conditional probability distribution function and/or conditional probability density function depending on the context;
$p[\mathbf{s}_n, \chi_n(\eta) = 1]$	joint PDF of precursor and event of a given magnitude η
$p[\mathbf{s}_n, \chi_n(\eta) = 0]$	joint PDF or precursor and non-event
$\rho[\mathbf{s}_n \chi_n(\eta) = 1]$	the <i>posterior</i> PDF of observing an event of a given event magnitude η
$\rho[\mathbf{s}_n \chi_n(\eta) = 0]$	the <i>posterior</i> PDF of observing a non-event
$p[\chi_n(\eta) = 1 \mathbf{s}_n]$	the <i>likelihood</i> of observing an event of a given magnitude η
$p[\chi_n(\eta) = 0 \mathbf{s}_n]$	the <i>likelihood</i> of observing a non-event
$p[\chi(\eta) = 1]$	the <i>prior probability</i> of <i>total probability</i> of observing an event of a given magnitude η
$p[\chi(\eta) = 0]$	the <i>prior probability</i> of <i>total probability</i> of observing a non-event

$\Lambda(\mathbf{s}_n, \eta)$	the <i>likelihood ratio</i> , in our notation in fact a ratio of posterior PDFs
$B(\mathbf{s}_n, \eta)$	the <i>Bayes classifier (ratio of odds)</i>
$r_c = r_c(\mathbf{u}, \delta, \eta)$	rate of correct predictions for predictions via precursors
$r_f = r_f(\mathbf{u}, \delta, \eta)$	rate of false alarms for predictions via precursors
$A_n(\mathbf{s}_n, \delta, \eta)$	the decision variable for predictions via precursors
$\hat{A}_n(\mathbf{s}_n, \delta, \eta)$	the decision variable for probabilistic forecast
$\hat{r}_c = \hat{r}_c(\mathbf{u}, \delta, \eta)$	rate of correct predictions for probabilistic forecasts
$\hat{r}_f = \hat{r}_f(\mathbf{u}, \delta, \eta)$	rate of false alarms for probabilistic forecasts
ROC	Receiver Operating Characteristic
AR(1) process	Autoregressive process of order one
ARMA(m, n) process	Autoregressive moving average process of orders m and n

Chapter 1

Introduction

Systems with a complex time evolution which generate a great impact event from time to time are ubiquitous. Examples include electrical activity of human brain with rare epileptic seizures, fluctuations of prices for financial assets in economy with rare market crashes, changing weather conditions with rare disastrous storms, floods or landslides, fluid flows with rare intermittent bursts, and also fluctuations of on-line diagnostics of technical machinery and networks with rare breakdowns or black-outs.

Since humans are typically adapted to the “normal”, i.e. non-extreme state of a system and not to its rare extremal behavior, many extreme events have a destructive impact on society. Hence, there should be no need to motivate the research of them. In the last years the interest in rare events, sometimes also titled “*black swans, kings*” [1] or “*outliers*” [2] significantly increased, since climate models suggest an increase in the frequency and the intensity of extreme climate events [3]. As a consequence researchers in different geo scientific disciplines now show an increased interest in research on natural hazards.

Besides the exploration of done for each type of events, e.g., landslides, earthquakes, floods, or hurricanes, there are approaches, which aim in accessing common features of extremes [4, 5]. Many of these generalizing approaches study extreme events by using concepts and methods from “complex systems science”, such as the assumption of long range correlations [6], cellular automata models [7], self organized criticality [4] or others [8]. A second type of generalizing approach concentrates on the statistical properties of extreme events. Many of these approaches (see, e.g., [9]) are based on the application of *extreme value theory* (see [10, 11, 12, 13, 14] and many others) or the *theory of large deviations* (see e.g., [15, 16]). The knowledge about the distribution of extreme events is then often used to calculate return time distributions and return time levels, see, e.g., [17, 18]. Although extreme value theory allows to access the statistical properties of extremes, it can not serve in order to predict the occurrence of single events.

In order to increase the knowledge about extreme events, both approaches, the detailed expert knowledge about the system generating extreme events, and the generalizing approach which searches for similarities in different phenomena are indispensable.

In this thesis we assume the point of view of the generalizing approaches, by investigating extreme events which are simply defined as structures in time series, independently of the mechanism which generated them. Nevertheless, we need a definition of the targets of our interest, in order to transfer the term “extreme event” from every day language into a scientific concept. Therefore a definition which characterizes extreme events in a general sense, i.e., independent from the mechanism, which generated them is developed in Sec. 2.1. Due to the variety of phenomena labeled as extreme events, this definition cannot claim to be exhaustive. However, it is able to describe many different types of events, such as strong turbulent wind gusts, errors in weather forecasts, threshold crossings, epileptic seizures and many others

The complexity of the systems generating extreme events, often prevents a complete modeling, either due to the huge number of degrees of freedom involved, or due to a lack of precise knowledge about the governing equations. Considering the great impact of some of the above

mentioned events, a prediction of their occurrence is nonetheless highly desirable. Therefore, there have been many attempts to investigate time series of a record of historical data and try to infer knowledge about the future. One possibility to do so, consists in predicting via identifying precursory structures. In order to keep the method of prediction as general as possible, we focus on predictions made via precursory structures, which are identified through the conditional probability that an event follows a given precursory structure. Hence, we do not use additional information about the details of the system under study, but simply rely on the estimates of conditional probabilities, obtained from the time series. Thus, this method explained in more detail in Sec.2.2, is generally applicable to a variety of phenomena. A very similar approach is e.g., successfully applied for the prediction of large earthquakes [19].

We keep the point of view of a joint approach, when focusing on a specific phenomenon, namely the observation, that the quality of a prediction is often dependent on the magnitude of the events under study. For the prediction of wind gusts [20], the prediction of precipitation events [21], prediction of collective phenomena in multi agent games [22] and also for the prediction of avalanches in systems which display self organized criticality [23] it has been reported that larger events are better predictable than smaller, although they are more rare. Furthermore it is known, that also in weather forecasts, large scale events are easier to predict [24].

It is the main goal of this thesis, to understand and describe this dependence on the event magnitude independently of the example under study. In Sec. 2.3.1 we will introduce different methods of forecast verification and then discuss, why we consider the Receiver Operator Characteristic curve to be a suitable measure for the quality of the prediction of extreme events. In contrast to many other measures for the quality of predictions, the Receiver Operator Characteristic curve is not designed and optimized to be “good in the average” and does not weight failures in the prediction of events according to their rareness.

Having specified the prediction method and the measure for the quality of a prediction, we develop in Chap. 4 a condition, which allows us to test for the dependence on the event magnitude. Next, we study the performance of this test condition for the prediction of events defined as increments in short and long-range correlated stochastic processes (in Chap. 3), in sequences of uncorrelated random numbers (in Chap. 5), in velocities, measured in a fluid flow experiment and wind speeds (Chap. 7). Analogously we also investigate the prediction of threshold crossings in short range correlated stochastic processes (Chap. 6) and the data sets named above. Furthermore we study the magnitude dependence also for the prediction of large failures of weather forecasts in Chap 8.

Parts of the results presented in this thesis are also published in [25],[26], [27], [28] and [29].

Chapter 2

Theoretical Background

2.1 Extreme Events in time series

2.1.1 What are extreme events?

In the introduction we already mentioned natural hazards, like floods or storms, breakdowns of electricity networks or epileptic seizures as popular examples for extreme events. Although these examples suggest an intuitive understanding of the term extreme events, a precise scientific definition of it is still missing. Due to the different context and circumstances in which extreme events are discussed, it is very difficult to provide a definition which covers all examples and is still precise enough to be useful. On the other hand, there is a strong need to specify the target of our interest. That is why we choose to work with a preliminary definition of extreme events which comprises the examples we are interested in. However, one has to keep in mind that this definition does exclude many other examples. We arrive at this definition by using the following considerations which are also proposed in [25].

In all types of phenomena where the event magnitude can assume any value inside some interval, one has to decide beyond which magnitude one calls an event “extreme”. This implies first of all that we can define a variable η which characterizes the magnitude of the event and which is a function $\eta(x)$ of a scalar or vector valued observable x of the system under study. Note that the magnitude is the essential key to transfer the term “extreme event” from the every day language into a scientific definition. We demand now that the explicit form of $\eta(x)$ is chosen such that $\eta(x)$ assumes a large value if the observable x is in a state to which we would refer as an extreme event. Typically the critical magnitude, beyond one would call an event extreme is chosen such that events of larger magnitude cause damage or harm to us.

Furthermore we demand that the events which we call extreme are related to observations x , which are different from the “typical” behavior of the system under study. This implies the rareness of extreme events. However, not every rare event is an extreme event. A counter example for a rare but not extreme event is winning the lottery. There is no way to distinguish the 6-tuple of drawn numbers from all other possible 6-tuples, such that we could call this observation a non-typical behavior of the dynamics of the system under study, i.e, the lottery wheel.

Additionally, we require some form of irregularity in the occurrence of extreme events, since otherwise prediction and adaptation to the events would be trivial. This irregularity in the occurrence of the extreme events requires that the underlying system possesses some complex dynamics.

Another issue is whether an event appears endogenously, generated by the system dynamics, or exogenously, being induced by some external perturbation. The distinction between exogenous and endogenous depends of course on the way in which you define the system under study, e.g., formerly endogenous events might become exogenous by excluding subsystems. In total, we will assume a slightly different point of view than Sornette [30] and not regard small perturbations as

external, which allows us to study also stochastic dynamics. Hence, small perturbations that act like noise are then considered as inherent to the system, and only externally controlled changes of system parameters or macroscopic perturbations that are much stronger than noise will be considered to be exogenous.

Finally we arrive at our working definition of Extreme Events.

Definition 1 (Extreme Events) *As an extreme event we understand an event which is generated by a system with an underlying (deterministic or stochastic) complex dynamic. We demand that it fulfills the following properties:*

Quantifiability: *It is possible to assign a magnitude η to the event, where $\eta(x)$ is a function of an observable x of the system under study. The explicit form of $\eta(x)$ is chosen such that $\eta(x)$ assumes a large value if the observable x is in a state to which we would refer as an extreme event,*

Rareness: *The event is rare, i.e., its occurrence is reflected by the tails of the probability distribution which describes the occurrence of all possible events.*

Irregularity: *The event does occur irregularly, i.e., it does not occur periodically.*

Endogeneity: *The event is generated by the dynamics of the system itself and not as a consequence of external shocks.*

As already mentioned this definition describes the examples of extreme events studied here, but it is not a general accepted definition. The two postulates in our definition which are the most ambiguous are rareness and endogeneity.

One can certainly not argue about the fact that extreme events are rare. However, one could be more restrictive about the conditions events characterized by the tails of probability distributions have to fulfill. Sometimes the term extreme events is only used if the events are characterized by tails, which decay slower than in the Gaussian or the exponential case, e.g., power-law tails. Following this idea, one could demand that the rareness of an event is quantifiable in order to call it an extreme event. An example for this would be demanding that the relative frequency of its occurrence is smaller than a given threshold.

The question of how to distinguish between an endogenous or an exogenous origin of the behavior of a system is typically linked with an example, in which these influences are in particular hard to distinguish: the stock market. However, one could also think of other events which consist simply in the response of a system to another primary event, e.g., a tsunami caused by an earthquake or a large landslide. While it is possible to recognize some causality between the events in the subsystems in such a well defined and simple example of the coupling of two systems (the ocean and the surface of the earth), it is much more difficult to identify this causality in systems which consist of many interacting subsystems, like markets. It was discussed since the 1930s whether stock-marked crashes are endogenous events, caused by the system of the market itself or exogenous events, which are stimulated by external shocks [31]. In the recent literature many authors implicitly assume the existence of an endogenous origin of crisis, by studying endogenous market crashes [32] or speculation bubbles.

2.1.2 Extreme Events in Time series

In this contribution we study the occurrence and prediction of extreme events which can be observed in a scalar time series [33, 34, 35], i.e., a set of N observations $\{x_n\}$ at discrete times t_n , where $t_n = t_0 + n\Delta$, with $n = 0, 1, \dots, N - 1$ with a fixed sampling interval Δ . We are not employing any prior knowledge about physical processes or models for the phenomenon under study, but rely only on recordings of past data. In case that the phenomena governing the evolution of the system under study are fairly well understood, and physical models have been

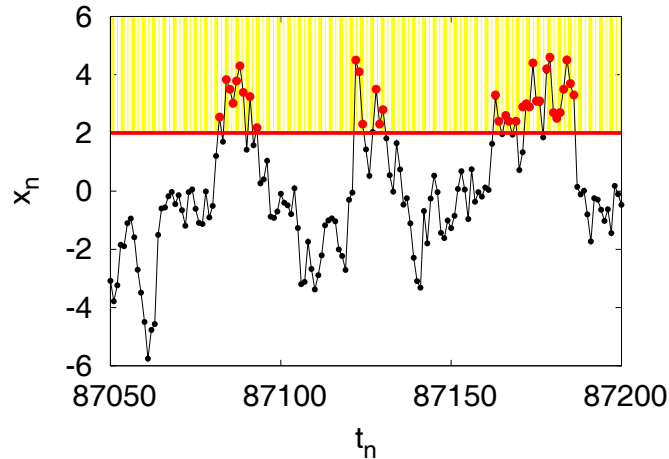


Figure 2.1: An example for threshold crossings for a threshold $\eta = 2$ which were observed in an autoregressive process of order one.

demonstrated to yield useful forecasts over a large spectrum of spatial and temporal scales, this approach would not be reasonable. In a variety of other circumstances though, a time series approach is the only possibility for predictions at all. A frequent reason is that good models for the specific situation are not available.

As a consequence of the approach via time series analysis we demand that the events under study are recurrent in the sense that we are able to extract statistical information about them from a time series. Note that this postulation does exclude events which change the dynamics of the system, as, e.g., events which terminate the life-time of a system, such as material fracture.

We trace the occurrence of an event in a second time series $\{\chi_m\}$ with the same sampling interval Δ and $m \in \mathbb{N}_0$. The variables χ_m of this tracer time series are indicators, i.e., they are restricted to the values unity and zero, in order to indicate, whether an event occurred or not. In many applications the tracer time series is derived from the original time series $\{x_n\}$. However, in principle the events could also be defined using the observation of some other quantity. (For simplicity we assume that this other observation is also recorded with the same sampling interval Δ .) In fact, the only important requirements are that at the time t_n when the forecast of an event has to be made, χ_m is unknown and that its actual value is revealed later. The relative time distance $\kappa = (m - n)\Delta$ between the time of the last observed value, which contributes to the forecast x_n and the time t_m at which the event occurs is usually called forecast horizon or lead time.

Since we are typically interested in comparing the forecast which was issued at t_n with the observation of an event happening at a future time t_m we decide to shift the time index of the tracer time series in such a way that an event (non-event) recorded in the tracer time series will have the same time index as the corresponding forecast,

$$\chi_n = \begin{cases} 1 & : \text{an event occurred at time } t_m = t_n + \kappa\Delta, \\ & \text{the corresponding forecast was issued at time } t_n; \\ 0 & : \text{no event occurred at time } t_m = t_n + \kappa\Delta, \\ & \text{the corresponding forecast was issued at time } t_n. \end{cases} \quad (2.1)$$

We are now going to take a closer look at some simple but interesting structures in time series, which are typically discussed as extreme events, if their magnitude is larger than expected.

Threshold Crossings

An important class of events are the exceedances of a given threshold η , see Fig.2.1. There are many examples in which a disastrous event is characterized by a variable which exceeds

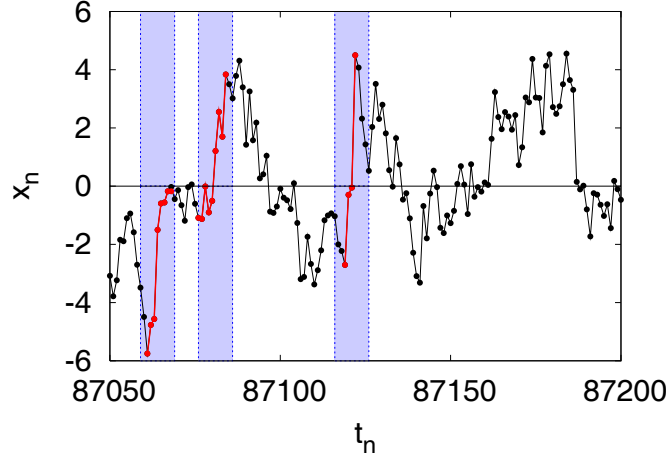


Figure 2.2: An example for increments occurring in an autoregressive process of order one. The colored regions highlight increments $x_{n+k} - x_n \geq 5$ occurring in less than 8 time steps.

a threshold as, e.g., the water level in a river which can exceed the height of a levee or the magnitude of a seismic activity which is stronger than a certain value.

If we assume that the tracer time series is derived from the one dimensional time series $\{x_n\}$, the corresponding events and non-events are defined as follows:

$$\chi_n(\eta) = \begin{cases} 1 & : \text{ if } x_{n+\kappa} \geq \eta, \\ 0 & : \text{ if } x_{n+\kappa} < \eta, \end{cases} \quad (2.2)$$

in both cases we assume that the index n of the tracer time series refers to a forecast made at a time t_n with $n < m$, as in Eq. 2.1.

Increments

Increments are events which consist in a sudden change of the observed variable within a few time steps. In this contribution we concentrate on sudden increases. An example for this kind of extreme events are wind gusts appearing as sudden increases in wind speed [20, 36] which can destroy wind turbines or influence aircraft take off and landing, but also stock market crashes [37] which consist in sudden decreases. Figure 2.2 shows an example for increments observed in an autoregressive process of order one. In this context we assume that the time series under study is a one dimensional time series and we define our extreme increment $x_{n+k} - x_n$ to be larger than a given distance η

$$\chi_n(\eta) = \begin{cases} 1 & : \text{ if } (x_{n+\kappa} - x_n) \geq \eta \\ 0 & : \text{ if } (x_{n+\kappa} - x_n) < \eta \end{cases} \quad (2.3)$$

Again we assume that the index n of the tracer time series refers to a forecast made at time t_n as explained in Eq. 2.1.

Increments and threshold crossings can be related to each other by creating a new time series $z_n = x_{n+k} - x_n$. If the time series under study is generated by a well defined stochastic process, e.g., an autoregressive moving average process (ARMA) of order (m,n) this transformation from the original time series to the time series of increments $z_n = x_{n+1} - x_n$ induces a transformation to another stochastic process. In the example above, which is in more detail discussed in appendix A the increment time series is described by an autoregressive moving average process of order $(m,n+1)$.

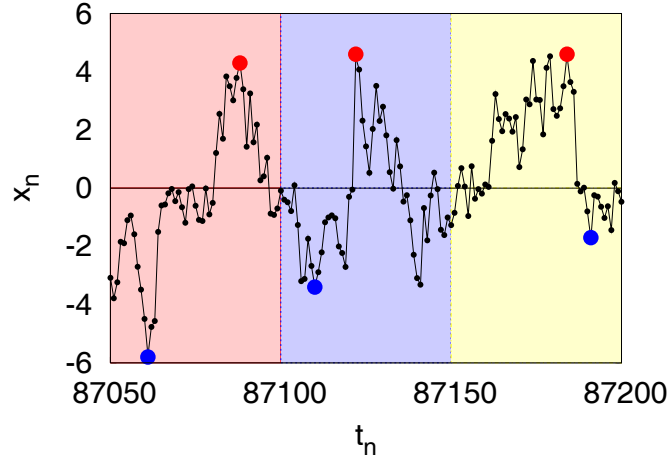


Figure 2.3: An example for the definition of extreme events as block maxima and minima i.e., the largest (smallest) values within 50 data points. The respective maxima and minima are highlighted by increased symbols and the color shaded regions indicate segments of 50 data points

Block Maxima

Another class of events, which is especially important for time series exhibiting periodic changes, e.g., seasonal fluctuations, are so called block maxima. Block maxima represent the largest value of a finite number of data points. They are of historical importance, since their study in river level data lead to the development of extreme value statistics [12].

Apart from the examples of event classes listed here, one can – with respect to the system under study and its interesting events – define many other different types of extreme events and thus construct $\{\chi(\eta)\}$ for $\{x_n\}$ in various ways.

2.1.3 Assigning a magnitude to the events under study

In the following we are going to measure the magnitude of an event in units of the standard deviation σ of the underlying stochastic process,

$$\eta = x_n/\sigma, \quad \text{for threshold crossings;} \quad (2.4)$$

$$\eta = \frac{(x_{n+k} - x_n)}{\sigma} = \frac{d}{\sigma}, \quad \text{for increments,} \quad (2.5)$$

where the variable d denotes the absolute size of the event. This definition for the event magnitude implies that the variance of the process is finite which is the case in the examples studied here. However, there exist also processes with infinite variance, e.g. the Weierstrass Random Walk[38, 39]. For processes with infinite variance there is no intuitive measure for the magnitude of an event and one has to describe the magnitude of the event indirectly, e.g., by its relative frequency.

2.2 Predictions via Precursory Structures

Et de eis quae significant illud [the aproaching of the plague] est ut videas mures et animalia quae habitant sub terra fugere ad superficiem terrae et pati sedar, id est, commoveri hinc inde sicut animalia ebria.

(Avicenna, Liber canonis, translation from 1556, original work written in Arabic between 1005 and 1024)

translation: And among those [signs] which announce it [the approaching of the plague] is, that you will see mice and animals which live under the earth flee to the surface and suffer, i.e. here, they will move like drunken animals.

The search for precursors which announce extreme events is a very old idea and it is the foundation of both, superstition and adaptive learning. If the assumption, that there is a correlation between the event and its precursor is wrong, we are very likely to call the corresponding prediction superstition, as in the famous example of the black cat which is interpreted as a precursor for bad luck. However, if a correlation between a precursor and an event exists, the interpretation of the precursor is very likely to turn into passed on knowledge, as for example the quote of Avicenna, listed above, which is according to Abel [40] the first reliable source of quoting anomalous behavior of rodents as a precursor for the plague.¹

Another example for passed on knowledge about precursors are the legends about precursory phenomena in clouds or animal behavior previous to large earthquakes, which were mainly told in Asian countries [41]. Today, there are different studies, trying to explain the observed phenomena as caused by electromagnetic signals [41]. However, naturally the distinction between superstition and passed on knowledge was not trivial in times where many scientific phenomena were not yet understood and is in some cases difficult even today. Thus, if we are interested in studying precursors from a scientific point of view, we have to be careful in examining the correlations between event and precursor and we have to examine the quality of our precursors by evaluating the quality of the corresponding predictions. In other words, we have to move away from sporadically observed precursors to precursors which are systematically identified by data driven methods.

Prediction via precursory structures on a scientific basis is discussed in a variety of fields. The most prominent example is the already mentioned task of earthquake prediction [42, 43, 44, 19], but precursors are also discussed for the prediction of epileptic seizures [45, 46, 47] and many others. Some authors even suggest that there could be “*similarities in the corresponding precursory features for earthquakes and epileptic seizures*” [4]. However, in both, earthquake and epileptic seizure prediction, it is not yet fully accepted that a practically applicable prediction of the events under study is indeed possible [47, 48, 49, 50]. Recent examples for prediction with precursory structures are the prediction of avalanches in SOC-systems[51] and the prediction of strong turbulent wind gusts[20, 36, 52, 53], which are known to destroy the rotor blades of wind turbines.

All the examples mentioned have in common that the predictions are made for a time in the “very near future”, although the time scale on which this “very near future” is defined can naturally differ with respect to the problem under study. The ideal prediction horizon for the prediction of strong wind gust would be e.g., in the order of tens of seconds, while earthquake warnings are most useful if they are issued days ahead. This is one of the typical aspects of a situation in which people prefer forecasting via identifying precursory structures, whereas model based forecasts are used to obtain information about the “far future”. Other indicators for the use of precursors are, e.g., a complicated dynamics of the system under study which cannot be sufficiently well described by a (deterministic) model, or if the corresponding model is too complex to allow simulations with reasonable computational effort or if initial conditions for the model.

¹Indeed, the plague is a disease which affects rodents and humans and was typically transmitted by rat fleas.

Due to the different fields in which precursor based predictions are discussed, there are also various interpretations of the term precursor, which we will not list here in detail. In the following we will assume that at least a part of the dynamics of the system under study is unknown and therefore has to be described by a stochastic term. The consequence of this is, that we cannot demand from the precursor to precede *every* individual event. However, we can expect the data structure we call precursor to *typically* preceding an event, allowing deviations from the given structure, but also allowing events without preceding structure. This interpretation of a precursor allows to determine the specific values of the precursory structure by statistical considerations.

In order to predict an event occurring at time $(n + 1)$ we compare the last τ observations

$$\mathbf{s}_n = (x_{n-\tau+1}, x_{n-\tau+2}, \dots, x_{n-1}, x_n) \quad (2.6)$$

to which we will refer as *precursory variable*, to a specific *precursor*², also called *predictor*,

$$\mathbf{u} = (u_{\tau-1}, u_{\tau-2}, \dots, u_0). \quad (2.7)$$

We can then define an *alarm volume* $V(\mathbf{u}, \delta)$ around each precursor as the set of all \mathbf{s}_n for which $\|\mathbf{s}_n - \mathbf{u}\| \leq \delta$, where $\|\cdot\|$ denotes a norm which can be, for example, the Euclidean norm or the maximum norm. When using the maximum norm, the alarm volume consists of all time series segments which fall into a δ -tube around the precursory pattern \mathbf{u} . We give an alarm for an event if we find the precursory structure \mathbf{s}_n within the alarm volume. In a formal way we can express this procedure by defining a *decision variable*

$$A(\mathbf{s}_n, \mathbf{u}, \delta) = \begin{cases} 1 & : \text{ if } \mathbf{s}_n \in V(\mathbf{u}, \delta), \\ 0 & : \text{ otherwise.} \end{cases} \quad (2.8)$$

Note that this approach is inspired from the ‘‘Lorenz method of analogs’’ [54, 55] which refers to prediction via so called *local predictors* which were defined in state space for predictions in chaotic systems. Furthermore Keilis-Borok and Kossobokov used a similar approach in order to construct their well known M8 algorithm for the prediction of earthquakes [19].

We will now consider the question, how we can identify precursors in time series. There are various approaches to do this, some of them are based on additional knowledge of the system under study, as e.g., an epileptic seizure is known to be a state of high synchronization and hence one tries to interpret increases in synchronization as precursors for seizures. However, there exist also purely data driven approaches which do not require any additional knowledge about the system and can hence be applied to arbitrary data. Osterhage and Lehnertz give a nice overview of such approaches in [56], where they list changes in the correlation sum, the correlation dimension, correlation density, correlation entropy, marginal predictability, Lyapunov exponents or local flow in state space as possible precursors of epileptic seizures.

We prefer to choose the precursor by using conditional probabilities. A similar approach was already used in the work of Ragwitz, Vitanov, Holstein and Kantz [52, 53, 20, 36], who studied the prediction of turbulent wind gusts via using conditional probabilities and probability densities. However, they preferred to use the mean of the conditional probability $p(\chi_n, \mathbf{s}_n)$ as a precursory structure, whereas we intend to maximize the respective conditional probabilities.

In the field of time series analysis the use of conditional probabilities is strongly linked with Markov-chains. In weather forecasting conditional probabilities form the so called probabilistic forecasts, and an analog use of conditional probabilities or probability densities is also the basis of statistical or Bayesian inference. Furthermore, artificial neural networks for pattern recognition are designed in a way that they approximate respective conditional probabilities.

Since for many considerations the distinction between probabilities and probability densities is not of relevance, as far as numerical estimates of probability densities are involved, we

²We are going to address the issue of how to identify such precursors in Subsec. 2.2.3. Suppose for the moment that we had already chosen a precursor \mathbf{u} in one way or another.

formulate the corresponding expressions mostly in terms of probabilities, denoted by $p(\cdot)$. If a distinction between probability and probability density is possible or desirable, we denote the corresponding probability densities by $\rho(\cdot)$. In general we will also use the abbreviation PDF for both, the term *probability density function* and the term *probability distribution function*. The same yields for conditional probability distribution functions and conditional probability densities which will both be denoted as CPDF.

2.2.1 Describing stochastic processes with conditional probabilities via Markov Chains

In the following we will interpret the time series under study as being generated by a discrete time stochastic process [33, 57, 58], regardless of what process and what measurement function created the data.

A Markov process is defined as a stochastic process with the property, that for any set of successive times $t_1 < t_2 < \dots < t_n$

$$\rho(x_n, t_n | x_1, t_1; \dots; x_{n-1}, t_{n-1}) = \rho(x_n, t_n | x_{n-1}, t_{n-1}) \quad (2.9)$$

holds [57]. This indicates that the process has only a finite memory (one time step) and one can hence infer the probability distribution/density function at time step t_n by knowing the value of x at the previous time step t_{n+1} . There are various textbooks on Markov processes and their properties [57, 59], thus we will highlight here just some aspects which are relevant in the context of forecasting.

Furthermore we assume the process to be *stationary*, although stationarity is a property which almost never applies to realistic processes such as atmospheric turbulence. Applying concepts from stationary processes to data which might originate from a non-stationary process could result in reduced performance of our prediction algorithms, but is not a fundamental problem in the examples of non-stationary data³, which we study in this contribution. Moreover, the methods proposed in this contribution should also be suitable in the special case of non-stationarity due to slowly varying system parameters, as it was argued in [60]. In terms of the prediction of wind-speeds in high frequency wind speed data those slowly varying system parameters are related to changing weather conditions or change of the time of the day.

For stationary and for homogeneous Markov processes the conditional probability $\rho(x_n, t_n | x_{n-1}, t_{n-1})$ does not depend explicitly on time, but just on the time interval r between the successive time steps, such that we can use $\rho(x_n | x_1, \dots, x_{n-1}) = \rho(x_n | x_{n-1})$ instead of Eq. (2.9). If we assume that the process under study is a discrete time process, we can construct a chain of CPDFs

$$\rho(x_n, x_{n-1}, \dots, x_0) = \rho(x_n | x_{n-1}) \rho(x_{n-1} | x_{n-2}) \dots \rho(x_1 | x_0) \rho(x_0), \quad (2.10)$$

in which the probability density for the present state of the process is given by the knowledge of the transition PDF $\rho(x_j | x_{j-1})$ and the initial function $\rho(x_0)$.

We implicitly assume that the Markov property (Eq. (2.9)) holds, whenever we choose the precursory variable to be scalar $s_n = x_n$, i.e., we condition only on the value occurring at the last time step. If we condition on more than the last time step, i.e. $\mathbf{s}_n = (x_{n-\tau+1}, x_{n-\tau+2}, \dots, x_{n-1}, x_n)$, we assume the process to be a Markov process of a higher order τ , i.e., we expect

$$\rho(x_{n+1} | x_n, x_{n-1}, \dots, x_0) = \rho(x_{n+1} | x_n, x_{n-1}, \dots, x_{n-\tau+1}) \quad (2.11)$$

to hold for any set of successive time indices $0, 1, \dots, n+1$. This process can be transformed into a corresponding (generalized) Markov process of order 1 by considering a vector valued process with $\mathbf{x}_j = \{x_n, x_{n-1}, \dots, x_{n-\tau+1}\}$, and the following consequence of Eq. (2.11) [57]

$$\rho(x_{n+1}, x_n, \dots, x_{n-\tau+1} | x_n, x_{n-1}, \dots, x_0) = \rho(x_{n+1}, x_n, \dots, x_{n-\tau+1} | x_n, x_{n-1}, \dots, x_{n-2\tau+1}), \quad (2.12)$$

³Data are called non-stationary, if the null-hypothesis of stationarity can easily be rejected.

such that

$$\rho(\mathbf{x}_j | \mathbf{x}_{j-1}, \mathbf{x}_{j-2}, \dots, \mathbf{x}_0) = \rho(\mathbf{x}_j | \mathbf{x}_{j-1}) \quad (2.13)$$

holds.

Another property, which holds for Markov processes and is of importance for forecasting is the fact that they fulfill the Chapman Kolmogorov Equation [57, 61]

$$\rho(x_{n+1} | x_{n-1}) = \int dx_n \rho(x_{n+1} | x_n) \rho(x_n | x_{n-1}). \quad (2.14)$$

By further marginalizing, we arrive at CPDFs connecting observations, which are κ time steps apart:

$$\rho(x_{n+\kappa} | x_n) = \int dx_{n+\kappa-1} \dots dx_{n+1} \rho(x_{n+\kappa} | x_{n+\kappa-1}) \rho(x_{n+\kappa-1} | x_{n+\kappa-2}) \dots \rho(x_{n+1} | x_n). \quad (2.15)$$

This allows us to use lead times $\kappa > 1$, i.e., to forecast based on $\rho(x_{n+\kappa} | x_n)$.

2.2.2 Applying finite conditioning to general (non Markovian) processes

CPDFs provide the information needed for (probabilistic) predictions: Knowing $p(x_{n+\kappa} | x_n, x_{n-1}, \dots, x_{n-\tau+1})$ and given specific values for the $x_n, x_{n-1}, \dots, x_{n-\tau+1}$'s, one can calculate the probability that the observation Δ time steps in the future will fall into a given interval. Generally, the probability density function or probability mass function of $x_{n+\kappa}$ will be the sharper, the further into the past the conditioning extends. Ideally, the entire past of the process would be observed and the CPDF for infinite conditioning would be known, thus yielding optimal knowledge of the future (which does not mean that this CPDF becomes necessarily sharp peaked like a δ -function). In practice this is absolutely out of reach. The difficulty here is to estimate the CPDF from the sample of N data points, as this estimate gets the worse the larger τ . If the observed time series were governed by a generalized Markov process of order τ_0 , then the τ_0 -step conditioning would be optimal and any additional conditioning would not improve (or in fact change) the forecast. In general, although the process is not Markovian, finite conditioning still provides a rather good *approximation* to infinite conditioning, or in more colloquial terms, there is nothing wrong about basing one's predictions on finite τ -conditioning, the worst to happen is that this is sub-optimal. In total one can say that whenever one constructs a CPDF with respect to a finite number of time steps in the past, and then uses this CPDF in order to infer the distribution of a future value, one implicitly assumes that the process under study is a Markov process. Although this assumption is not valid in general, it is impossible to avoid it, due to the fact, that every algorithm will require a conditioning on a *finite* number of time steps τ . Furthermore, the only consequence of the fact, that the conditioning was not done with respect to sufficiently many time steps into the past, are probably unsatisfactory forecasts.

2.2.3 From Hypothesis Testing to the Bayes Classifier

As already explained above, we will use conditional probabilities and probability densities in order to identify precursory structures. Conditional probabilities and probability densities (CPDFs) are used in various contexts, e.g., hypothesis testing and classification problems, but also for probabilistic forecasts. Therefore it is not surprising that we can find different names for the same quantities, but also different concepts, which are expressed by the same formulas. In this section we will try to sort out possible confusions of names and concepts. Thus, we start with highlighting the classical use of CPDFs in the form of hypothesis testing, e.g., used for Bayesian decision making, and then move on to a different interpretation in terms of the maximum likelihood method.

There are two possible choices of using CPDFs: One can either condition on an already observed event, such that the CPDF describes the probability or probability density to find

a specific precursory variable before the event/non-event occurs, namely $\rho[\mathbf{s}_n|\chi_n(\eta)]$, or one can condition on precursory variables, and evaluate the probability, that an event/a non-event precedes them $p[\chi_n(\eta)|\mathbf{s}_n]$.⁴ Both alternatives are linked via Bayes' theorem [62] and they are called *posterior* PDF (also a *posteriori* PDF) and *likelihood* [63, 64]. In order to understand the origin of these names one has to take a short glance at the framework of hypothesis testing. For simplicity let us consider just a single hypothesis H , which can be accepted due to the observation of the evidence E . Hence the PDF describing the acceptance of the hypothesis is given by

$$p(H|E) = \frac{p(E|H)p(H)}{p(E)}, \quad (2.16)$$

which is Bayes' theorem and goes back to Bayes and Laplace in the 18th century[64]. In this framework of hypothesis testing [63, 65]

$p(H|E)$ is called *posterior probability* or a *posteriori probability* of the hypothesis to be true, after observing the evidence E ;

$p(E|H)$ is referred to as *likelihood* of seeing the evidence E , if the hypothesis is true;

$p(E)$ has the names *predictive probability* or *marginal probability* and

$p(H)$ is called *prior probability*, that $p(H)$ is true, prior to (i.e., without) considering the evidence E .

For hypothesis testing, the estimates are then made by considering the posterior PDF. The names *prior probability* and *posterior probability* can only be understood in the sense of an update of the PDFs by using the information provided by the evidence E [66]. The prior PDF reflects in this context the information, which was available before the evidence E was observed. Key to the framework which is today known as *Bayesian statistics* is the idea, that the prior probability can also be obtained by other methods than the analysis of the available data [64, 63, 66], e.g. by the opinion of an expert or by an ad hoc choice of a probability distribution function, e.g. a Gaussian.

This practice is disliked by the so called "*frequentists*" which follow the idea of R. A. Fisher [67, 68] that statistical inference must be based entirely on probabilities with direct experimental verification [64]. This implies that they should be constructed by analyzing the relative frequencies of the occurrence of events. Furthermore, the school of the frequentists prefers to use the maximum of the likelihood $p(E|H)$ in order to estimate, e.g., parameter values, instead of the "posterior PDF". In this context, the evidence is assumed to be fixed and the hypothesis is variable. While the first difference to the Bayesian framework, the dislike of subjective probabilities, is a rather strong issue, the second one, the concentration on the likelihood, does not make such a big difference as it seems on the first glance. The reason for this is that the choice, to which conditional probability one assigns the names *likelihood* or *posterior probability* is rather subjective. We will illustrate this issue by transferring the framework from hypothesis testing to forecasting via precursory structures. It turns out quickly that there is some need for subjective interpretation, if we have to decide what we consider to be the hypothesis under study, and what the observed evidence. Naturally there are (at least) two ways of formulating hypotheses in terms of the prediction via precursory structures. The first one focuses on the hypothesis, that an event is expected to follow the precursor by specifying:

H : an extreme event will happen (after E is observed)

E : we observe a distinguished pattern \mathbf{u} in the precursory variable \mathbf{s}_n .

⁴Please do not worry about the usage of the brackets "[" and "]" as e.g., in $p_n[x_n|\chi_n(d) = 1]$ and in the following formulas, since it has no special meaning in this context and is simply introduced to improve the readability of the formulas.

The second approach focuses on the selection of the precursor:

H: a specific choice \mathbf{u} (precursor) of all possible patterns \mathbf{s}_n (precursory variables) is a precursor for an extreme event, i.e., we give an alarm after observing \mathbf{s}_n in $V(\mathbf{u}, \delta)$.

E: an event occurred (after the precursor was observed and we gave the alarm).

The corresponding hypothesis of observing a non-event and giving no alarm can of course be formulated analogously. Since we are interested in identifying precursors, we decide to consider the second interpretation which is the basis of the following nomenclature.

Definition 2 (Nomenclature for (C)PDFs in the context of forecasting)

In this contribution we call

$\rho[\mathbf{s}_n|\chi_n(\eta)]$ *the posterior PDF,*

$p[\chi_n(\eta)|\mathbf{s}_n]$ *the likelihood,*

$\rho(\mathbf{s}_n)$ *the marginal probability and*

$p[\chi_n(\eta)]$ *the total probability of observing an event.*

All PDFs can be formulated for both values of the event variable, $\chi_n(\eta) = 1$ and $\chi_n(\eta) = 0$.

This choice follows also the convention that the likelihood is typically chosen to be a function of its second argument, while the first argument is fixed [63, 64].⁵ Furthermore it expresses also our approval of frequentists approaches, i.e., we will construct the PDFs based on the analysis of the available data and not on subjective choices. Note however that in the literature of forecasting and signal detection one can also find the terms likelihood and posterior PDF to be assigned vice versa.

After settling the question of the nomenclature, we can now start thinking about whether it is more useful for predictive purposes to exploit the CPDF, which conditions on the precursor, or the CPDF, which conditions on the event. The two possible strategies which arise from this are the following:

Definition 3 (Strategies to identify the Precursor)

Strategy I: *The a posteriori PDF $\rho(\mathbf{s}_n|\chi_n(\eta) = 1)$ takes into account all events of size η and provides the probability density to find a specific precursory structure before an observed event. Hence the first strategy consists in defining the precursors in a retrospective way: once the extreme event χ_n has been identified, one asks for the signals right before it. We can then use the global maximum of the a posteriori PDF to define precursors*

$$\mathbf{u}_I := \mathbf{s}_n : \rho[\mathbf{s}_n|\chi_n(\eta) = 1] = \max \rho[\mathbf{s}_n|\chi_n(\eta) = 1]. \quad (2.17)$$

Strategy II: *The likelihood $p[\chi_n(\eta) = 1|\mathbf{s}_n]$ takes into account all possible values of precursory structures, and provides the probability that an event will follow them. The second strategy consists thus in determining those values of each component x_i of the precursory variable \mathbf{s}_n for which the likelihood is maximal.*

$$\mathbf{u}_{II} := \mathbf{s}_n : p[\chi_n(\eta) = 1|\mathbf{s}_n] = \max p[\chi_n(\eta) = 1|\mathbf{s}_n]. \quad (2.18)$$

⁵Since we consider either the conditional PDFs for the occurrence of an event or a non-event, i.e., $\chi_n = 1$ or $\chi_n = 0$ the first argument of the likelihood is indeed fixed.

Note that the posterior PDF in strategy I can be interpreted as probability or probability density whereas the likelihood in strategy II can (for our purposes) not be formulated as a density function with respect to the precursory structure, but as a probability distribution with respect to the event. This is due to the fact that we distinguish only between event and non-events and hence this probability distribution thus consist of only two bins.⁶ The precursory structure thus enters into the likelihood only as a parameter.

The two possible strategies which we address here, represent fundamental choices of strategies, in the sense, that they are not constructed with respect to additional requirements of possible applications. In more applied examples still based on CPDFs one looks for precursors which minimize or maximize more sophisticated quantities, e.g., discriminant functions or loss matrices, see, e.g., [64]. These quantities are usually functions of the posterior probability or the likelihood, but they take into account the additional demands of the specific problem, e.g., minimizing the loss due to a false prediction. The two strategies studied in this contribution are thus fundamental in the sense that they enter into most of the more sophisticated quantities which were used for predictions and decision making.

Despite the ambiguity of the names of the CPDFs, we can expect the respective prediction strategies to be very different in performance, since they follow different concepts, either conditioning on the precursor or conditioning on the event. Strategy I might seem a very intuitive approach to look for precursory structures and it is indeed applied in problems, where it would be (numerically) too expensive to apply strategy II. In this situation one prefers to concentrate on the precursory structures, which one can extract from the history of already observed events instead of taking all possible values of the precursory variable into account in order to generate $p[\chi_n(\eta) = 1|\mathbf{s}_n]$. However, according to the Neyman-Pearson lemma [69] it is actually strategy II, which is the optimal choice for a scheme to determine the precursor. Hence the performance obtained by identifying precursors with strategy I is expected to be worse or equal, but not better than with strategy II.

In order to show this, one needs a concept called (*Bayes*) *likelihood ratio test* in the context of statistical inference [64] and it is related to a very similar concept, known as the so called *Bayes classifier* (*Bayes test*, *Bayes factor*) for classification problems[65, 64]. As the name likelihood ratio test already tells us, the quantity, which an optimal strategy should maximize is the ratio

$$\Lambda[\mathbf{s}_n, \chi(\eta)] = \frac{p[\mathbf{s}_n|\chi_n(\eta) = 1]}{p[\mathbf{s}_n|\chi_n(\eta) = 0]}. \quad (2.19)$$

Although $\Lambda(\mathbf{s}_n)$ is called *likelihood ratio*, it is in our notation in fact a ratio of posterior PDFs. However we will use the common name likelihood ratio in the following.

Definition 4 (Likelihood ratio test for predictions via precursors) *The (randomized) test with test function ϕ_0 is said to be a likelihood ratio test if it is of the form*

$$\phi_0(\mathbf{s}_n) = \begin{cases} 1 & \text{if } p[\chi_n(\eta) = 1|\mathbf{s}_n] > Kp[\chi_n(\eta) = 0|\mathbf{s}_n] \\ g(\mathbf{s}_n) & \text{if } p[\chi_n(\eta) = 1|\mathbf{s}_n] = Kp[\chi_n(\eta) = 0|\mathbf{s}_n] \\ 0 & \text{if } p[\chi_n(\eta) = 1|\mathbf{s}_n] < Kp[\chi_n(\eta) = 0|\mathbf{s}_n] \end{cases} \quad (2.20)$$

where $K \geq 0$ is a constant and $g(\mathbf{s}_n)$ an arbitrary function satisfying $0 \leq g(\mathbf{s}_n) \leq 1$ for all \mathbf{s}_n .

In order to arrive at the Neyman-Pearson lemma, we need to specify furthermore the power and the size of a test. In general the *power-function* w of a test is the probability of rejecting a null hypothesis H_0 , when the tested hypothesis H_1 is true. For our purpose, we formulate the hypothesis under study H_1 and the corresponding null hypothesis H_0 as follows:

H_1 : an event will occur in the next step \rightarrow give an alarm

⁶According to [66] this testing for the outcome of binary events was indeed also the initial problem studied by Bayes.

H_0 : no event will occur in the next step \rightarrow no alarm

In our situation rejecting H_0 means to accept H_1 and hence

$$w(\mathbf{u}, \delta, \eta) = Pr[\text{reject } H_0 | \chi_n(\eta) = 1] = Pr[\text{accept } H_1 | \chi_n(\eta) = 1] \quad (2.21)$$

$$= Pr[\mathbf{s}_n \in V(\mathbf{u}, \delta)] \quad (2.22)$$

$$= \int_{V(\mathbf{u}, \delta)} d\mathbf{s}_n \rho[\mathbf{s}_n | \chi_n(\eta) = 1] = r_c(\mathbf{u}, \delta, \eta), \quad (2.23)$$

i.e., the power-function of a test corresponds in the context of forecasting to the rate of correct predictions r_c , which we will discuss in more detail in Subsec. 2.3.4. The size b of a test corresponds in this context to a fixed upper limit for the rate of false alarms r_f

$$Pr[\text{reject } H_0 | \chi_n(\eta) = 0] = Pr[\text{accept } H_1 | \chi_n(\eta) = 0] \leq b \quad (2.24)$$

$$Pr[\text{reject } H_0 | \chi_n(\eta) = 0] = \int_{V(\mathbf{u}, \delta)} d\mathbf{s}_n \rho[\mathbf{s}_n | \chi_n(\eta) = 0] = r_f(\mathbf{u}, \delta, \eta) \leq b. \quad (2.25)$$

The Neyman-Pearson lemma for simple hypothesis then states:

Theorem 1 (Neyman-Pearson lemma for simple hypothesis)

Optimality: For any K and $g(\mathbf{s}_n)$, the test $\phi_0(\mathbf{s}_n)$ has maximum power among all tests whose sizes b are no greater than the size of ϕ_0 .

Existence: Given $b \in (0, 1)$, there exist constants K and g_0 such that the likelihood ratio test defined by this K and $g(\mathbf{s}_n) = g_0$ for all \mathbf{s}_n has exactly the size b .

Uniqueness: If the test ϕ has size b , and is of maximum power among all possible tests of size b , then ϕ is necessarily a likelihood ratio test, except possibly on a set of values of \mathbf{s}_n which has probability 0 under both H_0 and H_1 .

A proof and further information about this theorem can, e.g., be found in [64] or in the original work of Neyman and Pearson from 1933 [69]. Thus according to the Neyman-Pearson theorem the strategy, which maximizes $\Lambda(\mathbf{s}_n)$ for a fixed rate of false alarms is the optimal strategy. One can see that strategy II is more suitable to maximize the likelihood ratio, by expressing the likelihood ratio in terms of the CPDFs, which enter into strategy I and II. Considering strategy II which consists in maximizing $p[\chi_n(\eta) = 1 | \mathbf{s}_n]$ we obtain by expressing the posterior PDF through Bayes' theorem

$$\Lambda[\mathbf{s}_n, \chi_n(\eta)] = \frac{p[\chi_n(\eta) = 1 | \mathbf{s}_n] \rho(\mathbf{s}_n)}{(1 - p[\chi_n(\eta) = 1 | \mathbf{s}_n]) \rho(\mathbf{s}_n)} \frac{(1 - p[\chi_n(\eta) = 1])}{p[\chi_n(\eta) = 1]} \quad (2.26)$$

$$= \frac{p[\chi_n(\eta) = 1 | \mathbf{s}_n]}{(1 - p[\chi_n(\eta) = 1 | \mathbf{s}_n])} \frac{(1 - p[\chi_n(\eta) = 1])}{p[\chi_n(\eta) = 1]} \quad (2.27)$$

$$= \Lambda(p[\chi_n(\eta) = 1 | \mathbf{s}_n], p[\chi_n(\eta) = 1]), \quad (2.28)$$

since $p[\mathbf{s}_n | \chi_n(\eta) = 0] = 1 - p[\mathbf{s}_n | \chi_n(\eta) = 1]$ holds. The probability $p[\chi_n(\eta) = 1]$ to find events is given by the process and cannot be influenced by the forecaster. Thus we find that the likelihood ratio can be maximized by maximizing $p[\chi_n(\eta) = 1 | \mathbf{s}_n]$ for a fixed $p[\chi_n(\eta) = 1]$. Expressing $\Lambda(\mathbf{s}_n)$ in terms of the posterior PDF (strategy I) reads

$$\Lambda[\mathbf{s}_n, \chi(\eta)] = \frac{\rho[\mathbf{s}_n | \chi_n(\eta) = 1]}{(1 - p[\chi_n(\eta) = 1 | \mathbf{s}_n])} \frac{(1 - p[\chi_n(\eta) = 1])}{\rho(\mathbf{s}_n)} \quad (2.29)$$

$$= \Lambda(\rho[\mathbf{s}_n | \chi_n(\eta) = 1], p[\chi_n(\eta) = 1 | \mathbf{s}_n], \rho(x_n), p[\chi_n(\eta) = 1]), \quad (2.30)$$

$$\text{since } \rho[\mathbf{s}_n | \chi_n(\eta) = 0] \neq 1 - \rho[\mathbf{s}_n | \chi_n(\eta) = 1], \quad (2.31)$$

$$\text{but } \rho[\mathbf{s}_n | \chi_n(\eta) = 0] = \frac{\rho(x_n)(1 - p[\chi_n(\eta) = 1 | \mathbf{s}_n])}{1 - p[\chi_n(\eta) = 1]}. \quad (2.32)$$

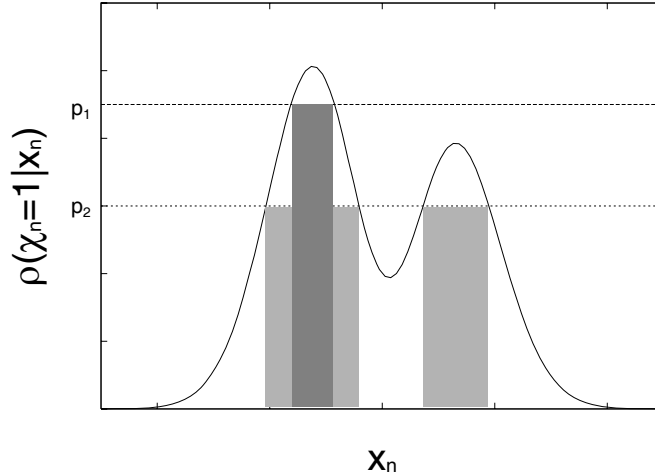


Figure 2.4: A multimodal CPDF leads to multiple precursors. In this cases one could also think about imposing probability thresholds p_1 and p_2 on the CPDF instead of constructing alarm volumes around the precursors. If the CPDF is symmetric with respect to its maxima the respective alarm volumes correspond exactly to the use of a probability threshold.

Reminding ourselves of the fact that $p[\chi_n(\eta) = 1]$ is determined by the problem under study and thus not under the influence of the forecaster, we see that the likelihood ratio is a monotonously increasing function of $p[\chi_n(\eta) = 1|\mathbf{s}_n]$ in the first case (Eq.2.28), whereas it is a function of both CPDFs in Eq.2.30. Therefore, strategy I can at best be equal to strategy II, in the special case that the precursor, which maximizes $\rho[\mathbf{s}_n|\chi_n(\eta) = 1]$ simultaneously maximizes $p[\chi_n(\eta) = 1|\mathbf{s}_n]$ and minimizes $\rho(x_n)$.

The *Bayes classifier* [70] (*posterior odds* [64], *ratio of odds*[71]) is a very similar concept, since it consists in our notation of the ratio of likelihoods⁷

$$B[\mathbf{s}_n, \chi(\eta)] = \frac{p[\chi_n(\eta) = 1|\mathbf{s}_n]}{p[\chi_n(\eta) = 0|\mathbf{s}_n]} = \frac{(1 - p[\chi_n(\eta) = 1])}{p[\chi_n(\eta) = 1]} \Lambda[\mathbf{s}_n, \chi_n(\eta)], \quad (2.33)$$

where $(1 - p[\chi_n(\eta) = 1]) / p[\chi_n(\eta) = 1]$ is called *prior odds* and the likelihood ratio $\Lambda(\mathbf{s}_n)$ is here referred to as *Bayes factor*. Since we assume the prior PDFs $p[\chi_n(\eta) = 1]$ and $p[\chi_n(\eta) = 0] = (1 - p[\chi_n(\eta) = 1])$ to be given by the process, and not under the influence of the forecaster, it is easy to see that the prediction strategy, which focus on maximizing the likelihood ratio also maximizes the Bayes classifier. In general, i.e., if subjective Bayesian estimates of the prior probabilities $p[\chi_n(\eta) = 1]$ and $p[\chi_n(\eta) = 0]$ are used, this it not necessarily the case.

In the definition of the precursors according to strategy I and II, we already took into account, that there might be more than one choice for a precursor which maximizes the CPDF, if the respective CPDF is multimodal, see e.g., Fig. 2.4. In this cases one could also think about imposing a threshold on the CPDF and give an alarm if the CPDF which corresponds to an observed precursory variable is larger then this threshold, instead of constructing alarm volumes around the precursors. Our approach of predicting via precursors is thus replaced by a probabilistic forecast. The decision variable in terms of a probability threshold $1 - \delta$ is then defined as

$$\hat{A}(\mathbf{s}_n, \delta, \eta) = \begin{cases} 1 & : \text{ if } p[\chi_n(\eta) = 1|\mathbf{s}_n] \geq 1 - \delta, \\ 0 & : \text{ if } p[\chi_n(\eta) = 1|\mathbf{s}_n] \leq 1 - \delta. \end{cases} \quad (2.34)$$

If the CPDF is symmetric with respect to its maxima the respective alarm volumes correspond exactly to the use of a probability threshold. Thus one can think of the use of precursors defined

⁷Keep in mind that in our notation the *likelihood ratio* is in fact a ratio of posterior PDF.

by maximizing CPDFs as an (numerically faster) approximation to the usage of the full CPDF, as it is done in probabilistic forecast.

decision variable A_n	Observation in terms of the event variable χ_n	
	event occurred, $\chi_n = 1$ marginal PDF $p[\chi_n = 1]$	no event occurred, $\chi_n = 0$ marginal PDF $p[\chi_n = 0]$
alarm for an event $A_n = 1$	hit (correctly predicted event) $r_c(\mathbf{s}_n, \mathbf{u}, \delta) = p[\mathbf{s}_n \in V(\mathbf{u}, \delta) \chi_n = 1]$	false alarm $r_f(\mathbf{s}_n, \mathbf{u}, \delta) = p[\mathbf{s}_n \in V(\mathbf{u}, \delta) \chi_n = 0]$
no alarm given $A_n = 0$	missed event $p[\mathbf{s}_n \notin V(\mathbf{u}, \delta) \chi_n = 1]$	correct rejection $p[\mathbf{s}_n \notin V(\mathbf{u}, \delta) \chi_n = 0]$

Table 2.1: The four possible outcomes of a binary forecast of events and the corresponding probabilities which describe them. The event variable χ_n and the decision variable A_n are defined in Eqs. (2.1) and (2.8). The variables r_c and r_f denote the the rates of correct predictions and false alarms.

2.3 Measuring the Quality of a Prediction

In this section, we address the question of how to quantify the performance of a forecast. Measures which try to judge the quality of a prediction are important not only in order to rank existing forecast schemes but also in the design of such schemes, for example in the tuning of free parameters. Such measures have been formulated for a variety of purposes and for a multitude of different situations[72]. For example the book of Jolliffe and Stephenson [73] discusses about 30 of them. In general we can distinguish between measures which understand the predictability as a property inherent to the system under study (*intrinsic measures of predictability*) and *verification measures* [72] which compare the results of a forecast with the observed outcome, the *verification*. We will mostly focus on the latter. Furthermore we introduce the idea of transforming measures which are traditionally considered to be intrinsic measures into verification measures.

The decision which verification measure to use is among other considerations influenced by the type of forecast, e.g., whether we forecast a specific value of an observable, e.g., the temperature forecasted by a weather model, or whether we simply do a *binary forecast (yes/no forecast)*, as it is the case for the prediction of (extreme) events. When forecasting a specific value, one typically tries to be “good in the average”, i.e., it is not important, if a single large mistake is hidden among sufficiently many good forecasts. The probably most frequently used verification method, the root mean square error $\bar{e} = \sqrt{\frac{1}{N} \sum_{n=1}^N (f_n - x_n)^2}$, with f_n representing the forecasted value and x_n the corresponding verification works according to this principle. However, the forecast of events, especially the forecast of large impact events, demands detailed information about the contribution of correctly predicted events, missed events, correctly predicted non-events and false alarms. Table 2.1 illustrates these four possible outcomes of a categorical forecast of events by using the event variable χ_n as defined in Eq. (2.1) and the decision variable A_n as defined in Eq. (2.8).⁸

In many applications it is furthermore important to weight these four options according to the loss which is associated with the occurrence of each of them.⁹ One then tries to find a prediction strategy which optimizes the corresponding *loss matrix*. Since we are not interested in a certain application, we will not consider losses and costs, but simply measure the quality of the prediction.

⁸Note that the definition of the rates r_c and r_f for probabilistic forecasts are used, i.e., if the decision variable was defined according to Eq.2.34.

⁹Note that losses and costs can in this context not only be counted in terms of money, but also in terms of e.g., injured humans

2.3.1 Predictability, Kullback Leibler distance and Mutual Information

Loosely speaking, the term *predictability* is used as a synonym for the quality of a forecast. However, it is worth mentioning, that in fact the term predictability is properly defined as a function of entropies. The predictability $P(t, \kappa)$ at a time t for future time $t + \kappa$ is defined as [74]

$$P(\kappa) = \frac{R(x(t), x(t + \kappa))}{H(x(t))}, \quad (2.35)$$

where $H(x(t)) = -\sum_x p(x(t)) \log(p(x))$ or respectively $H(x(t)) = -\int dx \rho(x(t)) \log(\rho(x))$ is the entropy of a marginal PDF and $R(x(t)) = H(x(t)) + H(x(t + \kappa)) - 2H(x(t), x(t + \kappa))$ denotes the redundancy. Typically entropies are evaluated for the marginal PDFs and the summation/integration is over all possible values of $x(t)$, which represent usually all accessible states of a system. Thus a function based on entropies provides us with information about the *possible* predictability of the system under study, without considering exactly how this prediction can be made. In other words, the predictability is assumed to be a property of the system under study and does not result from the interaction with the system and a prediction method. In particular it does not reflect the quality of a prediction algorithm.

Except from the predictability as defined above there are various other similar measures, which can specify the tendency of a system to reveal information about its future or other uncertain aspects. For example, if the dynamics of the system is sufficiently known, the time until which the system is expected to be predictable is often characterized by the inverse of the Lyapunov exponent.

However, we saw before that we can easily combine the forecast and the verification in terms of a joint PDF [72]. Hence it is in principle also possible to evaluate intrinsic measures based on probabilities on the joint PDF of precursor and event, or the respective CPDFs. Predictability measures which are in particular suitable for this are *Mutual Information* and the *Kullback Leibler distance*[75], also called *relative entropy*, which is often interpreted as a measure for the “distance” between two distributions. Another interpretation describes it as “*the inefficiency of assuming that the distribution is q when the true distribution is p* ” [75]. The Kullback Leibler distance is defined as

$$D(p||q) = \sum_x p(x) \log \left(\frac{p(x)}{q(x)} \right), \quad (2.36)$$

$$\text{or } D(p||q) = \int dx p(x) \log \left(\frac{p(x)}{q(x)} \right), \quad (2.37)$$

for probability densities.

Considering the Kullback Leibner distance of the a posteriori probabilities of an event and a non-event

$$D(\rho[\mathbf{s}_n|\chi_n(\eta) = 1] || \rho[\mathbf{s}_n|\chi_n(\eta) = 0]) = \int d\mathbf{s}_n \rho[\mathbf{s}_n|\chi_n(\eta) = 1] \log \left(\frac{\rho[\mathbf{s}_n|\chi_n(\eta) = 1]}{\rho[\mathbf{s}_n|\chi_n(\eta) = 0]} \right). \quad (2.38)$$

one can see that the Kullback Leibler distance is in fact the expected value of the logarithm of the likelihood ratio [75] with respect to the PDF $\rho[\mathbf{s}_n|\chi_n(\eta) = 1]$. In Subsec. 2.3.4 we will see that the likelihood ratio itself is closely linked with the ROC curve, which is not an intrinsic measure for predictability but a forecast verification methods. Hence, due to its link with the likelihood ratio, the Kullback Leibler distance might serve as method to related information obtained from the system under study and its dynamics and forecast verification methods.

In case that we are interested in using the joint PDF, rather than the CPDFs of precursors and events, we can compute the *mutual information*, which is defined as the relative entropy between the joint distribution and the product of the marginal distributions of two random

variables X and Y . Hence we can easily express the mutual information between event and precursor as follows

$$\begin{aligned} M(\mathbf{s}_n, \chi_n(\eta)) &= \int_{\mathbf{s}_n} d\mathbf{s}_n p[\mathbf{s}_n, \chi_n(\eta) = 1] \log \frac{p[\mathbf{s}_n, \chi_n(\eta) = 1]}{\rho[\mathbf{s}_n]p[\chi_n(\eta) = 1]} \\ &+ \int_{\mathbf{s}_n} d\mathbf{s}_n p[\mathbf{s}_n, \chi_n(\eta) = 0] \log \frac{p[\mathbf{s}_n, \chi_n(\eta) = 0]}{\rho[\mathbf{s}_n]p[\chi_n(\eta) = 0]}. \end{aligned} \quad (2.39)$$

2.3.2 Forecast Verification

In contrast to the measures discussed in the previous section, forecast verification consists by definition in comparing a forecast with the corresponding observation. Thus the quality of a single forecast can only be evaluated retrospectively, when the forecasted event was observed or not observed. From a practical point of view, in the moment in which the knowledge about the observation is available, the forecast also loses its value. However, by comparing many pairs of forecasts and observations one can derive information about the overall quality of a forecast method.

The framework proposed by Murphy [72] is based on the joint probability of forecast and observations. Based on this joint PDF and its possible factorizations in terms of marginal PDFs and CPDFs one can then derive properties, which one hopes a probabilistic forecast \hat{p} to fulfill. Note that it is the definition of the decision variable as shown in Eqn. 2.8 and 2.34 which makes a distinction between a forecast based on the full CPDF of a forecast based on the maximum of the CPDF (the precursor). As we discussed in Sec. 2.2.3, forecasts made by precursors which are identified by CPDs, can thus be interpreted as an approximation to a full probabilistic forecast.

Intuitively, one would hope that $\hat{p}(\mathbf{s}_n)$ gives the probability of $\chi_n(\eta) = 1$ given \mathbf{s}_n , or

$$\hat{p}(\mathbf{s}_n) = p[\chi_n(\eta) = 1 | \mathbf{s}_n]. \quad (2.40)$$

Many reasonable measures of forecast success support this intuition, that is, they give maximum possible scores if $\hat{p}(\mathbf{s}_n)$ indeed agrees with the probability of $\chi_n(\eta) = 1$ given \mathbf{s}_n .

A seemingly different way to motivate $p[\chi_n(\eta) = 1 | \mathbf{s}_n]$ as a good forecast probability is through *reliability*. Reliability means that on condition that the relative frequency of alarms (approximately) equals z , the event should occur with a relative frequency (approximately) equal to z , too. As an optimality criterion, reliability is not sufficient to single out a particular forecasting scheme, since *any* conditional probability of the form $p(\chi_n(\eta) = 1 | I)$ is reliable, independent of what I is. In particular, the unconditional probability $\hat{p}(\mathbf{s}_n) = \text{const.} = p[\chi_n(\eta) = 1]$ is reliable as well. Hence, in addition to reliability, the forecast should feature a high correlation with the actual event. This property is known as *sharpness*. It can be demonstrated (also as a consequence from the Neyman-Pearson theorem), that $p[\chi_n(\eta) = 1 | \mathbf{s}_n]$ is indeed the reliable forecast which features maximum sharpness among all functions of \mathbf{s}_n .

Furthermore one can propose so called Scoring schemes, which evaluate a scoring function on every pair of forecast and observation and then average in order to obtain information about the overall quality of the forecast method. In the following Section we will discuss the Brier score and the ignorance as examples for Scores.

2.3.3 Brier Score and Ignorance

A score is a ‘‘point-wise’’ (evaluated at every single time instance) measure of performance. It quantifies the success of individual forecast instances by comparing the forecasts \hat{p}_n and the observations $\chi_n(\eta)$ for every n . The general quality of a forecasting system is commonly measured by the average score $E[S(\hat{p}, \chi_n(\eta))]$, which can be estimated from the empirical mean

$$E[S(\hat{p}, \chi_n(\eta))] \cong \frac{1}{N} \sum_{i=1}^N S(\hat{p}_i, \chi_n(\eta)) \quad (2.41)$$

over a sufficiently large data set $(\hat{p}_i, \chi_n(\eta))$. A *scoring rule* [76, 77, 78, 79] is a function $S(\hat{p}, \chi_n(\eta))$ which effectively defines two functions $S(\hat{p}|\chi_n(\eta) = 1)$ and $S(\hat{p}|\chi_n(\eta) = 0)$ quantifying the score in case the forecast is \hat{p} and the event happens or does not happen. Two important examples are the ignorance score[80], given by the scoring rule

$$S(\hat{p}, \chi_n(\eta)) := -\log(\hat{p}) \cdot \chi_n(\eta) - \log(1 - \hat{p}) \cdot (1 - \chi_n(\eta)), \quad (2.42)$$

and the Brier score[81], given by the scoring rule

$$S(\hat{p}, \chi_n(\eta)) := (\chi_n(\eta) - \hat{p})^2 = (1 - \hat{p})^2 \cdot \chi_n(\eta) + \hat{p}^2 \cdot (1 - \chi_n(\eta)). \quad (2.43)$$

Both definitions imply the convention that a smaller score indicates a better forecast.

The rationale behind both scoring rules mentioned is rather obvious. If the event occurs, the score should become better (i.e., decrease) with increasing \hat{p} , while if it does not occur, the score should become worse (i.e. *increase*) with increasing \hat{p} . But why then not taking just $1 - \hat{p}$ if the event occurs, and \hat{p} if it does not? To see the problem with this “linear” scoring rule, define the *scoring function*

$$s(\hat{p}, q) := S(\hat{p}, 1) \cdot q + S(\hat{p}, 0) \cdot (1 - q) \quad (2.44)$$

where q is another probability, i. e. , a number in the unit interval. Note that the scoring function is the score averaged over cases where the forecast is \hat{p} but in fact q is the true distribution of χ . In view of the interpretation of the scoring function, it seems reasonable to require that the average score of the forecast \hat{p} should be best (i.e. minimal) if and only if \hat{p} in fact coincides with the true distribution of χ . This means that the *divergence function*

$$d(\hat{p}, q) := s(\hat{p}, q) - s(q, q) \quad (2.45)$$

has to be positive definite, i.e., it has to be nonnegative, and zero only if $\hat{p} = q$. A scoring rule with the corresponding divergence function having this property is called *strictly proper* [78, 82]. The divergence function of the Brier score for example is $d(\hat{p}, q) := (\hat{p} - q)^2$, demonstrating that this score is strictly proper. While the ignorance is proper as well, the linear score though is easily shown to be *improper*. In total, scores are designed to evaluate the forecast of multiple values. It is not in general obvious that we can apply them to evaluate binary forecast.

2.3.4 The Receiver Operating Characteristic and the Likelihood Ratio

A common method to verify a hypothesis or test the quality of a prediction via precursory structures is the receiver operating characteristic curve (ROC curve) [83, 84]. In the 1980s it became popular for medical diagnostic testing, nowadays there are many other fields of application as well. The idea of the ROC curve consists simply in comparing the rate of correctly predicted events r_c (*hit rate, rate of true positives, sensitivity*) with the rate of false alarms r_f (*rate of false positives, (1-specificity)*) by plotting r_c vs. r_f as it is illustrated in Fig. 2.5. Numerically, these rates can be computed from the original time series $\{x_n\}$ and the time series of the events $\{\chi_n(\eta)\}$ by simple counting. In order to calculate r_c and r_f analytically, we have to consider the respective definition of the decision variables, either in terms of the alarm volume according to Eq. (2.8) or in terms of a probability threshold according to Eq. (2.34). In the context of predictions with precursory structures, these rates of correct and false alarms can be calculated from the posterior PDFs of events and non-events

$$r_c(\mathbf{u}, \eta, \delta) = \int_{V_{pre}(\mathbf{u}, \delta)} d\mathbf{s}_n \rho[\mathbf{s}_n | \chi_n(\eta) = 1], \quad (2.46)$$

$$\text{and } r_f(\mathbf{u}, \eta, \delta) = \int_{V_{pre}(\mathbf{u}, \delta)} d\mathbf{s}_n \rho[\mathbf{s}_n | \chi_n(\eta) = 0], \quad (2.47)$$

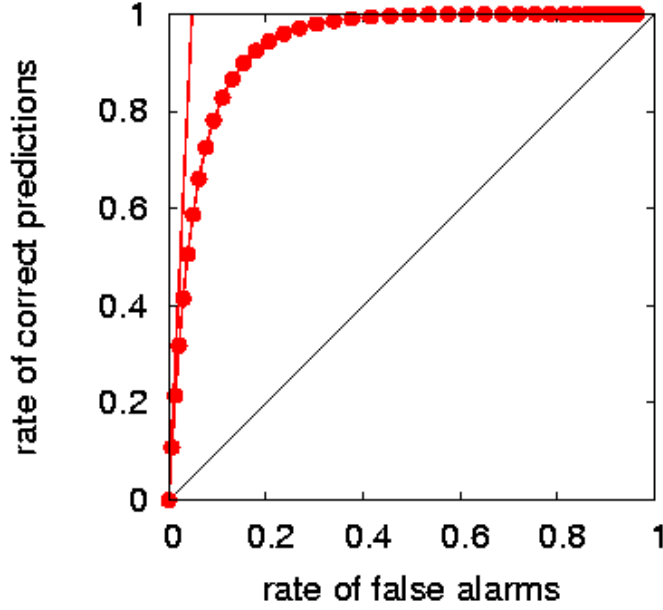


Figure 2.5: An example for an ROC curve and the slope of the ROC curve in the vicinity of the origin, i.e., the likelihood ratio.

where $V(\mathbf{u}, \delta)$ is the alarm volume as defined in Sec. 2.2

If the definition of the decision variable according to Eq. (2.34) is applied, then the rates of correct prediction and the rate of false alarms are given by the following expressions:

$$\hat{r}_c(\eta, \delta) = \int_{\mathbf{s}_n: p[\chi_n(\eta)=1|\mathbf{s}_n] \geq 1-\delta} d\mathbf{s}_n \rho[\mathbf{s}_n | \chi_n(\eta) = 1], \quad (2.48)$$

$$\text{and } \hat{r}_f(\eta, \delta) = \int_{\mathbf{s}_n: p[\chi_n(\eta)=1|\mathbf{s}_n] \geq 1-\delta} d\mathbf{s}_n \rho[\mathbf{s}_n | \chi_n(\eta) = 0], \quad (2.49)$$

which can only be evaluated analytically if the regions for which $p[\chi_n(\eta) = 1 | \mathbf{s}_n] \geq \delta$ holds are simply connected.

Thus for each value of δ one obtains a point in the ROC-coordinates and if δ is assumed to be a continuous variable one arrives at a curve parametrized by δ . The resulting curve in the unit-square of the r_f - r_c plane approaches the origin for $\delta \rightarrow 0$ and the point (1,1) in the limit $\delta \rightarrow \infty$. It follows readily from the definitions that both r_c and r_f are monotonously decreasing functions of the size of the tolerance volume δ with limits 0 for decreasing δ and 1 for increasing δ . Hence, the ROC curve is a monotonously *increasing* arc connecting the points (0,0) and (1,1).

A ROC curve with a monotonously decreasing slope (a convex ROC) is called proper ROC curve [84].

The shape of the curve characterizes the significance of the prediction. A curve above the diagonal reveals that the corresponding strategy of prediction is better than a random prediction which is characterized by the diagonal.

If we have to compare ROCs created from two different decision variables A and \tilde{A} , then it is not obvious to decide which is “the better” ROC as the two curves might cross. This is a problem if a criterion is required in order to optimize a prediction algorithm, in particular, when we search for optimal precursors. Hence, *summary indices* [85] of ROC curves are needed, for example the following:

Proximity to (0,1) (Kolmogorov Smirnov Distance): A good ROC should be close to the point (0,1), that is where the false-alarm rate is zero while the hit rate is 1. The point closest to (0,1) would simultaneously define an operation point for the algorithm.

Area under ROC curve: The area under the ROC curve (AUC) is a well established summary index for ROC curves, which should be maximal. It can be shown that this quantity gives the probability that on an instance when the event takes place, the decision variable is actually larger than on an independent instance when the event does not take place. It is a global quantity, averaging over all alarm rates.

Maximal Hit Rate for fixed Alarm Volume: Optimizing ROC for precursors by asking for a maximal hit rate without any further constraints is not a useful criterion, since all decision variables have a maximum hit rate of 1, achievable by just giving always alarms. Fixing the alarm volume this criterion leads to precursors according to strategy I of Sect. 3. Note that the false alarm rate is not considered at all in such an optimization, so that the optimal hit rate for fixed alarm rate might be achieved at the cost of an unreasonably large false alarm rate. Inverting the criterion to minimizing the false alarm rate for fixed alarm volume leads to the same precursor.

Ratio of Hit Rate and False Alarm Rate (Likelihood Ratio): A maximum ratio of hit rate versus false alarm rate in the limit of small false alarm rates yields another well established summary index, the slope m of the ROC curve at the vicinity of the origin, which we can identify as the likelihood ratio. The slope of the ROC curve is given by

$$m[\mathbf{u}, \delta, \chi(\eta)] = \frac{dr_c}{dr_f} = \frac{\frac{\partial r_c[\mathbf{u}, \delta, \chi(\eta)]}{\partial \delta}}{\frac{\partial r_f[\mathbf{u}, \delta, \chi(\eta)]}{\partial \delta}} \quad (2.50)$$

Expressing the partial derivatives of the rates by their definition, i.e.,

$$\frac{\partial r_c[\mathbf{u}, \delta, \chi(\eta)]}{\partial \delta} = \lim_{h \rightarrow 0} \frac{r_c[\mathbf{u}, \delta + h, \chi(\eta)] - r_c[\mathbf{u}, \delta, \chi(\eta)]}{h} \quad (2.51)$$

and then considering

$$\lim_{x \rightarrow p} \frac{f(x)}{g(x)} = \frac{\lim_{x \rightarrow p} f(x)}{\lim_{x \rightarrow p} g(x)}, \quad (2.52)$$

(see e.g., [86]) the slope of the ROC curve reads

$$m[\mathbf{u}, \delta, \chi(\eta)] = \lim_{h \rightarrow 0} \frac{r_c[\mathbf{u}, \delta + h, \chi(\eta)] - r_c[\mathbf{u}, \delta, \chi(\eta)]}{r_f[\mathbf{u}, \delta + h, \chi(\eta)] - r_f[\mathbf{u}, \delta, \chi(\eta)]}. \quad (2.53)$$

Assuming that the Euclidean norm was used to define the alarm Volumina used in Eqs. (2.46) and (2.47), the rates of correct prediction and false alarm for the decision variable according to Eq. (2.8) are given by

$$r_c(\mathbf{u}, \eta, \delta) = \int_{u_{\tau-1}-\delta/2}^{u_{\tau-1}+\delta/2} dx_{n-\tau+1} \int_{u_{\tau-2}-\delta/2}^{u_{\tau-2}+\delta/2} dx_{n-\tau+2} \dots \int_{u_0-\delta/2}^{u_0+\delta/2} dx_n \rho[\mathbf{s}_n | \chi_n(\eta) = 1], \quad (2.54)$$

$$r_f(\mathbf{u}, \eta, \delta) = \int_{u_{\tau-1}-\delta/2}^{u_{\tau-1}+\delta/2} dx_{n-\tau+1} \int_{u_{\tau-2}-\delta/2}^{u_{\tau-2}+\delta/2} dx_{n-\tau+2} \dots \int_{u_0-\delta/2}^{u_0+\delta/2} dx_n \rho[\mathbf{s}_n | \chi_n(\eta) = 0], \quad (2.55)$$

with $\mathbf{s}_n = (x_{n-\tau+1}, x_{n-\tau+2}, \dots, x_n)$.

Substituting with $\mathbf{y} = \mathbf{s}_n + \frac{1}{2}\vec{\delta}$, with the τ -dimensional vector $\vec{\delta} = (\delta, \delta, \dots, \delta)$ yields

$$r_c(\mathbf{u}, \eta, \delta) = \int_{u_{\tau-1}}^{u_{\tau-1}+\delta} dy_{n-\tau+1} \int_{u_{\tau-2}}^{u_{\tau-2}+\delta} dy_{n-\tau+2} \dots \int_{u_0}^{u_0+\delta} dy_n \rho \left[\mathbf{y} - \frac{1}{2}\vec{\delta} \mid \chi_n(\eta) = 1 \right], \quad (2.56)$$

$$r_f(\mathbf{u}, \eta, \delta) = \int_{u_{\tau-1}}^{u_{\tau-1}+\delta} dy_{n-\tau+1} \int_{u_{\tau-2}}^{u_{\tau-2}+\delta} dy_{n-\tau+2} \dots \int_{u_0}^{u_0+\delta} dy_n \rho \left[\mathbf{y} - \frac{1}{2}\vec{\delta} \mid \chi_n(\eta) = 0 \right], \quad (2.57)$$

which allows us to express the difference $r_c[\mathbf{u}, \delta + h, \chi(\eta)] - r_c[\mathbf{u}, \delta, \chi(\eta)]$ as

$$\begin{aligned} & r_c[\mathbf{u}, \delta + h, \chi(\eta)] - r_c[\mathbf{u}, \delta, \chi(\eta)] = \\ & = \int_{u_{\tau-1}+\delta}^{u_{\tau-1}+\delta+h} dy_{n-\tau+1} \int_{u_{\tau-2}+\delta}^{u_{\tau-2}+\delta+h} dy_{n-\tau+2} \dots \int_{u_0+\delta}^{u_0+\delta+h} dy_n \left[\mathbf{y} - \frac{1}{2}\vec{\delta} \mid \chi_n(\eta) = 1 \right], \end{aligned} \quad (2.58)$$

and the analog for $r_f[\mathbf{u}, \delta + h, \chi(\eta)] - r_f[\mathbf{u}, \delta, \chi(\eta)]$. Using the mean value theorem of the integral calculus, we can express the difference as

$$\begin{aligned} r_c[\mathbf{u}, \delta + h, \chi(\eta)] - r_c[\mathbf{u}, \delta, \chi(\eta)] &= h^\tau \rho \left[\mathbf{u} + \frac{1}{2}\vec{\delta} + \vec{\xi} \mid \chi_n(\eta) = 1 \right], \quad (2.59) \\ \text{with } \vec{\xi} &= (\xi_1, \xi_2, \dots, \xi_\tau), \quad (u_{i-1} + \delta < \xi_i < u_{i-1} + \delta + h), \quad (i = 1, 2, \dots, \tau), \end{aligned}$$

and the analog for $r_f[\mathbf{u}, \delta + h, \chi(\eta)] - r_f[\mathbf{u}, \delta, \chi(\eta)]$. Hence we obtain the following expression for the slope of the ROC curve

$$\begin{aligned} m[\mathbf{u}, \vec{\delta}, \chi(\eta)] &= \lim_{h \rightarrow 0} \frac{\rho \left[\mathbf{u} + \frac{1}{2}\vec{\delta} + \xi^c \mid \chi_n(\eta) = 1 \right]}{\rho \left[\mathbf{u} + \frac{1}{2}\vec{\delta} + \xi^f \mid \chi_n(\eta) = 0 \right]} = \frac{\rho \left[\mathbf{u} + \frac{1}{2}\vec{\delta} \mid \chi_n(\eta) = 1 \right]}{\rho \left[\mathbf{u} + \frac{1}{2}\vec{\delta} \mid \chi_n(\eta) = 0 \right]} \quad (2.60) \\ \text{with } \xi^c &= (\xi_1^c, \xi_2^c, \dots, \xi_\tau^c), \quad (u_{i-1} + \delta < \xi_i^c < u_{i-1} + \delta + h), \quad (i = 1, 2, \dots, \tau), \\ \text{and } \xi^f &= (\xi_1^f, \xi_2^f, \dots, \xi_\tau^f), \quad (u_{i-1} + \delta < \xi_i^f < u_{i-1} + \delta + h), \quad (i = 1, 2, \dots, \tau). \end{aligned}$$

Comparing with Sec. 2.2.3 we see, that this expression is indeed the likelihood ratio

$$m[\mathbf{u}, \vec{\delta}, \chi(\eta)] = \frac{\rho \left[\mathbf{u} + \frac{1}{2}\vec{\delta} \mid \chi_n(\eta) = 1 \right]}{\rho \left[\mathbf{u} + \frac{1}{2}\vec{\delta} \mid \chi_n(\eta) = 0 \right]} = \Lambda \left[\mathbf{u} + \frac{1}{2}\vec{\delta}, \chi(\eta) \right]. \quad (2.61)$$

In [87] it is shown that this relation also holds for tolerance volumes, which are not connected, i.e., the tolerance volume according to Eq.(2.34).

We will use the slope in the vicinity of the origin of the ROC curve, i.e. in the limit of small tolerance volumes

$$m[\mathbf{u}, 0, \chi(\eta)] = \frac{\rho[\mathbf{u} \mid \chi_n(\eta) = 1]}{\rho[\mathbf{u} \mid \chi_n(\eta) = 0]} = \Lambda[\mathbf{u}, \chi(\eta)] \quad (2.62)$$

as a summary index for the ROC curve. The slope of the ROC curve decreases monotonously with increasing tolerance volume, to the value of zero in the point (1,1). Thus the slope in the vicinity of the origin does not only corresponds to the region, in which we have a very low rate, but it also provides us with the maximal slope of the ROC curve. Alternatively one could also consider the slope for a fixed size of the alarm volume as a summary index.

2.3.5 Consequences of the Neyman-Pearson Lemma

We saw in the previous subsection, that the shape of an ROC curve can be characterized by the likelihood ratio. From Sec. 2.2 we know that due to the Neyman-Pearson lemma the strategy which maximizes the likelihood ratio is the optimal strategy for a prediction. Furthermore we argued that this optimal strategy is represented by strategy II which consists in predicting with the likelihood $p[\chi_n(\eta) = 1 \mid \mathbf{s}_n]$ or a precursor which maximizes it. Due to the connection of ROC curve and likelihood ratio, we can hence conclude, that strategy II is also the optimal strategy in terms of the ROC curve. This connection was already noticed by Egan in 1975 [84] in his studies about the use of ROCs in signal detection. Egan calls the respective decision rule, which optimizes the likelihood ratio *likelihood ratio observer*. Furthermore Egan found,

that “the ROC for the likelihood observer is never concave upwards”. This can be understood, by considering that the point (0,0) of the ROC corresponds to the maximum of the likelihood ratio. If we then lower the tolerance threshold δ , we move away from this maximum of the likelihood ratio, i.e., the slope of the ROC decreases. In other words, the slope of the ROC is a monotonously decreasing function and hence we can expect to observe a convex ROC or a ROC which corresponds to the diagonal, but not a concave ROC.

Another consequence derived from the likelihood ratio is the existence of monotone likelihood ratio families and corresponding families of ROC curves. Imagine a situation, where the likelihood ratio is given by two distributions which are in addition to the conditioning on the event/non-event parameterized by a variable θ . If we understand the likelihood ratio as a function of \mathbf{s}_n we now have a family of functions $\Lambda(\mathbf{s}_n, \theta)$ parameterized by θ . This situation can be reflected by a family of ROCs. Egan investigated this property for the special case, that the nominator and the denominator of the likelihood ratio are given by distributions which differ only in their means [84]. The resulting family of likelihood ratios and ROC curves are thus parameterized by the relative difference of the means. The main topic of interest in this contribution is very similar, since we study families of ROC curves which are parameterized by the event magnitude η .

Chapter 3

Predicting Increments in AR(1) and long range correlated ARMA Processes

In this chapter we study analytically and numerically predictions in a simple autoregressive process of order 1 [AR(1) process] in order to obtain a detailed understanding of some aspects on precursors and predictions. The AR(1) process is a simple stationary stochastic model process, that might not reflect all features of more complex processes occurring in nature, but it admits an analytic treatment. Additionally, we study similar prediction procedures numerically in long-range correlated data, verifying the same quantitative results. The aspects, which we intend to address are the following:

- A1, superiority of strategy II:** In the previous chapter we argued, that according to the Neyman-Pearson lemma, strategy II should be superior to strategy I. In this chapter we will illustrate this aspect for the example of the AR(1) process.
- A2, influence of the event magnitude:** In the introduction we listed examples from the literature in which a better predictability of larger events was reported. The simple AR(1) process allows us to study this effect analytically and numerically.
- A3, influence of the correlation:** Naturally one would expect a stronger correlation to lead to a better quality of the predictions. However, for events defined as increments, the reasoning has to be more subtle, especially if the forecast is made by statistical considerations. Thus we study the influence of the correlation of the data on the quality of predictions of extreme increments.

3.1 AR(1) process

In this chapter we study the prediction of increments, as defined in Subsec. 2.1.2, in a time-series $\{x_n\}$ which is generated by an auto-regressive model of order 1 [AR(1)] (see, e.g.,[33])

$$x_{n+1} = ax_n + \xi_n, \quad (3.1)$$

where ξ_n are uncorrelated Gaussian random numbers and $-1 < a < 1$ is a constant which represents the coupling strength. The size and the sign of the coupling strength determine whether successive values of x_t are clustered or spread, as illustrated in Fig. 3.1. In the case $a = 0$ the process reduces to uncorrelated random numbers with mean $\mu = 0$ and variance $\sigma^2 = 1$, whereas generally the process is exponentially correlated $\langle x_n x_{n+k} \rangle = |a|^k < 1$.

The joint PDF $p(\mathbf{s}_n, \chi_n(\eta))$ of event and precursory variable is derived from the joint PDF $\rho_j(x_0, x_1, \dots, x_{n+1})$ of the process as described in Subsec. B. The Markov property of the AR(1) process suggests to reduce the vector valued precursory variable \mathbf{s}_n to the scalar value x_n , i.e., $\mathbf{s}_n = x_n$ and $\kappa = \tau = 1$. For a stationary process the conditional PDF $\rho(x_{n+1}|x_n)$ is given by

$$\rho(x_{n+1}|x_n) = \sqrt{\frac{1}{2\pi}} \exp\left(-\frac{1}{2}(x_{n+1} - ax_n)^2\right), \quad (3.2)$$

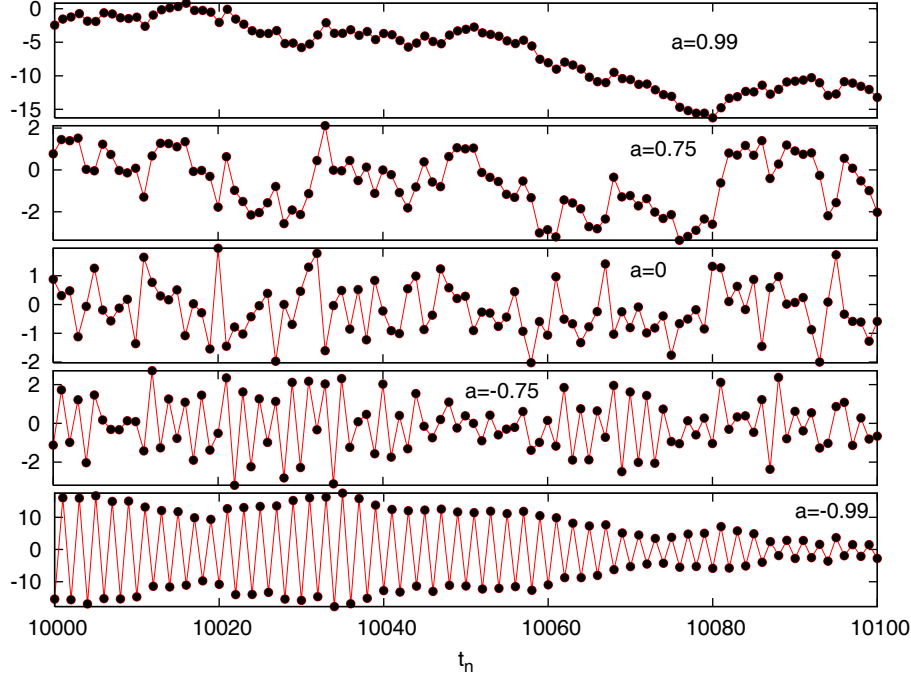


Figure 3.1: Parts of the time series generated by the AR(1) process for different values of a .

since it adapts the distribution of the noise term ξ_n , which has mean ax_n . The marginal probability for one variable x_n of an AR(1) process reads (see e.g., [33])

$$\rho(x_n, a) = \frac{\sqrt{1-a^2}}{\sqrt{2\pi}} \exp\left(-\frac{1-a^2}{2}x_n^2\right). \quad (3.3)$$

Since the magnitude of the events is naturally measured in units of the standard deviation $\sigma(a)$, the relative magnitude of the events η is given by $\eta = \frac{d}{\sigma(a)} = (\sqrt{1-a^2})d$, where d denotes the absolute magnitude of an increment.

Applying the filter mechanism developed in App. B we obtain the a posteriori PDF of extreme events and the joint PDF of non-extreme events as defined in Sec.2.2.3

$$\begin{aligned} p[x_n, \chi_n(\eta) = 1, a] &= \sqrt{\frac{1-a^2}{2\pi}} \exp\left(-\frac{(1-a^2)}{2}x_n^2\right) \\ &\int_0^\infty d\gamma \exp\left(-\frac{1}{2}((1-a)x_n + d + \gamma)^2\right) \\ &= \frac{1}{2}\sqrt{\frac{1-a^2}{2\pi}} \exp\left(-\frac{(1-a^2)}{2}x_n^2\right) \operatorname{erfc}\left(\sqrt{\frac{1}{2}}((1-a)x_n + d)\right). \end{aligned} \quad (3.4)$$

Replacing the absolute event magnitude $d = \frac{\eta}{\sqrt{1-a^2}}$ and normalizing with $p[\chi_n(\eta) = 1, a]$ leads

to the posterior PDFs

$$\rho[x_n|\chi_n(\eta) = 1, a] = \frac{\sqrt{1-a^2}}{2\sqrt{2\pi}p[\chi_n(\eta) = 1, a]} \exp\left(-\frac{1-a^2}{2}x_n^2\right) \cdot \operatorname{erfc}\left(\frac{(1-a)x_n}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right), \text{ and} \quad (3.5)$$

$$\rho[x_n|\chi_n(\eta) = 0, a] = \frac{\sqrt{1-a^2}}{2\sqrt{2\pi}(1-p[\chi_n(\eta) = 1, a])} \exp\left(-\frac{1-a^2}{2}x_n^2\right) \left(1 + \operatorname{erf}\left(\frac{(1-a)x_n}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right)\right). \quad (3.6)$$

Through Bayes' theorem we can calculate the corresponding likelihoods

$$p[\chi_n(\eta) = 1|x_n, a] = \frac{1}{2} \operatorname{erfc}\left(\frac{(1-a)x_n}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right) \quad (3.7)$$

$$p[\chi_n(\eta) = 0|x_n, a] = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{(1-a)x_n}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right)\right) \quad (3.8)$$

The total PDF $p[\chi_n(\eta) = 1, a]$ to find events can only be calculated exactly if the process under study is uncorrelated, i.e., for $a = 0$. (The corresponding calculation for a sequence of Gaussian random numbers will be presented in Sec. 5.2.) For the AR(1) process with $-1 < a < 1$ we can obtain an approximation by using an asymptotic expression for the mean of the posterior PDF,

$$\langle x_n \rangle_{\rho[x_n|\chi_n(\eta)=1, a]} = \int_{-\infty}^{\infty} dx_n x_n \rho[x_n|\chi_n(\eta) = 1, a]. \quad (3.9)$$

Inserting Eq. (3.5) and using the integral representation

$$\begin{aligned} & \int z \exp(-b^2 z^2) \operatorname{erfc}(az + c) dz = \\ &= \frac{-a}{2b^2\sqrt{a^2 + b^2}} \exp\left(-\frac{b^2 c^2}{a^2 + b^2}\right) \operatorname{erf}\left(z\sqrt{a^2 + b^2} + \frac{ac}{\sqrt{a^2 + b^2}}\right) \\ & \quad - \frac{1}{2b^2} \exp(-b^2 z^2) \operatorname{erfc}(az + c) \end{aligned} \quad (3.10)$$

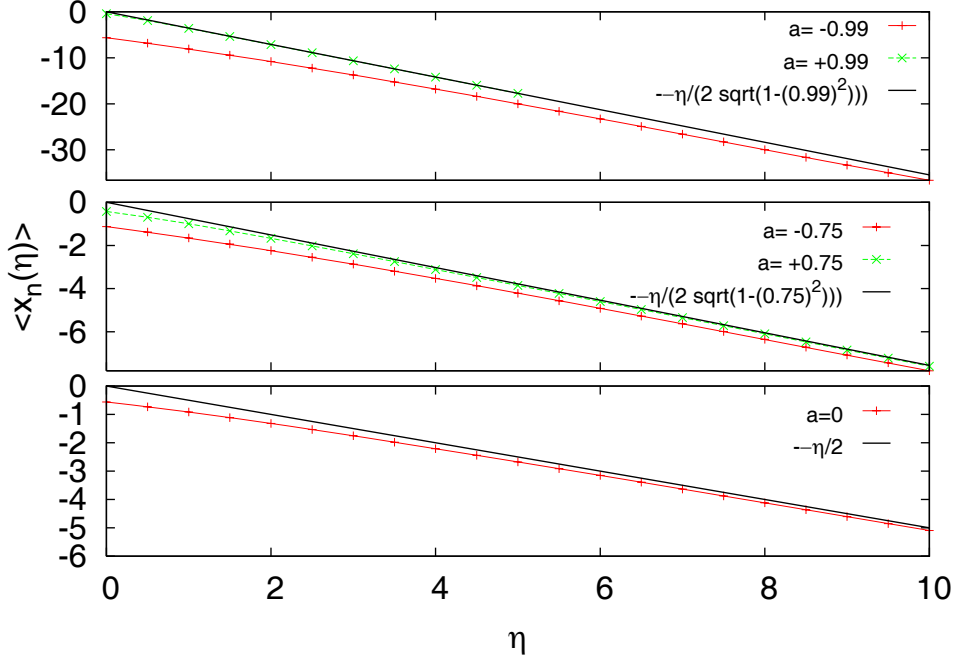


Figure 3.2: The means evaluated numerically by integration of Eq. (3.9) and the asymptotic expression according to Eq. (3.13). The symbols represent the results of the numerical calculation, the black lines represent Eq. (3.13).

from [88] we find that the mean of the posterior PDF is given by

$$\begin{aligned}
\langle x_n \rangle_{\rho[\mathbf{s}_n | \chi_n(\eta)=1]} &= \frac{\sqrt{1-a^2}}{2\sqrt{2\pi} p[\chi_n(\eta)=1, a]} \int_{-\infty}^{\infty} \exp\left(-\frac{1-a^2}{2} x_n^2\right) \\
&\quad \operatorname{erfc}\left(\frac{(1-a)x_n}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right) dx \\
&= -\frac{\sqrt{1-a^2}}{2\sqrt{2\pi} p[\chi_n(\eta)=1, a]} \frac{\sqrt{1-a}}{\sqrt{2}(1-a^2)} \exp\left(-\frac{(1+a)}{4} d^2\right) \\
&\quad \operatorname{erf}\left(x_n \sqrt{1-a} + \frac{\sqrt{1-a}}{2} d\right) \Big|_{-\infty}^{\infty} \\
&\quad - \frac{1}{(1-a^2)} \exp\left(-\frac{(1-a^2)}{2} x_n^2\right) \operatorname{erfc}\left(\frac{(1-a)}{\sqrt{2}} z + \frac{d}{\sqrt{2}}\right) \Big|_{-\infty}^{\infty} \\
&= \frac{-\exp\left(-\frac{(1+a)}{4} d^2\right)}{4\sqrt{\pi}\sqrt{1+a} p[\chi_n(\eta)=1, a]} \operatorname{erf}\left(x_n \sqrt{1-a} + \frac{\sqrt{1-a}}{2} d\right) \Big|_{-\infty}^{\infty} \\
&= \frac{-\exp\left(-\frac{(1+a)}{4} d^2\right)}{4\sqrt{\pi}\sqrt{1+a} p[\chi_n(\eta)=1, a]} (1 - (-1)) \\
&= \frac{-\exp\left(-\frac{\eta^2}{4(1-a)}\right)}{2\sqrt{\pi}\sqrt{1+a} p[\chi_n(\eta)=1, a]}. \tag{3.11}
\end{aligned}$$

We can then formulate the total probability in terms of the mean

$$p[\chi_n(\eta)=1, a] = \frac{-\exp\left(-\frac{\eta^2}{4(1-a)}\right)}{2\sqrt{\pi}\sqrt{1+a} \langle x_n \rangle_{\rho[\mathbf{s}_n | \chi_n(\eta)=1]}} \tag{3.12}$$

Since we do not know the analytic structure of the mean explicitly, we assume, that the mean of the posterior PDFs is similar to its maximum. In principle the maximum of a PDF equals the

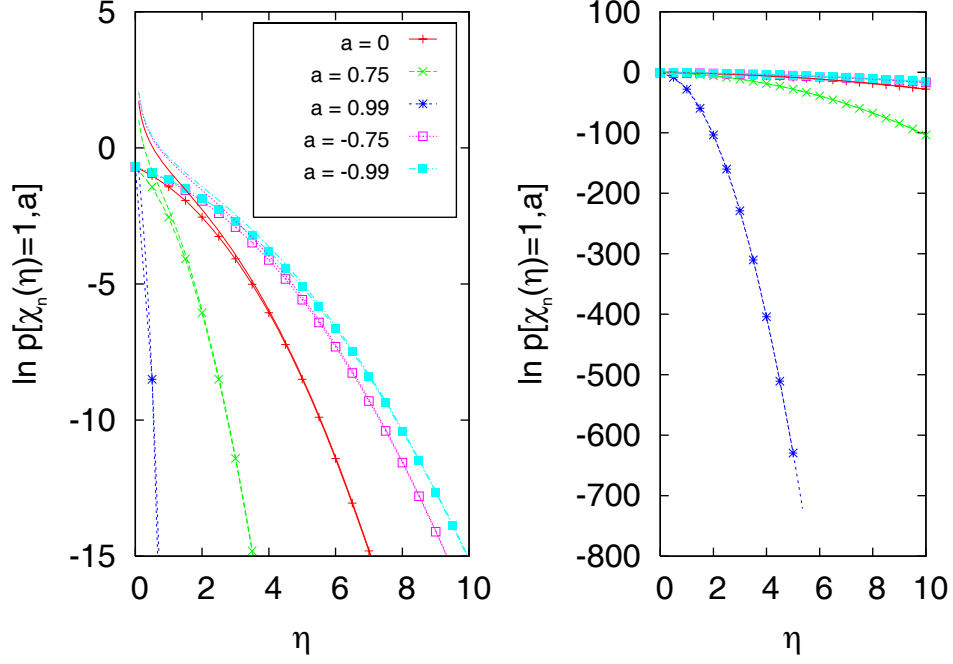


Figure 3.3: The lines with symbols represent the values of $p[\chi_n(\eta) = 1, a]$, which are obtained by numerically integrating the posterior PDF given by Eq. (3.5). The lines without symbols represent the expression for $p[\chi_n(\eta) = 1, a]$ according to Eq. (3.14).

mean only in the special case, that the distribution is symmetric and unimodal. Although our posterior PDFs given by Eq. 3.5 are unimodal, they are not symmetric as Fig. 3.4 illustrates. Thus, replacing the mean by the asymptotic expression for the maximum of the posterior PDF in Eq. (3.20) which will be derived in the following section is nothing else than an educated guess. Hence for large values of η we expect

$$\langle x_n \rangle \simeq u_I \sim -\frac{\eta}{2\sqrt{1-a^2} \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}, \quad (\eta \rightarrow \infty), \quad (3.13)$$

to hold, provided that $\rho[x_n | \chi_n(\eta) = 1, a]$ is not too asymmetric (i.e., a is not too small). Fig. 3.2 compares this asymptotic expression with specific values for the means obtained by numerical integration of Eq. (3.9) and reveals that the approximation in Eq. (3.20) holds well for positive values of a and it is not too bad for negative a .

Inserting the asymptotic expression for the mean into Eq. (3.11) to obtain an approximation for the total probability to observe extreme events

$$p[\chi_n(\eta) = 1, a] \sim \frac{\sqrt{1-a}}{\sqrt{\pi}} \frac{1}{\eta} \exp\left(-\frac{\eta^2}{4(1-a)}\right) \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right), \quad \eta \rightarrow \infty. \quad (3.14)$$

Figure 3.3 shows that for large values of η this expression complies well with the numerical results for $p[\chi_n(\eta) = 1, a]$, which were obtained by numerically integrating the posterior PDF as given by Eq. (3.5).

3.2 Determining the precursor value

Because of the Markov-property of the AR(1) model the precursory variable is reduced to the preceding value x_n and the alarm volume reduces to an alarm interval. Thus, we give an alarm for an extreme event when an observed value x_n is in an interval

$$I = [u - \delta/2, u + \delta/2]; \quad (3.15)$$

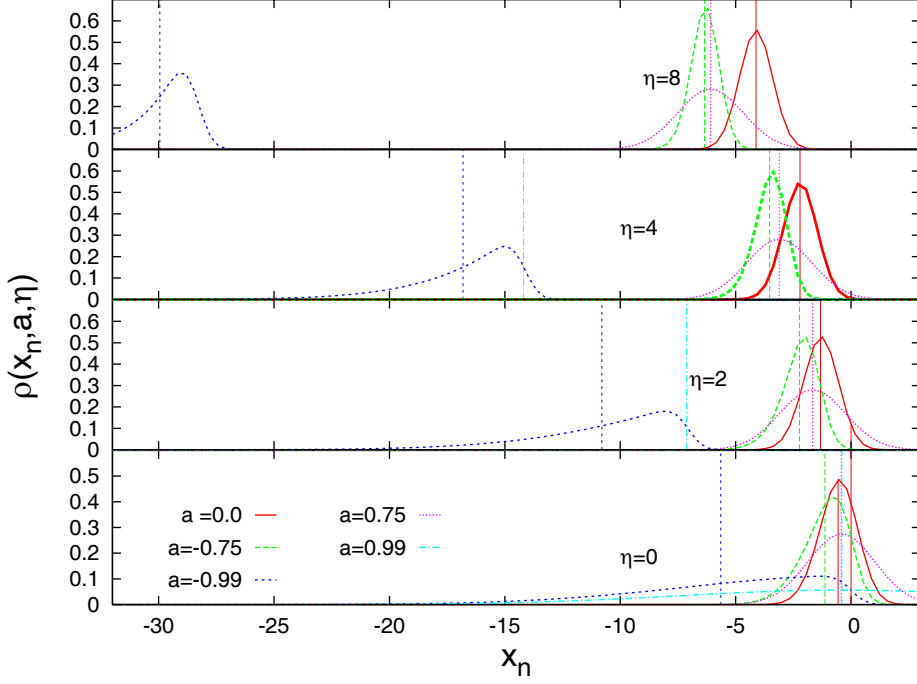


Figure 3.4: The a posteriori PDFs for the AR(1) process are depicted for different values of $a < 0$ and η . The vertical lines represent the means. The PDFs become asymmetric for $a \rightarrow -1$. (For $a = -0.99$ and $\eta \rightarrow \infty$ the marginal PDFs becomes very flat and hence can not be distinguished from the x-axis.)

around the precursor u .

Although we saw in Sec. 2.2.3 that strategy II is superior to strategy I we illustrate this property with the example of the AR(1) process. Therefore we compute the precursor values of both strategies, x_I , and x_{II} , according to Eqs. (2.17) and (2.18), i.e.,

$$u_I := \sup_{x_n} \rho[x_n | \chi_n(\eta) = 1], \quad (3.16)$$

$$u_{II} := \sup_{x_n} p[\chi_n(\eta) = 1 | x_n]. \quad (3.17)$$

3.2.1 The precursor according to strategy I

Differentiating Eq. (3.5) with respect to x_n , the maximum u_I of $p[x_n | \chi_n(\eta) = 1, a]$, is given by the solution of the transcendental equation

$$u_I = \frac{\sqrt{2}}{\sqrt{\pi}(1+a)} \frac{\exp\left(-\frac{1}{2}\left((1-a)x_I + \frac{\eta}{\sqrt{1-a^2}}\right)^2\right)}{\operatorname{erfc}\left(\frac{(1-a)x_I}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right)}. \quad (3.18)$$

Inserting the asymptotic expansion for large arguments of the complementary error function

$$\operatorname{erfc}(z) \sim \frac{\exp(-z^2)}{\sqrt{\pi}z} \left(1 + \sum_{m=1}^{\infty} (-1)^m \frac{1 \cdot 3 \dots (2m-1)}{(2z^2)^m}\right), \quad (z \rightarrow \infty, |\arg z| < \frac{3\pi}{4}), \quad (3.19)$$

which can be found in [89] we obtain:

$$u_I \sim -\frac{\eta}{2\sqrt{1-a^2} \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}, \quad (\eta \rightarrow \infty). \quad (3.20)$$

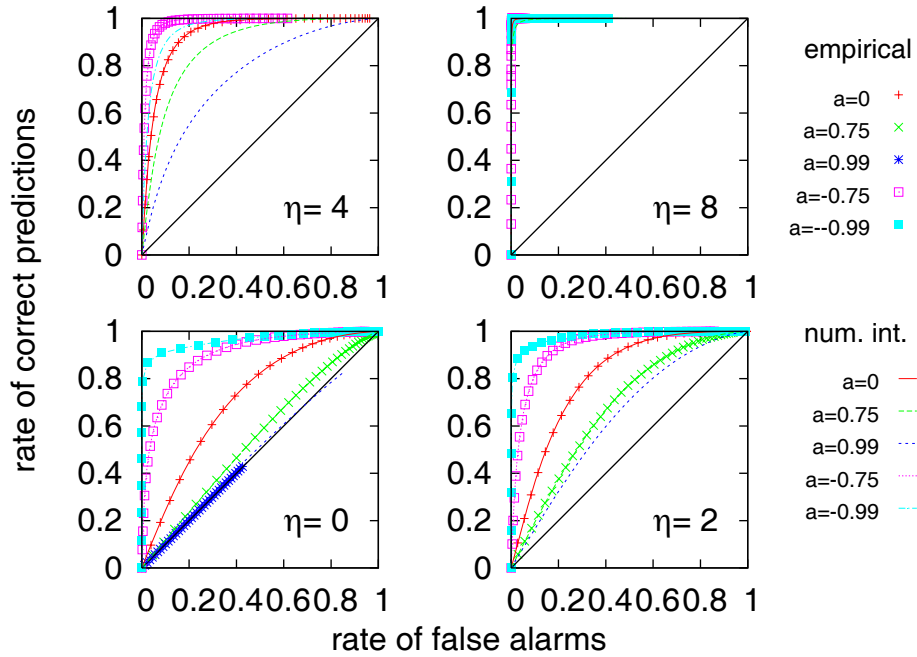


Figure 3.5: The ROC curves made for the precursors of strategy I. The lines represent the results of the evaluation of the integrals in Eqs (2.46) and (2.47), the symbols correspond to the results of the empirical test using the means of $\rho[x_n|\chi_n(\eta) = 1, a]$ as precursors and 10^7 AR(1)-correlated data. In the case $a = 0.99$ the data set contained no extreme events which were significantly larger than $\eta = 0$. Very large events ($\eta = 8$) were only observed for $a = -0.75$ and $a = -0.99$.

Fig. 3.4 shows the posterior PDFs according to Eq. (3.5) for different values of a and η . One can see that the maximum of $\rho[x_n|\chi_n(\eta) = 1, a]$ moves towards $-\infty$ with increasing size of η and decreasing a . Note that $\rho[x_n|\chi_n(\eta) = 1, a]$ becomes asymmetric if $a \rightarrow -1$ and its variance increases immensely if $a \rightarrow 1$. Although we can always formally define the maximum u_I one can argue that the maximum of the distribution has no predictive power if $a \rightarrow 1$. Since $\rho[x_n|\chi_n(\eta) = 1, a]$ becomes very flat in this limit, the value of $\rho[x_n|\chi_n(\eta) = 1, a]$ in its maximum does not considerably differ from the values in any other point.

3.2.2 The precursor according to strategy II

In order to determine u_{II} , the precursor for strategy II, we have to find the maximum of the likelihood in x_n as given in Eq. (3.7). Since the complementary error function is a monotonously decreasing function of x_n we see that we do not have a well defined maximum u_{II} and that the interval $I_- = [-\infty, \delta]$ with the upper limit δ represents the optimal strategy to raise alarms according to strategy II. In other words, the optimal precursor for strategy II is

$$u_{II} = -\infty. \quad (3.21)$$

Since we will never observe this optimal precursor value in a given data set, we consider the smallest observed value in the data set as an approximation for u_{II} , i.e., the “practically useable” value u_{II}^p is given by

$$u_{II}^p := \min\{x_n\}. \quad (3.22)$$

3.3 Testing the Performance of the Precursor

In order to test for the predictive power of u_I and u_{II} we create ROC curves, as discussed in chapter 2. We use two different ways of determining the rates of correct prediction and

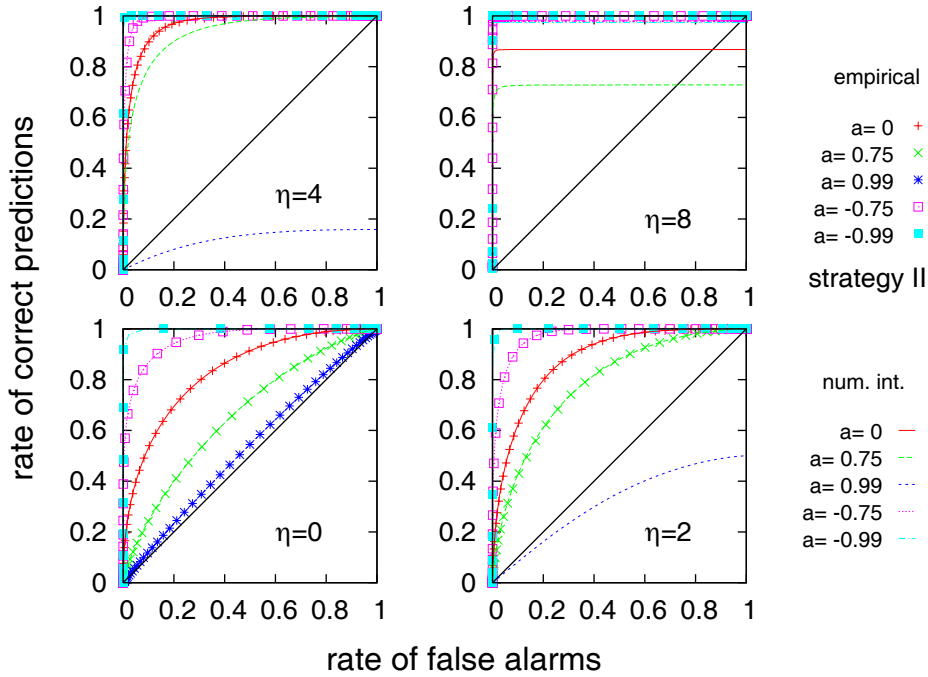


Figure 3.6: The ROC curves made for the precursors of strategy II. The lines represent the results of the evaluation of the integrals in Eqs (2.46) and (2.47), the symbols correspond to the results of the empirical test using the means of $\rho[x_n|\chi_n(\eta) = 1, a]$ as precursors and 10^7 AR(1)-correlated data. In the case $a = 0.99$ the data set contained no extreme events which were significantly larger than $\eta = 0$. Very large events ($\eta = 8$) were only observed for $a = -0.75$ and $a = -0.99$.

false alarms. In order to determine the rates from the analytically calculated posterior PDFs we evaluate the integrals in Eqs. (2.46) and (2.47) by using *Mathematica* [90]. As precursor values we inserted the means of the posterior PDFs as an approximation of the maxima for the calculations according to strategy I. In order to calculate the rates of false alarms and correct predictions according to strategy II we used the smallest values of the AR(1) correlated data sets as precursors. Note that this evaluation of strategy II, was made in order to compare the analytic results with the ROC curves obtained from the AR(1) correlated data sets. Thus the smallest value of the data sets was inserted for the numerical evaluation of the alarm rates, instead of the theoretical precursor value $u_{II} = -\infty$.

We also generated ROC-plots empirically by predicting within 10^7 AR(1)-correlated data. The second method consists in simply performing predictions on a time series of 10^7 AR(1) data, and counting the number of extreme increments which could be predicted by using the precursors specified above. For different values of the correlation coefficients the data sets contained the following numbers of extreme increments:

a	number of increments of size			
	$\eta \geq 0$	$\eta \geq 2$	$\eta \geq 4$	$\eta \geq 8$
-0.99	5000059	1579103	222858	310
-0.75	5000563	1425146	162405	107
0	5000417	786355	23370	0
0.75	5000818	23377	0	0
0.99	5001081	0	0	0

For both methods, the size of the precursory volume ranged from 10^{-6} to 4, measured in size of the standard deviation of the marginal PDF of the AR(1) process $\sigma(a) = 1/\sqrt{1-a^2}$.

The figures 3.5 and 3.6 reveal that the empirically determined rates comply well with the rates obtained via the evaluation of Eqs. (2.46) and (2.47). Hence, we tried to create the ROC curves

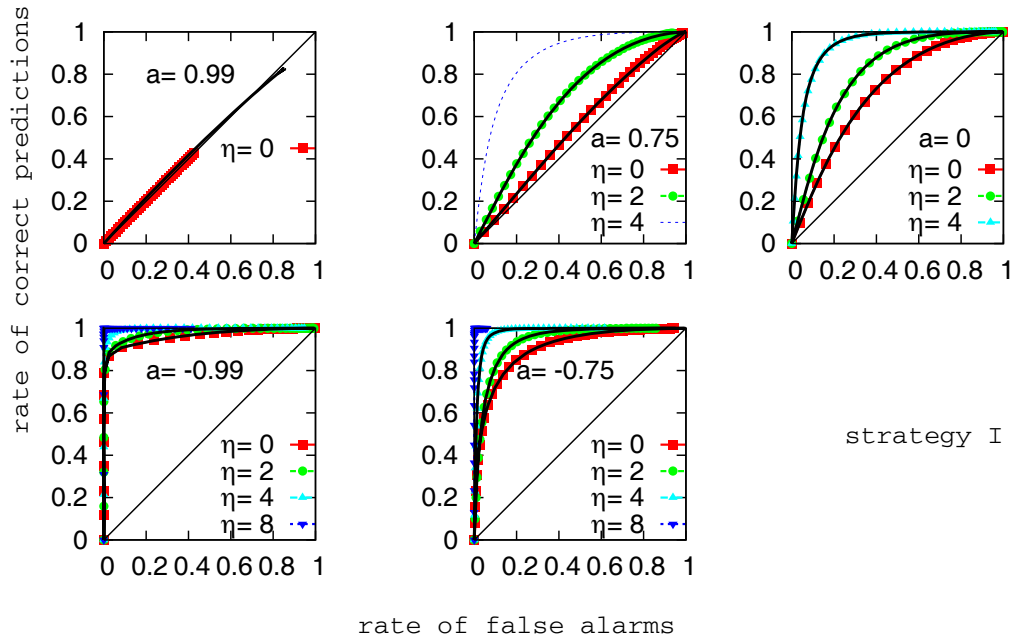


Figure 3.7: The influence of the correlation strength a and the event magnitude η on the ROC-Plots made for the precursors of strategy I.

for the values of a and η , which were not accessible for the empirical test, via evaluating the integrals in Eqs. (2.46) and (2.47) with *Mathematica*. However for very large values of η , and $a \rightarrow 1$ also the evaluation of the integrals was not successful, see e.g, the curves for $a = 0.99$ in Fig. 3.6. The reason for this, is that in this case, the probability to find large increments in the data set become smaller than 10^{-7} and thus the standard algorithms for the numerical integration fail to archive reasonable accuracy.

The resulting ROC curves are shown in Figs. 3.7- 3.9. Whereas Figs. 3.5 and 3.6 focus on comparing the ROC curves obtained via integrating the posterior PDFs and the empirically generated ROC curves, Figs. 3.7 and 3.8 reveal the dependence on the event magnitude for different values of a and Fig.3.9 focus on the comparison of strategy I and strategy II.

In total the resulting ROC curves display the following properties:

A1, superiority of strategy II:

Fig. 3.9 illustrate that the predictions according to strategy II are better than the predictions according to strategy I for all values of a and η . A detailed discussion of this phenomenon will be provided in Sec. 3.4.

A2, influence of the event magnitude:

The ROC curves display an increase of the quality of our prediction with increasing size of the events η . Explanations for this effect will be provided by the asymptotic expression for the likelihood ratio in Sec. 3.4.

A3, influence of the correlation:

The prediction within the strongly correlated random numbers with positive correlation strength ($a=0.99$) is *not* (significantly) better than any random prediction which corresponds to the diagonal. Hence our precursor does not have any predictive power in this case. This complies with the fact that $\rho[x_n|\chi_n(\eta) = 1, a]$ is very flat for this value of a , as discussed in the previous section. Surprisingly we observe a better quality of prediction for $a = 0$, although in this case the predictions were made within completely uncorrelated

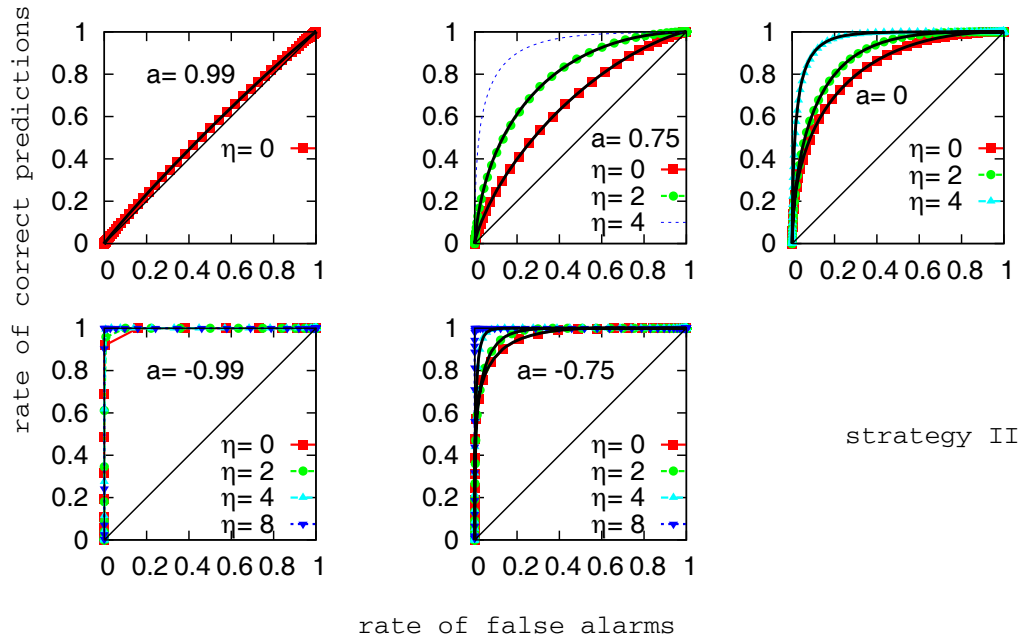


Figure 3.8: The influence of the correlation strength a and the event magnitude η on the ROC-Plots made for the precursors of strategy II.

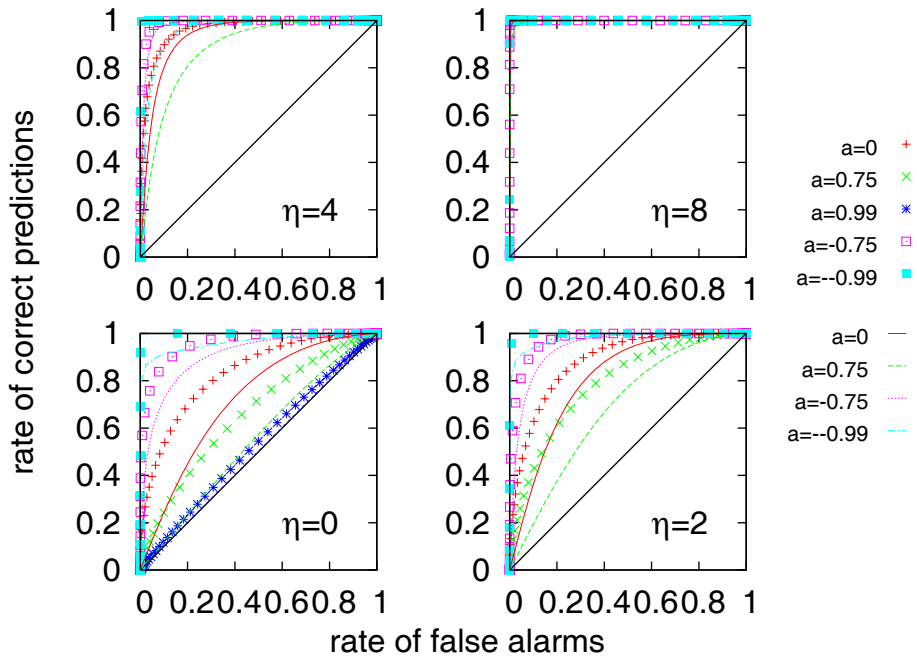


Figure 3.9: The ROC-Plots made for the precursors of strategy I and II. The lines represent the results of strategy I, the symbols correspond to predictions made according to strategy II. In both cases the predictions were made within 10^7 AR(1)- correlated data.

random numbers. This is contrary to the intuitive expectation that the quality of the prediction should increase with increasing correlation strength. In fact, we observe the opposite: the predictability increases with decreasing correlation strength a . This observation is in agreement with results which were reported by Sornette et al. in [91] for the

prediction of signs of increments in uncorrelated random numbers. The prediction of the sign of an increment within uncorrelated random numbers corresponds to the special case $\eta = 0$ and $a = 0$ of the AR(1) process we discuss here. We can understand this counter intuitive effect by considering that increments are not independent from the last observation. More precisely $x_{n+1} - x_n = (a - 1)x_n + \xi_n$, so that the absolute value of the known part of the increment $(a - 1)x_n$ is the larger, the smaller a . A formal explanation of the a -dependence is also given by an asymptotic expression for the slope $m(u, a, \eta)$, which we derive in Sec. 3.4.

3.4 Analytical investigation of the Precursor Performance

In this section we try to understand the results A1-A3 obtained by the ROCs by investigating the asymptotic behavior of the slope $m(u, \eta, a)$ of the ROC curve in the vicinity of the origin. As we saw in chapter 2 the slope of the ROC curves $m(u, \eta, a)$ is identical to the likelihood ratio $\Lambda(u, \eta, a)$. For the AR(1) processes under study the slopes in the vicinity of the origin are given by

$$\Lambda(a, \eta, u) = \frac{(1 - p[\chi_n(\eta) = 1, a])}{p[\chi_n(\eta) = 1, a]} B(u, \eta, a), \quad (3.23)$$

$$\text{with } B(u, \eta, a) = \frac{\operatorname{erfc}\left(\frac{(1-a)u}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right)}{1 + \operatorname{erf}\left(\frac{(1-a)u}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right)}, \quad (3.24)$$

where $B(u, \eta, a)$ denotes the ratio of odds, as introduced in the previous chapter (see Eq. (2.33)). Inserting the asymptotic expression for the total probability in Eq. (3.14) we obtain

$$\Lambda(a, \eta, u) \sim \left(\frac{\sqrt{\pi}}{\sqrt{1-a}} \frac{\eta \exp\left(\frac{\eta^2}{4(1-a)}\right)}{1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)} - 1 \right) B(u, \eta, a), \quad (\eta \rightarrow \infty). \quad (3.25)$$

According to the choice of the precursor u , we have to consider two different behaviors of the ratio of odds $B(u, \eta, a)$

$$\lim_{\eta \rightarrow \infty} B(u, \eta, a) = \begin{cases} 1, & \text{if } u = -\infty; \\ 0, & \text{if } u > -\infty. \end{cases} \quad (3.26)$$

3.4.1 The theoretical precursor given by strategy II

The first case in Eq. (3.26) corresponds to the theoretical precursor value for strategy II. The slope of the ROC curves obtained with strategy II can be computed by simply evaluating Eq. (3.23) at the value $u_{II} = -\infty$

$$\Lambda(a, \eta, u_{II}) = \infty \quad \text{for } x_{II} = -\infty, \quad \eta \neq 0, \quad (3.27)$$

$$\text{since } \operatorname{erf}(\infty) = -1. \quad (3.28)$$

Thus, if we make predictions according to strategy II, with a size independent precursor $u_{II} = -\infty$, we should in theory expect to obtain ROC curves which coincide with the vertical axis of the plot and hence represent an ideal predictability for all size of events and all possible correlation strengths. However, since in any finite data set, the precursor is not $-\infty$, but the smallest accessible data point, we find also for strategy II ROC curves which depend on a and η .

3.4.2 Finite (realistic) precursors

For any finite precursor value, i.e., here u_I and any finite u_{II}^p , it is necessary to ascertain, if the contribution of the ratio $B(u, \eta, a)$ or the exponential function in Eq. (3.25) prevails. Inserting the asymptotic expansion of the complementary error function in Eq. (3.19), the ratio $B(u, \eta, a)$ reads

$$\begin{aligned} B(z) &= \frac{\operatorname{erfc}(z)}{1 + \operatorname{erf}(z)} = \frac{\operatorname{erfc}(z)}{2 - \operatorname{erfc}(z)} \\ &\propto \frac{\exp(-z^2) \left(1 + \mathcal{O}\left(\frac{1}{z^2}\right)\right)}{2\sqrt{\pi}z - \exp(-z^2) \left(1 + \mathcal{O}\left(\frac{1}{z^2}\right)\right)}, \quad (z \rightarrow \infty) \\ \text{with } z &= z(\eta, a, u) = \frac{(1-a)}{\sqrt{2}}u + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}} \end{aligned} \quad (3.29)$$

Note that the limit $z \rightarrow \infty$ corresponds to the limit $\eta \rightarrow \infty$ for fixed values of a and u , but it can also be interpreted as the limit $a \rightarrow \pm 1$ for a fixed (sufficiently large) value of η . The later case will be discussed in more detail in Sec. 3.4.3, whereas we focus in this section on the dependence of the coupling strength. We therefore demand in the following that the selected precursor value u is fixed and finite such that $\eta \rightarrow \infty$ implies $z \rightarrow \infty$. Inserting the consideration made for the ratio of odds, we obtain the following asymptotic expression for the slope

$$\Lambda(a, \eta, u) \sim \frac{\sqrt{\pi}}{\sqrt{1-a}} \left(\frac{\eta \exp\left(\frac{\eta^2}{4(1-a)}\right)}{\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)} - 1 \right) B(u, \eta) \quad (3.30)$$

$$\sim \frac{\sqrt{\pi}}{\sqrt{1-a}} \left(\frac{\eta \exp\left(\frac{\eta^2}{4(1-a)}\right)}{\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)} \frac{\exp(-z^2) \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}{2\sqrt{\pi}z - \exp(-z^2) - \mathcal{O}\left(\frac{\exp(-z^2)}{\eta^2}\right)} - B(u, \eta) \right) \quad (3.31)$$

$$\sim \frac{\eta \exp\left(\frac{\eta^2}{4(1-a)} - z^2\right)}{2\sqrt{1-a}z - \frac{\sqrt{1-a}}{\sqrt{\pi}} \exp(-z^2) - \mathcal{O}\left(\frac{\exp(-z^2)}{\eta^2}\right)} - \frac{\sqrt{\pi}B(u, \eta)}{\sqrt{1-a}} \quad (3.32)$$

$$\sim \frac{\eta \exp(f(u, \eta, a))}{2\sqrt{1-a}z - \frac{\sqrt{1-a}}{\sqrt{\pi}} \exp(-z^2) - \mathcal{O}\left(\frac{\exp(-z^2)}{\eta^2}\right)} - \mathcal{O}\left(\frac{\exp(-z^2)}{z^2}\right),$$

$$\text{with } f(u, \eta, a) = \frac{\eta^2}{4(1-a)} - \left(\frac{(1-a)u}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}} \right)^2,$$

$$\text{and } \mathcal{O}(B(u, \eta)) = \mathcal{O}\left(\frac{\exp(-z^2)}{z^2}\right). \quad (3.33)$$

The denominator of Eq. (3.33) tends to $2\sqrt{1-a}z$ in the limit $z \rightarrow \infty$, hence the asymptotic behavior of the whole expression is mainly influenced by the argument $f(u, \eta, a)$ of the exponential function in Eq. (3.33). If $f(u, \eta, a)$ is positive, we can expect $\Lambda(a, \eta, u)$ to increase exponentially with increasing event magnitude η . Whereas we can expect $\Lambda(a, \eta, u)$ to approach a constant in the limes $\eta \rightarrow \infty$ if $f(u, \eta, a)$ is negative. Hence we can use the zeros of this parabolae $f(u, \eta, a)$, namely

$$(u^0(a, \eta))_{+,-} = \frac{\eta}{1-a} \left(-\frac{1}{\sqrt{1-a^2}} \pm \frac{1}{\sqrt{2(1-a)}} \right) \quad (3.34)$$

to discuss the regimes in which a chosen precursor value leads to an increase of $\Lambda(u, a, \eta)$ with increasing η . Figure 3.10 illustrates the following discussion. For a fixed value of η , we can clearly distinguish specific intervals $[(u^0)_-, (u^0)_+]$, for which $f(u, \eta, a)$ is nonnegative. However for a fixed value of η it is also pointless to discuss the limit $\eta \rightarrow \infty$. As we see in Eq. (3.34)

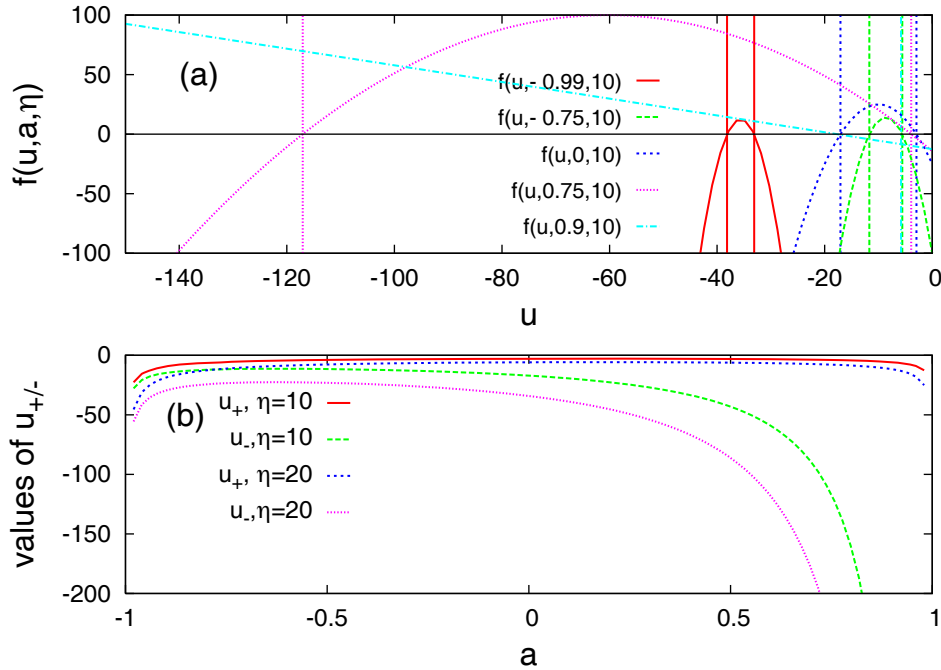


Figure 3.10: (a) The parabolae $f(a, u, \eta)$ for $\eta = 10$ and different values of a . (b) The zeros of the parabolae given by Eq. (3.34) in dependence on the coupling strength a .

and in Fig. 3.10, the width of the intervals $[(u^0)_{-}, (u^0)_{+}]$ is a function of the coupling strength a and especially for $a \rightarrow -1$, the intervals become quite small. Thus for a fixed range of $\eta \in [\eta_{-}, \eta_{+}]$, and a precursor value $u \notin [(u^0(a, \eta))_{-}, (u^0(a, \eta))_{+}]$ the slope of the ROC-plot could also decrease instead of increase with increasing η , which is in contrast to the observations made in the previous section. However, we should not forget, that the borders of the intervals $[(u^0)_{-}, (u^0)_{+}]$ are functions of η themselves, i.e., they decrease linearly with increasing η , as the explicit expression for the zeros of in Eq. (3.34) reveals and as it is illustrated in Fig. 3.11. Furthermore, the width $w(a, \eta)$ of the intervals in which $f(u, a, \eta)$ is positive also increases linearly with increasing η

$$w(a, \eta) = (u^0(a, \eta))_{+} - (u^0(a, \eta))_{-} \quad (3.35)$$

$$= \sqrt{\frac{2}{(1-a)(1+a)}} \eta, \quad (3.36)$$

which is also displayed in Fig. 3.11 (b). Hence, in the limit $\eta \rightarrow \infty$, $\eta \neq \infty$ also the width of the interval for which $f(u, a, \eta)$ is positive increases, i.e., it becomes more likely that we find a specific precursor value u within this interval. Only if $\eta = \infty$ the interval size reduces to zero and we arrive at the single point $u = -\infty$, which is the optimal (theoretical) precursor according to strategy II. Thus discussing the limit $\eta \rightarrow \infty$ implies that $w(a, \eta) \rightarrow \infty$ and $u \in [(u^0(a, \eta))_{+}, (u^0(a, \eta))_{-}]$. The consequence of this is

$$\Lambda(a, \eta, u) \rightarrow \infty, \quad (u \in [(u^0(a, \eta))_{+}, (u^0(a, \eta))_{-}], \text{ and } \eta \rightarrow \infty). \quad (3.37)$$

The special case of strategy I

There are (at least) three ways to show, that for strategy I larger events are the better predictable, the larger they are, i.e., the slope m of the ROC curve increases, with increasing η . The first approach consist in simply showing, that the asymptotic expression for u_I according

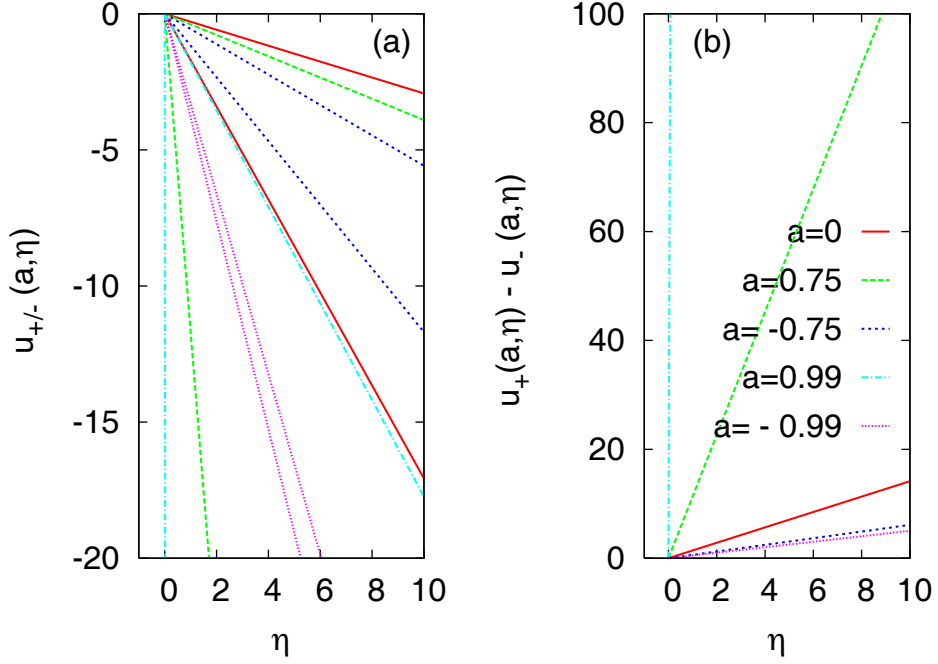


Figure 3.11: The zeros of the parabolae given by Eq. (3.34) in dependence on the event magnitude η (a) and the width $w(a, \eta) = (u^0(a, \eta))_+ - (u^0(a, \eta))_-$ of the intervals in which the expression in Eq. (3.34) is positive as functions of η .

to Eq. (3.13) lies within the interval $[(u^0(a, \eta))_-, (u^0(a, \eta))_+]$ for all value of a and η . As the discussion in the previous section suggests, in this case $m \rightarrow \infty$, as $\eta \rightarrow \infty$. This can be also shown by inserting the asymptotic expression, given by Eq. (3.13) in Eq. (3.33). The third approach starts even before Eq. (3.33) and focuses on separating the influences of the posterior PDF to find events, after the precursor was observed $\rho[x_n | \chi_n(\eta) = 1, a]$, and the corresponding PDF for non-events $\rho[x_n | \chi_n(\eta) = 0, a]$. In order to illustrate the role of each posterior PDF we will choose a third approach. Therefore we study first the asymptotic behavior of the posterior PDFs $\rho[u_I | \chi_n(\eta) = 1, a]$ and $\rho[u_I | \chi_n(\eta) = 0, a]$ and then the behaviour of the complete likelihood ratio. Inserting the asymptotic expression for the total probability to find events according to Eq. (3.14), the approximation of u_I in Eq. (3.13) and the asymptotic expansion of the complementary error function Eq. (3.19) we find the following asymptotic expression

$$\begin{aligned}
\rho[u_I | \chi_n(\eta) = 1, a] &= \frac{\sqrt{1-a^2}}{2\sqrt{2\pi} p[\chi_n(\eta) = 1, a]} \exp\left(-\frac{1-a^2}{2} u_I^2\right) \\
&\quad \operatorname{erfc}\left(\frac{(1-a)u_I}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right) \\
&\sim \frac{\sqrt{1-a^2}}{2\sqrt{2\pi} p[\chi_n(\eta) = 1, a]} \exp\left(-\frac{\eta^2}{8} \frac{1}{\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}\right) \\
&\quad \operatorname{erfc}\left(\left(\frac{1+a + \mathcal{O}\left(\frac{1}{\eta^2}\right)}{2\sqrt{2}\sqrt{1-a^2} \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}\right) \eta\right), \quad (\eta \rightarrow \infty)
\end{aligned} \tag{3.38}$$

In the following we will denote the asymptotic expression for the argument of the complementary error function by

$$g(a, \eta) = \left(\frac{1+a + \mathcal{O}\left(\frac{1}{\eta^2}\right)}{2\sqrt{2}\sqrt{1-a^2} \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}\right). \tag{3.39}$$

By using the asymptotic expression of the complementary error function in Eq. (3.19), we arrive at

$$\rho[u_I|\chi_n(\eta) = 1, a] \sim \frac{\sqrt{1-a^2}}{2\sqrt{2\pi} p[\chi_n(\eta) = 1, a]} \frac{\exp\left(-\frac{\eta^2}{8} \frac{1}{\left(1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right)} - (\eta g(a, \eta))^2\right)}{\sqrt{\pi} (\eta g(a, \eta))} \left(1 + \sum_{m=1}^{\infty} (-1)^m \frac{1 \cdot 3 \dots (2m-1)}{2 (\eta g(a, \eta))^{2m}}\right), \quad (\eta \rightarrow \infty). \quad (3.40)$$

Simplifying the argument of the exponential function

$$\begin{aligned} & \left(-\frac{\eta^2}{8} \frac{1}{\left(1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right)} - \left(\left(\frac{1+a+\mathcal{O}\left(\frac{1}{\eta^2}\right)}{2\sqrt{2}\sqrt{1-a^2}\left(1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}\right)\eta\right)^2\right) \\ &= -\frac{\eta^2}{8\left(1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right)} \left(1 + \frac{1+2a+a^2+2(1+a)\mathcal{O}\left(\frac{1}{\eta^2}\right)+\mathcal{O}\left(\frac{1}{\eta^4}\right)}{1-a^2}\right) \\ &= -\frac{\eta^2}{8\left(1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right)} \left(1 + \frac{1+a+\mathcal{O}\left(\frac{1}{\eta^2}\right)}{1-a}\right) \\ &= -\frac{\eta^2}{8(1-a)\left(1+\mathcal{O}\left(\frac{1}{\eta^2}\right)\right)} \left(2 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right) = -\frac{\eta^2}{4(1-a)}. \end{aligned}$$

and truncating the expansion yields

$$\rho[u_I|\chi_n(\eta) = 1, a] \sim \frac{\sqrt{1-a^2}}{2\sqrt{2\pi} p[\chi_n(\eta) = 1, a]} \frac{\exp\left(-\frac{\eta^2}{4(1-a)}\right)}{\eta g(a, \eta)} \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right), \quad (\eta \rightarrow \infty). \quad (3.41)$$

If we now replace the total probability to find events by its asymptotic expression from Eq. (3.14), i.e.,

$$p[\chi_n(\eta) = 1, a] \sim \frac{\sqrt{1-a}}{\sqrt{\pi}} \frac{1}{\eta} \exp\left(-\frac{\eta^2}{4(1-a)}\right) \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right), \quad (\eta \rightarrow \infty),$$

we arrive at

$$\rho[u_I|\chi_n(\eta) = 1, a] \sim \frac{\sqrt{1+a} \eta}{2\sqrt{2\pi} \eta} \frac{\exp\left(+\frac{\eta^2}{4(1-a)} - \frac{\eta^2}{4(1-a)}\right)}{g(a, \eta)} \frac{\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}{\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}, \quad (\eta \rightarrow \infty) \quad (3.42)$$

$$\sim \frac{\sqrt{1+a}}{2\sqrt{2\pi}} \frac{1}{g(a, \eta)} \quad (3.43)$$

$$\sim \frac{\sqrt{1-a^2}\sqrt{1+a}}{\sqrt{\pi}} \frac{\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}{\left(1 + a + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}, \quad (\eta \rightarrow \infty). \quad (3.44)$$

Hence the value of the PDF at the precursor approaches a constant if terms of the order $1/\eta^2$ can be neglected. The asymptotic expression for $\rho[u_I|\chi_n(\eta) = 0, a]$ can be evaluated analogously

$$\rho[u_I|\chi_n(\eta) = 0, a] = \frac{\sqrt{1-a^2}}{2\sqrt{2\pi}(1-p[\chi_n(\eta) = 1, a])} \exp\left(-\frac{1-a^2}{2}u_I^2\right) \left(1 + \operatorname{erf}\left(\frac{(1-a)u_I}{\sqrt{2}} + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}\right)\right) \quad (3.45)$$

$$\sim \frac{\sqrt{1-a^2}}{\sqrt{2\pi}(1-p[\chi_n(\eta) = 1, a])} \exp\left(-\frac{\eta^2}{8} \frac{1}{\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}\right) \quad (3.46)$$

$$\left(1 - \frac{1}{2}\operatorname{erfc}(\eta g(a, \eta))\right), \quad (\eta \rightarrow \infty) \quad (3.47)$$

$$\sim \frac{\sqrt{1-a^2} \exp\left(-\frac{\eta^2}{8} \frac{1}{\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}\right)}{\sqrt{2\pi} \left(1 - \frac{\sqrt{1-a}}{\sqrt{\pi}} \frac{1}{\eta} \exp\left(-\frac{\eta^2}{4(1-a)}\right) \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)\right)} \left(1 - \frac{\exp\left(-(\eta g(a, \eta))^2\right)}{2\sqrt{\pi}(\eta g(a, \eta))} \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)\right), \quad (\eta \rightarrow \infty) \quad (3.48)$$

$$\sim \frac{\sqrt{1-a^2} \exp\left(-\frac{\eta^2}{8} \frac{1}{\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}\right)}{\sqrt{2\pi} \left(1 - \mathcal{O}\left(\frac{\exp(-\eta^2)}{\eta}\right)\right)} \left(1 - \mathcal{O}\left(\frac{\exp(-\eta^2)}{\eta}\right)\right), \quad (\eta \rightarrow \infty)$$

$$\sim \frac{\sqrt{1-a^2}}{\sqrt{2\pi}} \exp\left(-\frac{\eta^2}{8} \frac{1}{\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}\right), \quad (\eta \rightarrow \infty). \quad (3.49)$$

Thus, we can conclude, that $\rho[u_I|\chi_n(\eta) = 1, a]$ decreases in the leading order as $\exp(-\eta^2)$ with increasing η . The numerically evaluated PDFs $\rho[x_n|\chi_n(\eta) = 1, a]$ and $\rho[x_n|\chi_n(\eta) = 0, a]$ at $x_n = u_I$ and their approximations given by Eqs. (3.44) and (3.49) are compared in Fig. 3.12. The approximations become worse for $a \rightarrow -1$, since in this case $\rho[x_n|\chi_n(\eta) = 1, a]$ becomes asymmetric and hence the assumption in Eq. (3.13) does not hold. Note that the value of the failure PDF at the precursor value decreases as a squared exponential with increasing η .

Inserting the asymptotic expressions for the posterior PDFs in Eqs. (3.44) and (3.49) yields

$$\Lambda(x_I, a, \eta) \sim \sqrt{\frac{2}{1+a}} \exp\left(\frac{\eta^2}{8 \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}\right) \frac{\left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}{\left(1 + a + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)} \quad \eta \rightarrow \infty. \quad (3.50)$$

The numerical evaluation of this expression leads to the straight lines in Fig. 3.13 which indicate the slopes of the ROC-plots. Since the posterior PDF of the events approaches a constant, whereas the posterior PDF for the non-events decreases, with increasing η one can argue, that strategy I's main contribution to the slope of the ROC curves is the minimization of the rate of false alarms. Or more precisely, at least, the η -dependence of the slope can be explained as a decrease of the failure PDF (see Eq. (3.44) and Fig. 3.12). This effect is illustrated in Fig. 3.14 which compares for families of posterior PDF $\rho[x_n|\chi_n(\eta, a) = 1, a]$ $\rho[x_n|\chi_n(\eta, a) = 0, a]$ parametrized by the event magnitude η for $a = -0.75$. Since the maximum of the posterior PDF $\rho[x_n|\chi_n(\eta, a) = 0, a]$ remains at the origin, but the precursor of strategy I, i.e., the maximum of $\rho[x_n|\chi_n(\eta, a) = 0, a]$, tends to infinity, the values of $\rho[x_n|\chi_n(\eta, a) = 0, a]$ at the precursor value x_I decrease according to the decrease of $\rho[x_n|\chi_n(\eta, a) = 0, a]$ as $x_n \rightarrow -\infty$.

Since the precursor value obtained by strategy II is even further away, from x_I in which the marginal failure PDF $\rho[x_n|\chi_n(\eta) = 0, a]$ has its maximum, strategy II focus on the minimization

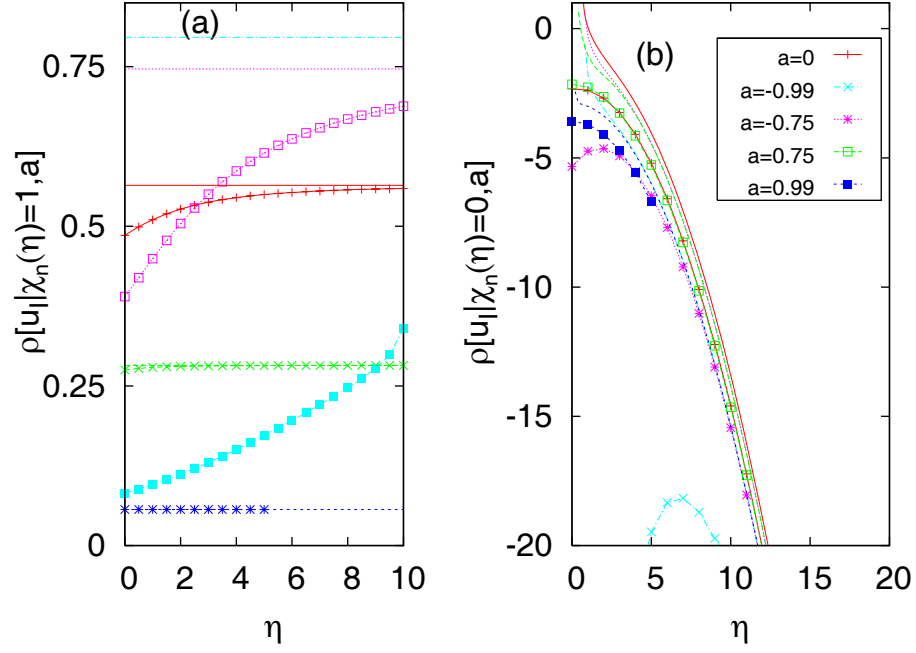


Figure 3.12: The posterior PDFs to observe or not to observe events, evaluated for the precursor of strategy I. The lines without symbols represent the approximations of PDFs $\rho[u_I|\chi_n(\eta) = 1, a]$ in Fig. (a) and $\rho[u_I|\chi_n(\eta) = 0, a]$, Fig. (b) for different values of $0 \leq a < 1$. The symbols are the results of the numerical evaluation of Eqs. (3.38) and (3.46). The constant lines represent the limit for large η according to Eqs. (3.44) and (3.49).

of the failure rate, rather than on the maximization of the rate of correct predictions. The fact, that in this point the corresponding value of the marginal PDF $\rho[x_n|\chi_n(\eta) = 1, a]$ is also far away from the maximum of $\rho[x_n|\chi_n(\eta) = 1, a]$ does apparently not significantly influence the quality of the prediction. Since strategy II leads to better predictions than strategy I, the minimization of the rate of false alarm seems to be more relevant for the quality of the predictions, than maximizing the rate of correct predictions.

3.4.3 The dependence on the coupling strength

As we saw before for $u_{II} = -\infty$, the theoretical precursor of strategy II, the asymptotic behavior of $\Lambda(a, \eta, u_{II}) \rightarrow \infty$. Hence the slope would be independent of the value of the coupling strength if the “theoretical” precursor of strategy II could be used.

For any finite precursor value, the asymptotic expression for the slope of the ROC-plots in Eq. (3.33) can also serve to investigate the dependence on the variable a , as we already mentioned in Sec. 3.4.2. Starting again from Eq. (3.33), namely

$$\Lambda(u, a, \eta) \sim \frac{\eta \exp(f(u, \eta, a))}{\sqrt{1-a} \left(2z - \frac{1}{\sqrt{\pi}} \exp(-z^2)\right) - \mathcal{O}\left(\frac{\exp(-z^2)}{\eta^2}\right)} - \mathcal{O}\left(\frac{\exp(-z^2)}{z^2}\right), \quad (z \rightarrow \infty)$$

$$\text{with } f(u, \eta, a) = -\frac{\eta^2}{4(1+a)} - \frac{\sqrt{1-a}}{\sqrt{1+a}} u \eta - \frac{(1-a)^2}{2} u^2,$$

$$\text{and } z = z(u, a, \eta) = \frac{(1-a)}{\sqrt{2}} u + \frac{\eta}{\sqrt{2}\sqrt{1-a^2}}. \quad (3.51)$$

we now interpret the limit $z \rightarrow \infty$ in the following way. Let η have an arbitrary fixed value, large enough to ensure that Eq. (3.14) is a useful approximation for the total probability and thus the limit $z \rightarrow \infty$ corresponds to $a \rightarrow \pm 1$. Hence, for any finite precursor the asymptotic

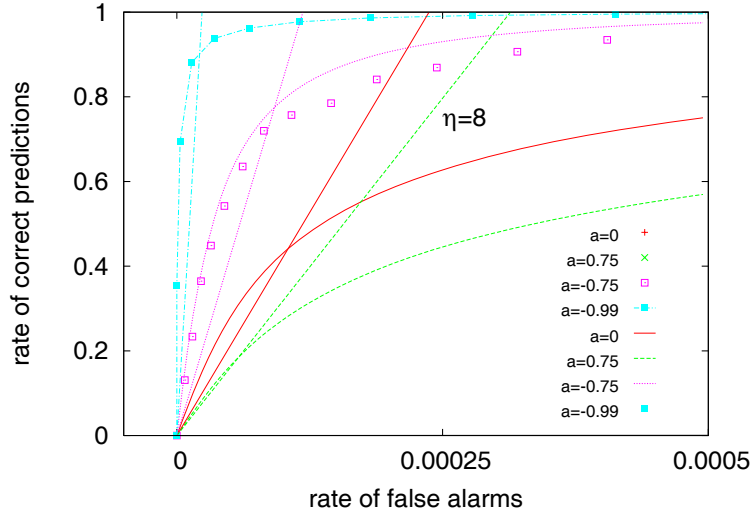


Figure 3.13: The ROC-plots made according to strategy I for $\eta = 8$ in the vicinity of the origin. The straight lines indicate the slope according to Eq. (3.50).

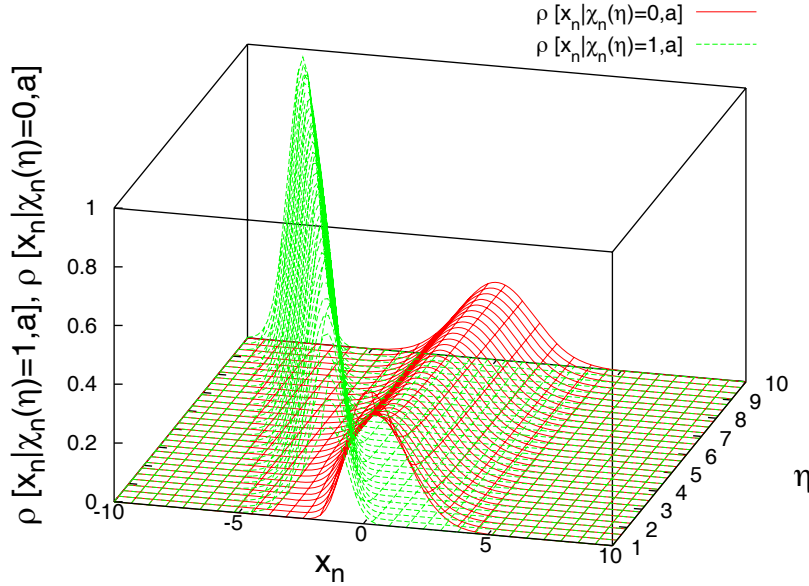


Figure 3.14: $\rho[x_n|\chi_n(\eta) = 1, a]$ and $\rho[x_n|\chi_n(\eta) = 0, a]$ for $a = -0.75$. The maximum of the marginal PDF to observe extreme events $\rho[x_n|\chi_n(\eta) = 1, a]$, which is used as precursor, moves towards $-\infty$ with increasing η since $x_I \sim -\eta/(2\sqrt{1-a^2})$. Because the maximum of the marginal failure PDF $\rho[x_n|\chi_n(\eta) = 0, a]$ remains at the origin, the values of $\rho[x_n|\chi_n(\eta) = 0, a]$ which are observed at the precursor value x_I decrease according to the decrease of $\rho[x_n|\chi_n(\eta) = 0, a]$ as $x_n \rightarrow -\infty$.

expressions of $f(u, a, \eta)$ and $z^2(u, a, \eta)$ read as follows

$$f(u, a, \eta) \sim -\frac{\eta^2}{4(1+a)} + \mathcal{O}(\sqrt{1-a}), \quad (a \rightarrow 1), \quad (3.52)$$

$$z^2(u, a, \eta) \sim \frac{\eta^2}{2(1-a^2)} + \mathcal{O}(\sqrt{1-a}), \quad (a \rightarrow 1), \quad (3.53)$$

and we would thus expect $f(u, a, \eta) \rightarrow \text{const.}$ in the limit $a \rightarrow 1$.

The asymptotic expression for the likelihood ratio $\Lambda(u, a, \eta)$ is then given by

$$\begin{aligned} \Lambda(u, a, \eta) &\sim \frac{\eta \exp(f(u, \eta, a))}{2\sqrt{1-a}z(u, a, \eta) - \mathcal{O}\left(\exp\left(-\frac{1}{1-a^2}\right)\right)} - \mathcal{O}\left(\frac{\exp\left(-\frac{1}{1-a^2}\right)}{1-a^2}\right), (a \rightarrow 1), \quad (3.54) \\ &\sim \frac{\eta \exp\left(-\frac{\eta^2}{4(1+a)} + \mathcal{O}(\sqrt{1-a})\right)}{\frac{\sqrt{2}\eta}{\sqrt{1+a}} + \sqrt{2}(1-a)^{3/2}u - \mathcal{O}\left(\exp\left(-\frac{1}{1-a^2}\right)\right)} - \mathcal{O}\left(\frac{\exp\left(-\frac{1}{1-a^2}\right)}{1-a^2}\right), (a \rightarrow 1), \\ &\sim \frac{\eta \mathcal{O}(1)}{\frac{\sqrt{2}\eta}{\sqrt{1+a}} + \mathcal{O}((1-a)^{3/2})} - \mathcal{O}\left(\frac{\exp\left(-\frac{1}{1-a^2}\right)}{1-a^2}\right), (a \rightarrow 1), \end{aligned} \quad (3.55)$$

and hence $\Lambda(u, a, \eta) \rightarrow \mathcal{O}(1)$.

In the limit $a \rightarrow -1$ the terms $-\sqrt{\frac{1-a}{1+a}}\eta u$ and $\frac{(1-a)^2}{2}u^2$ in $f(u, a, \eta)$ and $z^2(u, a, \eta)$ do not vanish, i.e.,

$$f(u, a, \eta) \sim -\frac{\eta^2}{4(1+a)} - \sqrt{\frac{1-a}{1+a}}\eta u + \mathcal{O}(1), \quad (a \rightarrow -1), \quad (3.56)$$

$$z^2(u, a, \eta) \sim \frac{\eta^2}{2(1-a^2)} + \sqrt{\frac{1-a}{1+a}}\eta u + \mathcal{O}(1), \quad (a \rightarrow -1), \quad (3.57)$$

although the leading order terms are still $-\frac{\eta^2}{4(1+a)}$ and $\frac{\eta^2}{2(1-a^2)}$. In order to understand the role of the term $\sqrt{\frac{1-a}{1+a}}\eta u$ we have to keep in mind, that suitable precursors for increments are typically negative, i.e., $u < 0$. Thus the behavior of the leading order terms is not enforced by $\sqrt{\frac{1-a}{1+a}}\eta u$, but inhibited. In other words $f(u, a, \eta) \sim o(1/(1+a))$, if $u > 0$ and $f(u, a, \eta) \sim \mathcal{O}(1/(1+a))$, if $u < 0$ and analog for $z^2(u, a, \eta)$. In any case $f(u, a, \eta) \rightarrow -\infty$ and $z(u, a, \eta) \rightarrow -\infty$ in the limit $a \rightarrow -1$ and hence

$$\Lambda(u, a, \eta) \sim \frac{\eta \exp\left(-\frac{\eta^2}{4(1+a)} - \sqrt{\frac{1-a}{1+a}}\eta u + \mathcal{O}(1)\right)}{\frac{\sqrt{2}\eta}{\sqrt{1+a}} + \mathcal{O}(1) - \mathcal{O}\left(\exp\left(-\frac{1}{1-a^2}\right)\right)} - \mathcal{O}\left(\frac{\exp\left(-\frac{1}{1-a^2}\right)}{1-a^2}\right), (a \rightarrow -1) \quad (3.58)$$

$$\sim \frac{\eta \exp\left(-\frac{\eta^2}{4(1+a)} - \sqrt{\frac{1-a}{1+a}}\eta u + \mathcal{O}(1)\right)}{\frac{\sqrt{2}\eta}{\sqrt{1+a}} + \mathcal{O}(1)} - \mathcal{O}\left(\frac{\exp\left(-\frac{1}{1-a^2}\right)}{1-a^2}\right), (a \rightarrow -1) \quad (3.59)$$

$$\sim \frac{\mathcal{O}(\exp(-1/(1+a)))}{\frac{\sqrt{2}\eta}{\sqrt{1+a}} + \mathcal{O}(1)} - \mathcal{O}\left(\frac{\exp\left(-\frac{1}{1-a^2}\right)}{1-a^2}\right), (a \rightarrow -1). \quad (3.60)$$

Thus $\Lambda(u, a, \eta) \rightarrow 0$, as $a \rightarrow -1$.

If we demand now, that $u \in [(u^0)_-, (u^0)_+]$ (which also implies $u < 0$), we know, that $f(u, a, \eta) \geq 0$, as discussed in the previous section. This implies, that the absolute value of the term $\sqrt{\frac{1-a}{1+a}}\eta u$ is larger than $-\frac{\eta^2}{4(1+a)}$. In other words, a suitable chosen value of u ensures, that the term $\sqrt{\frac{1-a}{1+a}}\eta u$ dominates the overall behavior of $f(a, u, \eta)$ and not the term, which is of leading order in terms of $a \rightarrow \pm 1$. Hence since

$$\sqrt{\frac{1-a}{1+a}}\eta u \rightarrow 0, \quad (a \rightarrow 1), \quad (3.61)$$

$$\text{and } \sqrt{\frac{1-a}{1+a}}\eta u \rightarrow \infty, \quad (a \rightarrow -1), \quad (3.62)$$

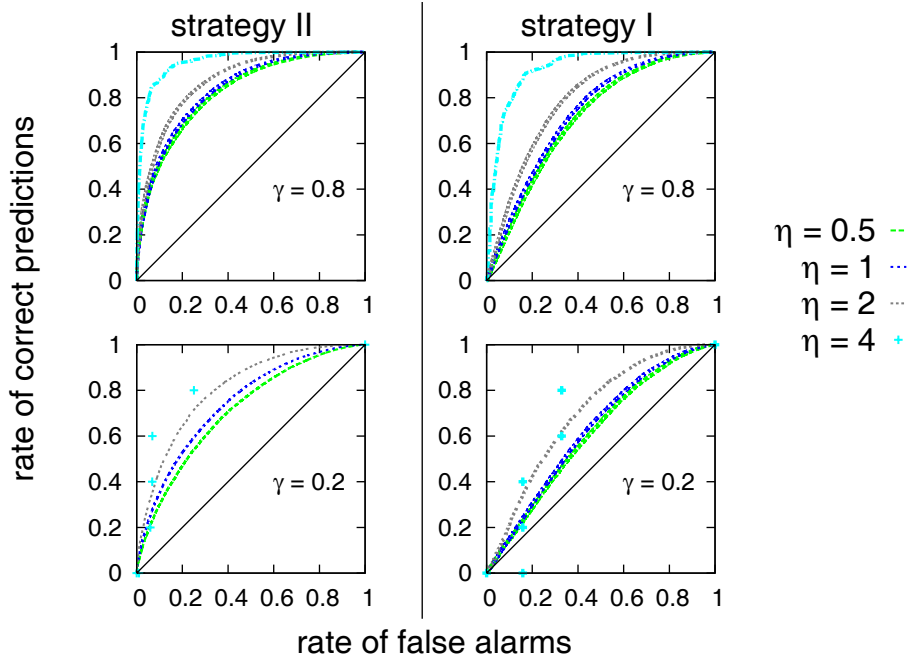


Figure 3.15: ROC curves for the ARMA(∞, ∞) processes with $\gamma_c = 0.2$ and $\gamma_c = 0.8$.

$f(u, a, \eta) \rightarrow \infty$, ($a \rightarrow 1$) and $f(u, a, \eta) \rightarrow 0$, ($a \rightarrow -1$). Consequently $\Lambda(u, a, \eta) \rightarrow 1$, ($a \rightarrow 1$) and $\Lambda(u, a, \eta) \rightarrow 0$, ($a \rightarrow -1$).

In total we can say, that the asymptotic expressions of the likelihood ratio reflect the behavior of the ROC curves, if the chosen precursor is within the interval $[(u^0)_-, (u^0)_+]$: The predictions become better as $a \rightarrow -1$ and worse for $a \rightarrow 1$.

However, the result for precursors outside the interval in the limit $a \rightarrow -1$, namely $\Lambda(u, a, \eta) \rightarrow 0$, ($a \rightarrow -1$), does not comply to the observations and is also a result which is difficult to understand in terms of the ROC curves.

3.5 Predicting extreme increments in long-range correlated processes

We study the same aspects as we investigated before in the short range correlated AR(1) processes, in long-range correlated processes. The idea is, to test whether the long-range correlation does qualitatively alter the results obtained for the prediction of extreme increments. Since the precursors we were interested in live on a very short time scale (one step before the event), one should not expect long-range correlations to lead to qualitatively different results for the aspects we were interested in.

There are various definitions of long-range correlation. Typically long-range correlation in a time series is characterized by the exponent $0 < \gamma_c < 1$ of the power-law decay of the autocorrelation function as a function of the time t

$$C_x(t) = \langle x_n x_{n+t} \rangle = \frac{1}{N-t} \sum_{n=1}^{N-t} x_n x_{n+t} \sim t^{-\gamma_c} \quad (3.63)$$

The correlation coefficient γ_c is controlling, how fast the correlations decay.

We study the predictability of increments numerically by applying the prediction strategies described in Sec. 3.2 and in chapter 2. The data used for this numerical study were generated as described in [92] and applied in [?]: Imposing a power-law decay on the Fourier spectrum,

$$f_x(k) \propto k^{-\beta} \quad (3.64)$$

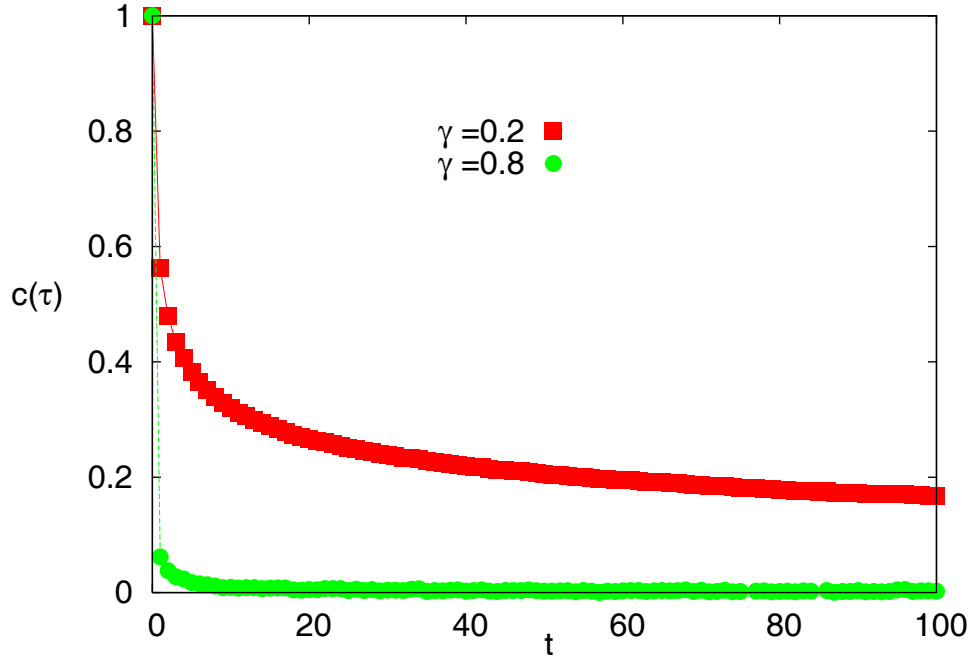


Figure 3.16: The autocorrelation function of the ARMA(∞, ∞) processes with $\gamma_c = 0.2$ and $\gamma_c = 0.8$.

with $0 < \beta < 0.5$ and choosing phase angles at random one obtains through an inverse Fourier transform the long-range correlated time series in x with $\gamma_c = 1 - 2\beta$. The data are Gaussian distributed with $\langle x \rangle = 0$, $\sigma = 1$. Having specified the power spectrum or, correspondingly, the autocorrelation function for sequences of Gaussian random numbers means to have fixed all parameters of a linear stochastic process. Hence, in principle the coefficients of an autoregressive or moving average process can be uniquely determined, where, due to the power-law nature of the spectrum and autocorrelation function the order of either of these models have to be infinite [34, 33]. Thus, the effects which we observed for this ARMA(∞, ∞) model should be valid for the whole class of linear long-term correlated processes.

The ROC curves in Fig. 3.15, which are generated from the long-range correlated data are very similar to the ones for the AR(1) process in terms of the question we want to study.

A1, superiority of strategy II:

The ROC curves obtained by using strategy II are superior to the curves resulting from strategy I.

A2 and A3, influence of the event magnitude and of the correlation

The quality of the prediction also increases with increasing event size and decreasing correlation. The decrease of the correlation is explicitly shown in Fig. 3.16

Hence we observe the same effects which we described before for the AR(1) process in a long range correlated ARMA(∞, ∞) process.

3.6 Summary

We studied the predictability of extreme increments in an AR(1) correlated process and in long-range correlated ARMA processes. To measure the quality of the prediction we used the ROC curve and additionally the slope of the ROC curve in the vicinity of the origin as a summary index. This so called *likelihood ratio*, characterizes particularly the behavior in the limit of low false-alarm rates.

In the case of the AR(1) process we could construct the posterior PDF and the likelihood analytically from a given joint PDF and hence we were able to obtain the asymptotic behavior

of the likelihood ratio analytically. In the case of the two other examples, we constructed the posterior PDFs numerically. The resulting distributions were then used to determine precursors according to two different strategies of prediction.

In all examples we studied the aspects : **(A1)** Which is the best strategy to choose precursors? **(A2)** How does the predictability depend on the event size? **(A3)** And how does the predictability depend on the correlation? The results can be summarized as follows:

superiority of strategy II

Strategy I, the a posteriori approach, maximizes the rate of correct predictions, while strategy II focuses on the minimization of the rate of false alarms.¹ For the example of the AR(1) process one can show that strategy II is the optimal strategy to make predictions. For other stochastic processes, it is not in general clear which of the two strategies leads to a better predictability. However, the application to the prediction of wind speeds and the numerical study within long-range correlated data reveals that also for these examples better results are obtained by predicting according to strategy II.

influence of the event magnitude:

The examples studied, we observe an increase of predictability with increasing size of the events. This phenomenon which is also reported for specific examples [20, 21, 22], can be discussed by investigating the asymptotic behavior of our summary index. In the case of the AR(1) process we showed explicitly that the likelihood ratio increases as a squared exponential with increasing event size.

influence of the correlation:

For the AR(1) process and the long-range correlated data we observe that the correlation of the data is inversely proportional to the quality of the predictions. This effect is due to the special definition of the events as increments. The asymptotic expression for the likelihood ratio in Eq. (3.33) provides us also with a formal understanding of the a -dependence.

¹Note that the terms maximization and minimization refer to changes in the integrand, which enters into the alarm rates as given by Eqs. (2.46) and (2.47) and not to changes in the integration ranges V_{pre} and V_- .

Chapter 4

Investigating the Influence of the Event Size

In the previous chapter, we saw that large increments are in the examples of the AR(1) processes and in the long-range correlated linear processes, the better predictable the larger they are. In the following we will call this effect *positive magnitude dependence* of the quality of a prediction. The previous chapter also revealed that we could describe the magnitude dependence investigating the asymptotic behavior of the likelihood ratio. In Chap. 2 we noticed that the likelihood ratio is essentially given by the likelihood to observe an event after a precursor occurred and by the total probability to find events. Both probabilities are derived from the joint PDF of event and precursor. Hence, we can expect to predict the occurrence of a (positive) magnitude dependence by simply knowing the joint PDF of event and precursor or the likelihood and the total probability to find events. In this chapter we will formulate a condition, based on the likelihood and the total probability, which can indeed predict the occurrence of a positive or negative magnitude dependence.

4.1 Magnitude Dependence of a Prediction

A survey of the literature reveals, that the better predictability of larger events was sporadically already reported by other authors. Lamper et al. [22] report an enhanced predictability prior to large changes in multi-agent games. They “measure” the predictability by visual comparison of the forecasted behavior of a time series and the verification. Furthermore Snirman et al. [23] found hints for the fact that larger avalanches in the Bak-Tang-Wiesenfeld sand pile are better predictable than smaller ones. They consider the error-curve which is typically used to characterize the efficiency of earthquake prediction algorithms as a measure for the quality of their predictions. For the prediction of wind speeds it was reported [36] that predictions focusing on larger wind gusts lead to better results in terms of better ROC curves.

An example from weather forecasting, which also uses the quantity which we call likelihood ratio is the work of Göber [21], who investigates “that forecasts of more extreme (rare) events have more skill than forecasts for more ‘normal’ events” for the prediction of precipitation. Events and non-events are in this case rainfall or no rainfall and Göber discusses the observed effect as an influence of the total probability to observe events. Furthermore he proposes a measure in which the influence of the total probability is eliminated. As a critical remark it has to be said that removing this explicit dependence on the total probability in the measure does not eliminate the implicit influence on the likelihood to find extreme events, as one can easily see by expressing the likelihood in terms of Bayes’ theorem. Furthermore one could argue, that the total probability plays the role of a normalizing factor in the likelihood ratio and in the ROC curve .

This last example raises also the interesting question whether the observed dependence on

the event magnitude is in fact not simply a dependence on the rareness, i.e., the low relative frequency or the low probability to observe events. We will discuss this question in more detail in Sec. 4.4.

Inspired from the above mentioned examples we dare to name the effect that larger events are observed to be better predictable in the following definition.

Definition 5 (Magnitude Dependence of a the Prediction Quality)

Suppose, that we have specified a certain measure \mathcal{L} for the quality of a prediction. Then we call a prediction of an event magnitude dependent under \mathcal{L} if the quality of the prediction changes as the magnitude of the events under study is varied.

positive magnitude dependence *If the quality of the prediction increases with increasing event magnitude, we call the prediction positively dependent on the event magnitude.*

negative magnitude dependence *If the quality of the prediction decreases with increasing event magnitude, we call the prediction negatively dependent on the event magnitude.*

magnitude independence *If we cannot find any dependence on the event magnitude we will call the prediction not magnitude dependent.*

Note that the magnitude dependence depends sensitively on the measure of the quality which is used. We will see in Sec. 4.3 that considerations made for one quality measure do in general **not** describe the magnitude dependence of another measure, even if both measures are applied to the same problem under study.

4.2 The Test-condition for Likelihood Ratios

We are now interested in learning how the predictability depends on the event magnitude η if we use ROC curves and their slopes in the vicinity of the origin, to characterize the quality of a prediction. As discussed already in Sec.2.3.4 for prediction via precursory structures, the slope of the ROC curve is given by the likelihood ratio

$$m[\mathbf{u}, \delta, \chi(\eta)] = \frac{\rho[\mathbf{u} + \delta | \chi_n(\eta) = 1]}{\rho[\mathbf{u} + \delta | \chi_n(\eta) = 0]} = \Lambda[\mathbf{u} + \delta, \chi(\eta)]. \quad (4.1)$$

and we can use the the slope in the vicinity of the origin

$$m[\mathbf{u}, \delta = 0, \chi(\eta)] = \Lambda[\mathbf{u}, \chi(\eta)] = \frac{\rho(\mathbf{u} | \chi_n(\eta) = 1)}{\rho(\mathbf{u} | \chi_n(\eta) = 0)} \quad (4.2)$$

as a summary index. In the following we will discuss the magnitude dependence of the likelihood ratio $\Lambda[\mathbf{s}_n, \chi(\eta)]$.

Via Bayes' theorem the likelihood ratio as introduced in Eq. 2.28 can be expressed in terms of the likelihood $p[\chi_n(\eta) = 1 | \mathbf{s}_n]$ and the total probability to find events $p[\chi_n(\eta) = 1]$.

$$\Lambda[\mathbf{s}_n, \chi_n(\eta)] = \frac{(1 - p[\chi_n(\eta) = 1])}{p[\chi_n(\eta) = 1]} \frac{\rho[\chi_n(\eta) = 1 | \mathbf{s}_n]}{(1 - \rho[\chi_n(\eta) = 1 | \mathbf{s}_n])}. \quad (4.3)$$

Keeping in mind that the likelihood and the total probability were in fact calculated from the joint PDF $p(\mathbf{s}_n, \chi_n)$ ¹ of the precursor time series $\{\mathbf{s}_n\}$ and the time series of events and non-events $\{\chi_n\}$ (see App. B), we can reformulate the expression as

$$\Lambda[\mathbf{s}_n, \chi_n(\eta)] = \frac{\left(1 - \int_{-\infty}^{\infty} d\mathbf{s}_n \Theta(f(\chi_n)) p(\mathbf{s}_n, \chi_n)\right)}{\int_{-\infty}^{\infty} d\mathbf{s}_n \Theta(f(\chi_n)) p(\mathbf{s}_n, \chi_n)} \frac{\Theta(f(\chi_n)) p(\mathbf{s}_n, \chi_n)}{\rho(\mathbf{s}_n) \left(1 - \frac{\Theta(f(\chi_n)) p(\mathbf{s}_n, \chi_n)}{\rho(\mathbf{s}_n)}\right)}. \quad (4.4)$$

¹Note that the notation $p[\mathbf{s}_n, \chi_n]$ describes the joint PDF of the precursor and of both, events and non-events, i.e., the filtering for events did not take place yet.

$\Theta(f(\chi_n))$ denotes the event filter as described in App. B and the function $f(\chi_n)$ has to be specified according to the definition of the event. The filter $\Theta(\cdot)$ for events defined as increments and threshold crossings as it is used in App. B leads to the following expression

$$\Lambda[\mathbf{s}_n, \chi_n(\eta)] = \frac{\left(1 - \int_{-\infty}^{\infty} d\mathbf{s}_n p(\mathbf{s}_n, \chi_n(\eta) = 1)\right)}{\int_{-\infty}^{\infty} d\mathbf{s}_n p(\mathbf{s}_n, \chi_n(\eta) = 1)} \frac{p(\mathbf{s}_n, \chi_n(\eta) = 1)}{\rho(\mathbf{s}_n) \left(1 - \frac{p(\mathbf{s}_n, \chi_n(\eta) = 1)}{\rho(\mathbf{s}_n)}\right)},$$

with $p(\mathbf{s}_n, \chi_n(\eta) = 1) = \int_{\mathcal{M}} dx_{n+\kappa} p(\mathbf{s}_n, x_{n+\kappa})$, $\mathcal{M} = \{x_{n+\kappa} : \chi_n = 1\}$, (4.5)

and $\rho(\mathbf{s}_n) = p[\mathbf{s}_n, \chi_n(\eta) = 1] + p[\mathbf{s}_n, \chi_n(\eta) = 0]$.

Looking at the rather technical formula in Eq. (4.5), there are three remarkable aspects:

Characterization via the joint PDF: Once the event and precursory variable are defined, (thus implicitly also set \mathcal{M} is defined,) the slope of the ROC-plot is fully characterized by the knowledge of the joint PDF $p[\mathbf{s}_n, \chi_n]$ of the precursory variable \mathbf{s}_n and the event variable χ_n .

Dependence on the correlation: The characterization via the joint PDF implies that in the framework of statistical predictions, all kind of (long-range) **correlations** which might be present in the time series **influence the quality of the predictions only through their influence on the joint PDF of events and precursors.**

Influence of the event type: The definition of the event, e.g., as a threshold crossing or an increment enters only into the set on which the integrals in Eq. (4.4) are carried out. Both $\chi_n(\eta)$ and the borders of the integrals in Eq. (4.4) have to be defined according to the type of events under study.

Exploiting Eq. (4.4) we can hence determine the dependence of the likelihood ratio and the ROC curve on the event magnitude η , via the dependence of the joint PDF of the process under study.

In order to simplify the expressions we prefer to work with the PDFs derived from the joint PDF, namely the likelihood to observe an extreme event after a certain precursor value $p[\chi_n(\eta) = 1|\mathbf{s}_n]$ and the total probability to observe events $p[\chi_n(\eta) = 1]$. Our aim is to find constraints or conditions which the likelihood and the total probability to find events have to fulfill in order to find a better predictability of larger (smaller) events. A simple way to arrive at such a condition consists in asking for the change of the likelihood ratio with increasing event magnitude η .

$$\frac{\partial}{\partial \eta} \Lambda[\mathbf{s}_n, \chi_n(\eta)] \underset{\leq}{\underset{\geq}} 0. \quad (4.6)$$

Before we continue, we introduce the following notations for the posterior PDFs, the likelihood and the total probability to find events, in order to improve the readability of the following calculations,

$$L := p[\chi_n(\eta) = 1|\mathbf{s}_n], \quad (4.7)$$

$$L' := \frac{\partial}{\partial \eta} p[\chi_n(\eta) = 1|\mathbf{s}_n], \quad (4.8)$$

$$P := p[\chi_n(\eta) = 1], \quad (4.9)$$

$$P' := \frac{\partial}{\partial \eta} p[\chi_n(\eta) = 1], \quad (4.10)$$

$$\Lambda' = \frac{\partial}{\partial \eta} \Lambda[\mathbf{s}_n, \chi_n(\eta)]. \quad (4.11)$$

Computing the derivative and rearranging the equation yields

$$\begin{aligned}
\Lambda' &= \frac{\partial}{\partial \eta} \frac{L - LP}{P - PL}, \\
&= \frac{(L' - L'P - P'L)(P - PL) - (P' - P'L - L'P)(L - LP)}{(P - PL)^2} \\
&= \frac{PL' - P^2L' - PLP' - PLL' + P^2LL' + PL^2P'}{(P - PL)^2} \\
&\quad + \frac{-LP' + L^2P' + PLL' + LPP' - L^2PP' - LP^2L'}{(P - PL)^2} \\
&= \frac{L'(P - P^2 - PL + P^2L + PL - P^2L)}{(P - PL)^2} \\
&\quad + \frac{P'(-PL + PL^2 - L + L^2 + LP - L^2P)}{(P - PL)^2} \\
&= \frac{PL'(1 - P) - LP'(1 - L)}{P^2(1 - L)^2} \\
&= \frac{L'(1 - P)}{P(1 - L)^2} - \frac{P'L}{P^2(1 - L)} \\
&= \frac{1}{P(1 - L)} \left(\frac{L'(1 - P)}{(1 - L)} - \frac{P'L}{P} \right) \tag{4.12}
\end{aligned}$$

Typically $0 < L(\mathbf{s}_n, \eta) < 1, 0 < P(\mathbf{s}_n, \eta) < 1$ for all finite values of the precursory variable due to their definition as PDFs and hence we can manipulate the equations:

$$\Lambda' \begin{matrix} \geq \\ < \end{matrix} 0 \Leftrightarrow \frac{L'(1 - P)}{P(1 - L)} - \frac{P'L}{P^2} \begin{matrix} \geq \\ < \end{matrix} 0, \tag{4.13}$$

$$\begin{aligned}
&\Leftrightarrow \frac{L'}{L} - \frac{(1 - L)P'}{(1 - P)P} \begin{matrix} \geq \\ < \end{matrix} 0, \quad (0 < L < 1), (0 < P < 1). \\
&\tag{4.14}
\end{aligned}$$

Thus if $0 < L(\mathbf{s}_n, \eta) < 1, 0 < P(\mathbf{s}_n, \eta) < 1$, the derivative of the likelihood ratio is positive (negative, zero), if the following sufficient condition $c(\eta, \mathbf{s}_n)$ is fulfilled

$$c(\eta, \mathbf{s}_n) = \frac{\partial}{\partial \eta} \ln p[\chi_n(\eta) = 1 | \mathbf{s}_n] - \frac{(1 - p[\chi_n(\eta) = 1 | \mathbf{s}_n])}{(1 - p[\chi_n(\eta) = 1])} \frac{\partial}{\partial \eta} \ln p[\chi_n(\eta) = 1] \begin{matrix} \geq \\ < \end{matrix} 0. \tag{4.15}$$

In the special case $L = 1$ and $0 < P < 1$ Eq. (4.12) diverges either to ∞ or $-\infty$ in dependence on the sign of L' . In case that $L = 0$ and $0 < P < 1$ we are left with the following expression

$$\Lambda' = \frac{L'(1 - P)}{P}, \quad (L = 0), (0 < P < 1). \tag{4.16}$$

The case $P = 1$ corresponds to a tracer time series which consists only of events, thus we can conclude, that in this case also $L = 1$ holds for all possible precursors. Both assumptions lead to

$$\Lambda' \rightarrow \frac{L'(1 - P)}{(1 - L)^2} - \frac{P'}{(1 - L)}, \quad (P \rightarrow 1), (L \rightarrow 1). \tag{4.17}$$

Since the first term in Eq. (4.17) is not well defined, and we cannot exclude that $(P \rightarrow 1)$ faster than $(L \rightarrow 1)$, it is hard to say which is the dominating term. However since the case $P = 1, L = 1$ corresponds to the perfect prediction scenario (always correct prediction, no false alarms), represented by a ROC curve situated on the vertical axis, we know that the likelihood ratio has in this case the value ∞ and thus no improvement due to a change in the event size is possible.

Analogously the case $P = 0$ also implies $L = 0$ which leads as well to a derivative of Λ' , which contains not well defined expressions,

$$\Lambda' \rightarrow \frac{L'}{P} - \frac{P'L}{P^2}, \quad (P \rightarrow 0), (L \rightarrow 0). \quad (4.18)$$

In total we can summarize these considerations to the following conditions to test the magnitude dependence.

Condition 1 (Condition to test the magnitude dependence)

(a) If $0 \leq p[\chi_n(\eta) = 1|\mathbf{s}_n] < 1$ and $0 < p[\chi_n(\eta) = 1] < 1$,

$$\text{sign} \left(\frac{\partial}{\partial \eta} \Lambda[\mathbf{s}_n, \chi_n(\eta)] \right) = \text{sign} c(\eta, \mathbf{s}_n), \quad (4.19)$$

$$\text{with } c(\eta, \mathbf{s}_n) = \frac{\partial}{\partial \eta} \ln p[\chi_n(\eta) = 1|\mathbf{s}_n] - \frac{(1 - p[\chi_n(\eta) = 1|\mathbf{s}_n])}{(1 - p[\chi_n(\eta) = 1])} \frac{\partial}{\partial \eta} \ln p[\chi_n(\eta) = 1],$$

holds. Consequently,

if $c(\eta, \mathbf{s}_n) > 0$, the corresponding families of likelihood ratios and ROC curves, parametrized by η show a positive magnitude dependence,

if $c(\eta, \mathbf{s}_n) < 0$, the corresponding families of likelihood ratios and ROC curves show a negative magnitude dependence and

if $c(\eta, \mathbf{s}_n) = 0$, likelihood ratio and ROC curves are not dependent on the event magnitude.

(b) If $p[\chi_n(\eta) = 1|\mathbf{s}_n] = 0$ and $0 < p[\chi_n(\eta) = 1] < 1$,
or if $p[\chi_n(\eta) = 1|\mathbf{s}_n] = 1$ and $0 < p[\chi_n(\eta) = 1] < 1$, then

$$\text{sign} \left(\frac{\partial}{\partial \eta} \Lambda[\mathbf{s}_n, \chi_n(\eta)] \right) = \text{sign} \left(\frac{\partial}{\partial \eta} p[\chi_n(\eta) = 1|\mathbf{s}_n] \right). \quad (4.20)$$

Consequently,

if $\frac{\partial}{\partial \eta} p[\chi_n(\eta) = 1|\mathbf{s}_n] > 0$ the corresponding families of likelihood ratios and ROC curves, parametrized by η show a positive magnitude dependence,

if $\frac{\partial}{\partial \eta} p[\chi_n(\eta) = 1|\mathbf{s}_n] < 0$ the corresponding families likelihood ratios and ROC curves show a negative magnitude dependence and

if $\frac{\partial}{\partial \eta} p[\chi_n(\eta) = 1|\mathbf{s}_n] = 0$ the likelihood ratio and the ROC curve are not dependent on the event magnitude.

(c) For $(p[\chi_n(\eta) = 1] = 1)$ which implies $(p[\chi_n(\eta) = 1|\mathbf{s}_n] = 1)$ and $(p[\chi_n(\eta) = 0])$ which implies $(p[\chi_n(\eta) = 1|\mathbf{s}_n] = 0)$ the derivative of the likelihood ratio is not well defined.

Remarks:

addressing (a) The conditions leading to case (a) describe a typical scenario, in which prediction via precursory structures is required. Hence in most examples for the prediction via precursory structures the condition $c(\eta, \mathbf{s}_n)$ is needed, whereas the (b) and (c) are less common.

addressing (b) In theory the case $p[\chi_n(\eta) = 1|\mathbf{s}_n] = 1$ only occurs for predictions in a deterministic system. However we can also find it as a numerical artefact in any finite time series containing a finite number of events. Note that the condition in (a) is formulated in a way which prevents this numerical artefact to cause a divergence of $c(\eta, \mathbf{s}_n)$, i.e., in principle also the formulation

$$\tilde{c}(\eta, \mathbf{s}_n) = \frac{(1 - p[\chi_n(\eta) = 1])}{(1 - p[\chi_n(\eta) = 1|\mathbf{s}_n])} \frac{\partial}{\partial \eta} \ln p[\chi_n(\eta) = 1|\mathbf{s}_n] - \frac{\partial}{\partial \eta} \ln p[\chi_n(\eta) = 1] \quad (4.21)$$

could have been considered.

addressing (c) Note, that the cases of $p[\chi_n(\eta) = 1] = 1$, i.e., every data point is an extreme event, and $p[\chi_n(\eta) = 1] = 0$, no extreme events are observed so far do not correspond to situations in which one would ask for predictions based on precursors in time series.

Although the considerations leading to this theorem are extremely simple, Eq. (4.19) turns out to be a useful tool in order to characterize the occurrence of the magnitude dependence. Consequently one can tell for an arbitrary process, if extreme events are better predictable, by simply testing, if the marginal PDF of the event and the likelihood of event and precursor fulfill Eq. (4.19).

4.3 The magnitude dependence of the Kullback Leibler distance

Having formulated a condition which describes the magnitude dependence for the likelihood ratio and for ROC curves, we could in principle derive similar conditions for other quality measures. Since we saw that the likelihood ratio is closely linked to the Kullback Leibler distance we can also ask, whether the corresponding magnitude dependences can be described by the same condition. Evaluating the magnitude dependence for the Kullback Leibler distance leads to

$$\begin{aligned} \frac{\partial}{\partial \eta} D(p(\eta, \mathbf{s}_n) || q(\eta, \mathbf{s}_n)) &= \int d\mathbf{s}_n \log \left(\frac{p(\eta, \mathbf{s}_n)}{q(\eta, \mathbf{s}_n)} \right) \left(\frac{\partial}{\partial \eta} p(\eta, \mathbf{s}_n) \right) \\ &+ \int d\mathbf{s}_n p(\eta, \mathbf{s}_n) \frac{q(\eta, \mathbf{s}_n)}{p(\eta, \mathbf{s}_n)} \left(\frac{\partial}{\partial \eta} \frac{p(\eta, \mathbf{s}_n)}{q(\eta, \mathbf{s}_n)} \right). \end{aligned}$$

Substituting $p(\eta, \mathbf{s}_n)$ by $\rho[\mathbf{s}_n|\chi_n(\eta) = 1]$ and $q(\eta, \mathbf{s}_n)$ by $\rho[\mathbf{s}_n|\chi_n(\eta) = 0]$ yields

$$\begin{aligned} \frac{\partial}{\partial \eta} D(\rho[\mathbf{s}_n|\chi_n(\eta) = 1] || \rho[\mathbf{s}_n|\chi_n(\eta) = 0]) &= \int d\mathbf{s}_n \log (\Lambda(\mathbf{s}_n, \chi_n(\eta))) \left(\frac{\partial}{\partial \eta} \rho[\mathbf{s}_n|\chi_n(\eta) = 1] \right) \\ &+ \int d\mathbf{s}_n \rho[\mathbf{s}_n|\chi_n(\eta) = 0] \left(\frac{\partial}{\partial \eta} \Lambda(\mathbf{s}_n, \chi_n(\eta)) \right). \end{aligned} \quad (4.22)$$

This Kullback Leibler distance evaluates the separability of the CPDFs according to strategy I. Equation (4.22) displays, that the sign of $\frac{\partial}{\partial \eta} \Lambda(\mathbf{s}_n, \chi_n(\eta))$ does not determine the sign of $\frac{\partial}{\partial \eta} D(\rho[\mathbf{s}_n|\chi_n(\eta) = 1] || \rho[\mathbf{s}_n|\chi_n(\eta) = 0])$. The analog holds for the Kullback Leibler distance, which describes the separability of the CPDFs according to strategy II:

$$\begin{aligned} \frac{\partial}{\partial \eta} D(p[\chi_n(\eta) = 1|\mathbf{s}_n] || p[\chi_n(\eta) = 0|\mathbf{s}_n]) &= \int d\mathbf{s}_n \log B(\mathbf{s}_n, \chi_n(\eta)) \left(\frac{\partial}{\partial \eta} p[\chi_n(\eta) = 1|\mathbf{s}_n] \right) \\ &+ \int d\mathbf{s}_n p[\chi_n(\eta) = 0|\mathbf{s}_n] \left(\frac{\partial}{\partial \eta} B(\mathbf{s}_n, \chi_n(\eta)) \right), \end{aligned} \quad (4.23)$$

with $B(\mathbf{s}_n, \chi_n(\eta))$ denoting the ratio of odds, as defined in Chap. 2.

4.4 Frequently asked Questions

- **Is the better predictability of larger events not simply caused by an increased signal to noise ratio?**

It is sometimes suggested, that the better detectability of larger events can be simply described by an increased signal to noise ratio. In other words, larger events are expected to be better predictable since they and their precursors differ significantly from the typical behavior of the system under study. Following this intuitive explanation we would expect to observe a better predictability of larger events in any case. This is in contrast to the previous investigation for the AR(1) process. We saw, that there are regimes of parameters for which we observe a positive magnitude dependence in terms of a diverging likelihood ratio and an improved ROC statistics. However, we also discussed that for other values of e.g., the precursor or the coupling strength, the likelihood ratio does not show a positive magnitude dependence in any case. Furthermore Chap. 5 will provide us with more counter examples. Thus, the intuitive explanation via an increased signal to noise ratio is not sufficient to describe the phenomenon under all circumstances and instead the condition in Eq. (4.19) should be applied.

- **Is the better predictability of larger events not simply a better predictability of rarer events?**

Eq. (4.19) can also serve to discuss the question, whether the dependence on the magnitude is a dependence on the rareness, i.e., the decreasing total probability to observe events. Most PDFs have the property that larger events are rarer and this dependence enters into the explicit formula of the joint PDF of event and precursor via the definition of the event magnitude. Since the likelihood and the total probability to find events are derived from the joint PDF, this dependence consequently propagates into the likelihood.

Visualizing the dependence of the likelihood on the “rareness”, i.e., the total probability of events, one can simply express the likelihood via Bayes’ theorem

$$p[\chi_n(\eta) = 1 | \mathbf{s}_n] = \frac{\rho[\mathbf{s}_n | \chi_n(\eta) = 1] p[\chi_n(\eta) = 1]}{\rho(\mathbf{s}_n)}. \quad (4.24)$$

Inserting this and the analog for the likelihood that a precursor is followed by a non-event into the likelihood ratio, the terms involving the total probabilities cancel

$$m[\mathbf{s}_n, \chi_n(\eta)] = \frac{(1 - p[\chi_n(\eta) = 1])}{p[\chi_n(\eta) = 1]} \frac{\rho[\mathbf{s}_n | \chi_n(\eta) = 1] p[\chi_n(\eta) = 1]}{\rho(\mathbf{s}_n)} \frac{\rho(\mathbf{s}_n)}{\rho[\mathbf{s}_n | \chi_n(\eta) = 0] (1 - p[\chi_n(\eta) = 1])}. \quad (4.25)$$

Thus, the “rareness” in terms of the total probability to find events plays an important role in the likelihood ratio. However, according to our understanding this role consists more in normalizing the likelihood than in giving an own contribution to the observed effect. One can argue that it is in fact the normalization with the total probability which insures that ROC curves are independent on the total probability, as we argued already in Sec. 2.3.4.

- **How do the results for the ROC curve relate to predictions which are evaluated using the “(ad your favorite measure of predictability here)-score”?**

As we saw on the example of the Kullback Leibler distance discussed in Sec. 4.3, one cannot conclude that the magnitude dependence found for the ROC curve or the likelihood ratio can describe the magnitude dependence of other scores.

4.5 Summary

In this chapter we saw, that for different examples of prediction tasks and different measures for the quality of the predictions the effect, that larger events are better predictable was reported. We therefore call this effect the positive magnitude dependence of the prediction quality, allowing also the existence of the opposite effect, called a negative magnitude dependence of the prediction quality. Furthermore we discussed that the likelihood ratio and hence the ROC curves are fully determined by the joint PDF of precursor variables and tracer time series and we derived a condition which can predict, whether likelihood ratios and ROC curves display a positive/negative or no magnitude dependence. Unfortunately the statements derived from this condition do not describe the magnitude dependence for other measures for the quality of a prediction, such as the Kullback Leibler distance. Having all preliminary requisites specified we can now in the following chapter determine the magnitude dependence of ROC curves for the prediction of increments and threshold crossings in stochastic processes and real world data.

Chapter 5

Predictions of Increments in i.i.d. random numbers

In this chapter we test the condition $c(\eta, \mathbf{s}_n)$ derived in the previous chapter for the prediction of increments in Gaussian, exponentially, symmetrized exponentially, power-law and symmetrized power-law distributed i.i.d. random numbers. As in chapter 3 we concentrate on extreme events which consist of a sudden increase (or decrease) of the observed variable within the next time steps, namely

$$x_{n+1} - x_n \geq \eta, \quad (5.1)$$

where x_n and x_{n+1} denote the observed values at two consecutive time steps and the event magnitude η is again measured in units of the standard deviation.

Since the first part of the increment x_n can be used as a precursory variable, the definition of the event as an increment introduces a correlation between the event and the precursory variable x_n . We already saw this in the previous studies on the AR(1) process, since we could even predict increments in an AR(1) process with correlation strength $a = 0$, i.e., a sequence of random numbers. Hence the prediction of increments in random numbers provides a simple, but not unrealistic example which allows us to study the influence of the distribution of the underlying process on the event-magnitude dependence of the quality of prediction. In the examples which we study in this section the joint PDF of precursory variable and event is known and we can hence evaluate $c(\eta, x_n)$ analytically. However we will also develop an algorithm which evaluates $c(\eta, x_n)$ numerically. We focus in this chapter on predictions made according to strategy II, as described in Sec. 2.2.3.

5.1 Numerical evaluation of the test condition

Since in most cases the structure of the PDF is not known analytically, we are also interested in developing a numerical method of evaluating $c(\eta, x_n)$. For the numerical computation of $c(\eta, x_n)$ the approximations of the total probability and the likelihood are obtained by “binning and counting” and their numerical derivatives are calculated using Savitzky-Golay filters [93, 94]. The numerical evaluation is done within 10^7 data points. In order to check the stability of this procedure, we compute $c(\eta, x_n)$ also on 20 bootstrap samples which are generated from the original data set. These bootstrap samples consist of 10^7 pairs of event and precursory variable, which were drawn randomly from the original data set. Thus, their PDFs are slightly different and they contain different numbers of events. Evaluating $c(\eta, x_n)$ on the bootstrap samples thus shows how sensitive our numerical evaluation procedure is towards fluctuations in the distributions and in the numbers of events. This is especially important for large and therefore rare events for which only few data are available.

In order to check the results obtained by the evaluation of $c(\eta, x_n)$, we compute also the corresponding ROCs analytically and numerically. Note that for both, the numerical evaluation

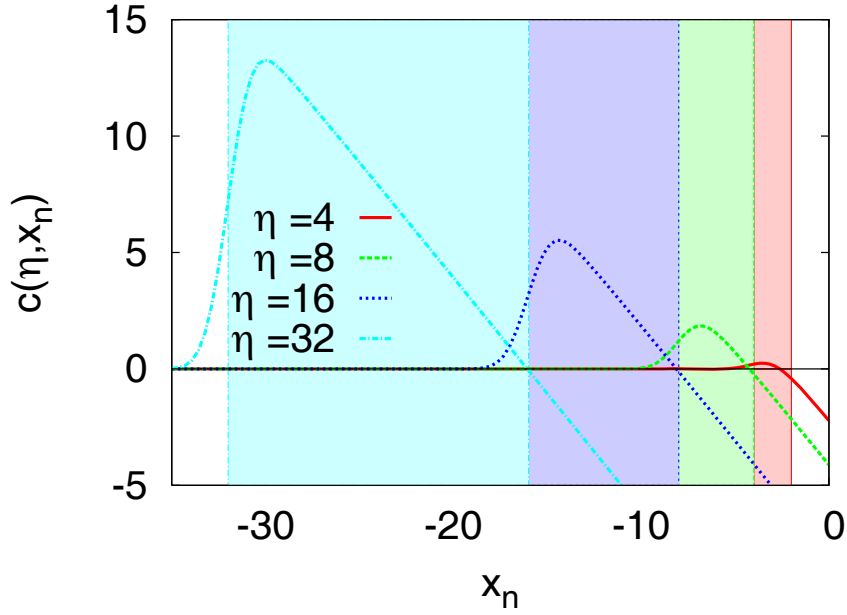


Figure 5.1: The condition $c(\eta, x_n)$ for the Gaussian distribution as given by Eq. (5.9). The color shaded regions indicate the intervals $[-\sigma\eta, -\eta/2]$ for which we can expect $c(\eta, x_n)$ to be positive according to Eq. (5.10). If $x_n < -\sigma\eta$, $\eta > 2\sqrt{\pi}$ and terms of the order of $\exp(-(x_n + \sigma\eta)^2)$ are sufficiently small, the condition is also positive according to Eq. (5.11). If terms of the order of $\exp(-(x_n + \sigma\eta)^2)$ cannot be neglected one also might find small regions in $(-\infty, -\sigma\eta]$ for which $c(\eta, x_n) < 0$. However, the influence of these regions is negligible, since our alarm interval is defined as $[-\infty, \delta]$ which implies an averaging over several possible values of the precursory variable.

of the condition and the numerically evaluated ROC curves, we used only event magnitudes η for which we found at least 1000 events, so that the observed effects are not due to a lack of statistics of the large events.

5.2 Gaussian distributed random numbers

In the first example we assume the sequence of i.i.d. random numbers which forms our time series to be normal distributed. As we know from chapter 3, increments within Gaussian random numbers are the better predictable, the more extreme they are. In this section we will show that their PDFs fulfill also the condition in Eq. (4.19). Applying the filter mechanism developed in App. B we obtain the following expressions for the a posteriori PDFs

$$\rho[x_n | \chi_n(\eta) = 1, \sigma] = \frac{\exp\left(-\frac{x_n^2}{2\sigma^2}\right)}{2\sqrt{2\pi}\sigma p[\chi_n(\eta) = 1]} \operatorname{erfc}\left(\frac{x_n + \sigma\eta}{\sigma\sqrt{2}}\right), \quad (5.2)$$

and the likelihood

$$\rho[\chi_n(\eta) = 1 | x_n, \sigma] = \frac{1}{2} \operatorname{erfc}\left(\frac{x_n + \sigma\eta}{\sigma\sqrt{2}}\right). \quad (5.3)$$

We do not show the explicit calculation for this PDFs here, since they can be obtained easily from the considerations made for the AR(1) process in Chap. 3 (see Sec. 3.1) by setting the coupling strength a of the AR(1) process to zero. We recall that the optimal precursor is given by the value of x_n which maximizes the likelihood. We refer to this special value of the precursory variable x_n by u and find for the likelihood according to Eq. (5.3) $u = -\infty$. Thus, instead of a finite alarm volume δ here is the upper limit of the interval $[-\infty, \delta]$. In contrast to the example

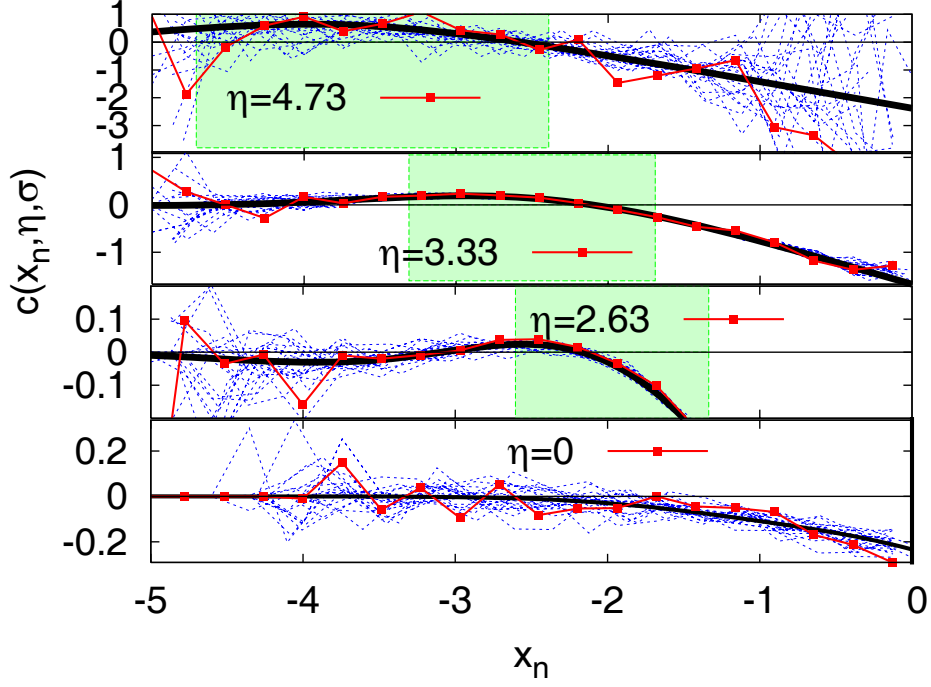


Figure 5.2: Comparison of the numerically evaluated condition $c(\eta, x_n, \sigma)$ for the Gaussian distribution and the expression given by Eq. (5.9). The black curves denote the evaluation of the analytic result in Eq. (5.9), the curves plotted with lines and symbols represent the numerical results obtained from the original data set, and the dashed lines represent the results obtained from the corresponding bootstrap samples. The green shaded regions indicate the regime $-\sigma\eta < x_n < -\sigma\eta/2$ for which $c(\eta, x_n, \sigma)$ is positive in the limit $\eta \rightarrow \infty$ according to Eq. (5.10). If $x_n < -\sigma\eta$, $\eta > 2\sqrt{\pi}$ and terms of the order of $\exp(-(x_n + \sigma\eta)^2)$ are sufficiently small, the condition is also positive according to Eq. (5.11). The numerical evaluation of $c(\eta, x_n, \sigma)$ was done by sampling the likelihood and the total probability of events from 10^7 random numbers.

of the AR(1) process we can calculate the total probability to find events exactly:

$$\begin{aligned}
 p[\chi_n(\eta) = 1] &= \frac{1}{2\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} dx_n \exp\left(-\frac{x_n^2}{2\sigma^2}\right) \operatorname{erfc}\left(\frac{x_n + \sigma\eta}{\sigma\sqrt{2}}\right) \\
 &= \frac{1}{2\pi\sigma} \int_{-\infty}^{\infty} dx_n \exp\left(-\frac{x_n^2}{2\sigma^2}\right) \int_{x_n + \sigma\eta}^{\infty} dy \exp\left(-\frac{y^2}{2\sigma^2}\right), \quad (5.4)
 \end{aligned}$$

due to the definition of the complementary error function and substitution with $y = x_n + \sigma\eta$. Substitution with $v = y - x_n$ and $u = y + x_n$ yields

$$p[\chi_n(\eta) = 1] = \frac{1}{2\pi\sigma} \int_{-\infty}^{\infty} du \exp\left(-\frac{u^2}{4\sigma^2}\right) \int_{\sigma\eta}^{\infty} dv \exp\left(-\frac{v^2}{4\sigma^2}\right) \quad (5.5)$$

$$= \frac{1}{2} \int_{\sigma\eta}^{\infty} dv \exp\left(-\frac{v^2}{4\sigma^2}\right) \quad (5.6)$$

$$= \frac{\sigma}{2} \int_{\eta/2}^{\infty} dt \exp(-t^2) \quad (5.7)$$

$$= \frac{\sigma}{2} \operatorname{erfc}\left(\frac{\eta}{2}\right). \quad (5.8)$$

Hence the condition in Eq. (4.19) reads

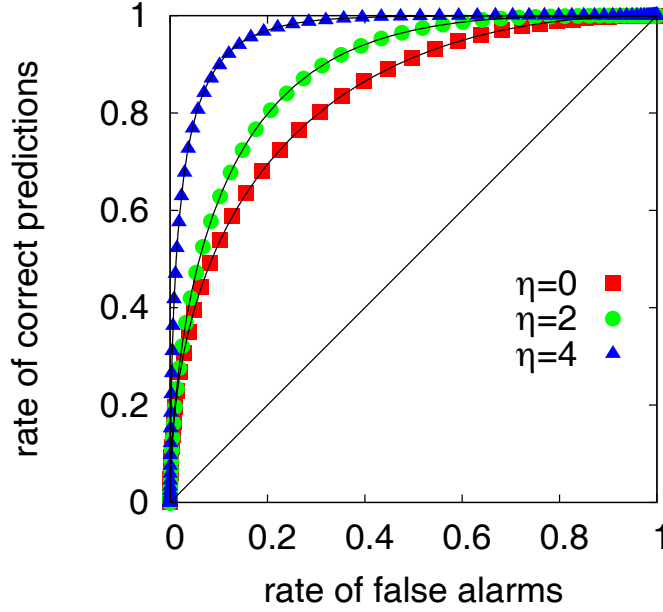


Figure 5.3: ROCs for Gaussian distributed i.i.d. random variables. The symbols represent ROC curves which were made via predicting increments in 10^7 normal i.i.d. random numbers. The predictions were made according to strategy II as described in Sec. 2.2.3. The lines represent the results of evaluating the rate of correct predictions and false alarms according to Eqs. (2.46) and (2.47) for the Gaussian case. Note that the quality of the prediction increases with increasing event magnitude.

$$c(\eta, x_n, \sigma) = -\sqrt{\frac{2}{\pi}} \frac{\exp(-z^2)}{\operatorname{erfc}(z)} + \frac{1}{\sqrt{\pi}} \frac{\exp\left(-\frac{\eta^2}{4}\right)}{\operatorname{erfc}\left(\frac{\eta}{2}\right)} \frac{\left(1 - \frac{1}{2}\operatorname{erfc}(z)\right)}{\left(1 - \frac{1}{2}\operatorname{erfc}\left(\frac{\eta}{2}\right)\right)},$$

with $z = \frac{x_n + \sigma\eta}{\sqrt{2}\sigma}$ (5.9)

Fig. 5.1 illustrates this expression and Fig. 5.2 compares it to the numerical results. For the ideal precursor $x_n = u = -\infty$ the condition $c(\eta, x_n)$ is—according to Eq. (5.9)—zero, since in this case, the slope of the ROC-curve tends to infinity (see chapter 3) and does not react to any variation in η . For any finite value of the precursory variable $x_n < 0$ we have to distinguish three regimes of $z = (x_n + \sigma\eta)/\sqrt{2}\sigma$, namely, $z \rightarrow \infty$ or $z \rightarrow -\infty$ and finally also the case $z = 0$.

In the first case we study the behavior of $c(\eta, x_n, \sigma)$ for a fixed value of the precursory variable $-\sigma\eta < x_n$ and $\eta \rightarrow \infty$. This implies that $z \rightarrow \infty$ and we can use the asymptotic expansion for large arguments of the complementary error function

$$\operatorname{erfc}(z) \sim \frac{\exp(-z^2)}{\sqrt{\pi}z} \left(1 + \sum_{m=1}^{\infty} (-1)^m \frac{1 \cdot 3 \dots (2m-1)}{(2z^2)^m} \right),$$

$\left(z \rightarrow \infty, |\arg z| < \frac{3\pi}{4} \right)$

as given in [89] to obtain

$$c(\eta, x_n, \sigma) \propto -\frac{x_n}{\sigma} + \frac{\eta}{2}, \quad -\sigma\eta < x_n < 0. \quad (5.10)$$

This expression is appropriate for $x_n > -\sigma\eta$ since the asymptotic expansion in Eq. (3.19) holds only if the argument of the complementary error function is positive. In this case $c(\eta, x_n, \sigma)$ is larger than zero, if x_n is fixed and finite and $-\sigma\eta < x_n < -\sigma\eta/2$.

In the second case, we assume $\eta \gg 1$ to be fixed, $x_n < -\sigma\eta$ and $x_n \rightarrow -\infty$. Hence we can use the expansion in Eq. (3.19) only to obtain the asymptotic behavior of the dependence on η

and not for the dependence on z . An asymptotic expression of $c(\eta, x_n, \sigma)$ hence reads

$$c(\eta, x_n, \sigma) \propto \frac{\eta}{2(1 - \frac{1}{2}\text{erfc}(\eta/2))} \left(\frac{\text{erf}(z)}{\sqrt{\pi}} + \frac{\eta}{2} \right) - \mathcal{O}(\exp(-z^2)), \quad x_n < -\sigma\eta. \quad (5.11)$$

Since $\text{erf}(z)$ tends to minus unity as $z \rightarrow -\infty$ the expression in Eq. (5.11) is positive if $\eta > 2\sqrt{\pi}$ and if the squared exponential term is sufficiently small. If the latter assumption is not fulfilled one might observe some regions of intermediate values of $-\infty < x_n < -\sigma\eta$, for which $c(\eta, x_n, \sigma)$ is negative.

However the ROC curves in Fig. 5.3 suggest that the influence of these regions is sufficiently small, if the alarm volume is chosen to be $[-\infty, \delta]$. Hence we can expect that the influence of the regions, where $c(\eta, x_n, \sigma)$ is negative, is suppressed since we average over many different values of x_n and the condition is positive as $x_n \rightarrow -\infty$. (Positive is meant here in the sense, that $c(\eta, x_n, \sigma)$ approaches the value zero for $x_n = -\infty$ from small positive numbers.)

In the third case, for $x_n = -\sigma\eta$ and hence $z = 0$ we find that $c(\eta, x_n, \sigma)$ is positive if $\eta > 2\sqrt{\frac{2}{\pi}}(1 - \frac{1}{2}\text{erfc}(\eta/2))$.

In total we can expect larger increments in Gaussian random numbers to be easier to predict the larger they are. The ROCs in Fig. 5.3 support these results.

5.3 Exponentially distributed random variables

In this section we investigate the prediction of increments in random numbers, which follow a one-sided exponential distribution

$$\rho(x) = \lambda \exp(-\lambda x), \quad x > 0, \quad (5.12)$$

with mean $\mu = 1/\lambda$ and variance $\sigma = 1/\lambda^2$. We can obtain the corresponding joint and conditional PDFs of event and precursor by filtering according to App. B. The joint PDF of precursor x_n and event reads

$$p[x_n, \chi_n(\eta) = 1] = -\lambda \exp(-2\lambda x_n - \lambda^2 \eta). \quad (5.13)$$

The likelihood and the total probability to find events are given by

$$p[\chi_n(\eta) = 1 | x_n] = \exp(-\lambda x_n - \lambda^2 \eta), \quad (5.14)$$

$$\text{and } p[\chi_n(\eta) = 1] = \frac{1}{2} \exp(-\lambda^2 \eta). \quad (5.15)$$

Using Bayes' theorem, we can compute the posterior probabilities to find and not to find events

$$\rho[x_n | \chi_n(\eta) = 1] = 2\lambda \exp(2\lambda x_n), \quad (5.16)$$

$$\rho[x_n | \chi_n(\eta) = 0] = \lambda \exp(-\lambda x_n) \frac{(1 - \exp(-\lambda x_n - \lambda^2 \eta))}{(1 - \frac{1}{2} \exp(-\lambda^2 \eta))}. \quad (5.17)$$

The condition determining the dependence on the event magnitude as specified in Sec. 4.2 is thus given by

$$c(x_n, \eta, \lambda) = \lambda^2 \exp(-\lambda^2 \eta) \left(-\exp(-\lambda x_n) + \frac{1 - \exp(-\lambda x_n - \lambda^2 \eta)}{(1 - \frac{1}{2} \exp(-\lambda^2 \eta))} \right) \quad (5.18)$$

This expression is compared with its numerically evaluated approximations in Fig. 5.4. The test condition for the magnitude dependence is smaller than zero for very small values of x_n and $\eta \approx 0$. As η increases, the value of the condition approaches zero.¹ Hence we should

¹Small values of x_n are of special importance for the prediction of increments since our precursor u is the smallest value of the data set.

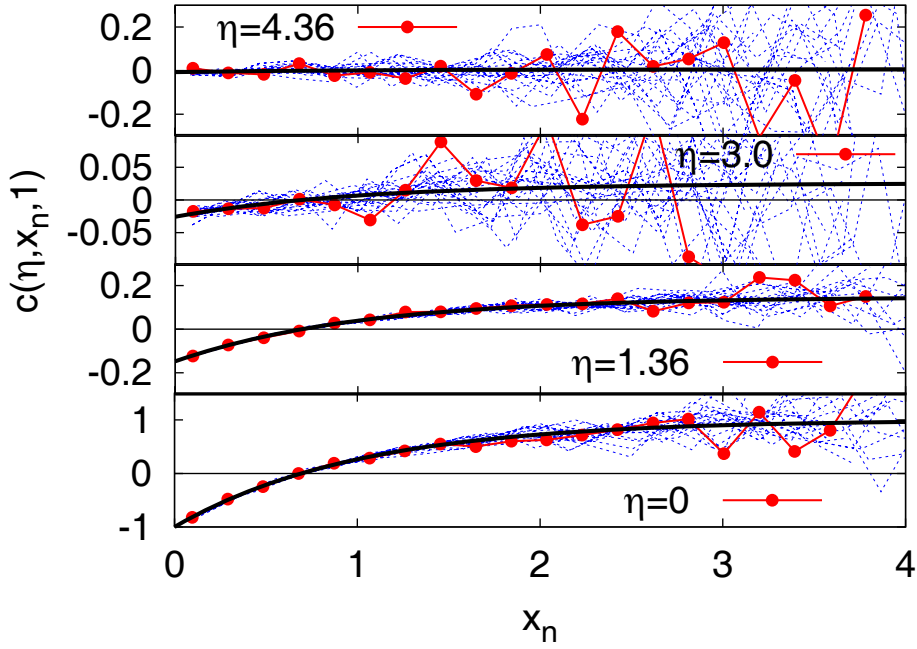


Figure 5.4: The numerically and analytically evaluated condition for the exponentially distributed random variables with $\lambda = 1$. The black line is the result of the analytical evaluation according to Eq. (5.18), the curves plotted with lines and symbols represent the numerical results obtained from the original data set, and the dashed lines represent the results obtained from the corresponding bootstrap samples. Note that for small values of x_n the condition $c(\eta, x_n, \lambda)$ is for all values of η close to zero.

expect a better predictability of smaller events in the region $\eta \approx 0$, and no dependence on the event size as η increases. The corresponding ROC-curves in Fig. 5.5 support these findings. The ROC-curves obtained for events with $\eta \neq 0$ indicate that events are harder to predict than just the sign of the change, characterized by $\eta = 0$. However within the bundle of curves for $\eta \neq 0$ we find almost no dependence on the event magnitude, which corresponds to the fact, that $c(\eta, x_n)$ approaches zero for small values of x_n and larger values of η .

5.4 Symmetrized exponentially distributed random variables

We now consider a sequence of random variables having a symmetrized exponential distribution. The PDF of the symmetrized exponential reads

$$\rho(x) = \frac{\lambda}{2} \exp(-\lambda|x_n|) = \begin{cases} \frac{\lambda}{2} \exp(-\lambda x_n) & : x_n > 0, \\ \lambda/2 & : x_n = 0, \\ \frac{\lambda}{2} \exp(\lambda x_n) & : x_n < 0. \end{cases}$$

The factor $\frac{\lambda}{2}$ normalizes the distribution $\rho(x)$ which has mean $\mu = 0$, and standard deviation $\sigma = \sqrt{2}/\lambda$. Since both, the variables x_n and $x_{n+1} \geq x_n + \sigma\eta$ have these marginal probabilities we have to distinguish three different regimes in which we can find increments, namely

$$I : (x_n > 0) \Rightarrow (x_{n+1} > 0), \quad (5.19)$$

$$II : (-\sigma\eta \leq x_n < 0) \Rightarrow (x_{n+1} \geq 0) \quad (5.20)$$

$$III : (x_n < -\sigma\eta) \Rightarrow (x_{n+1} \geq 0) \vee (x_{n+1} < 0). \quad (5.21)$$

Hence we have to calculate all PDFs separately for the three different regimes. We apply the filtering mechanism according to App. B. If it simplifies the calculation we use the absolute

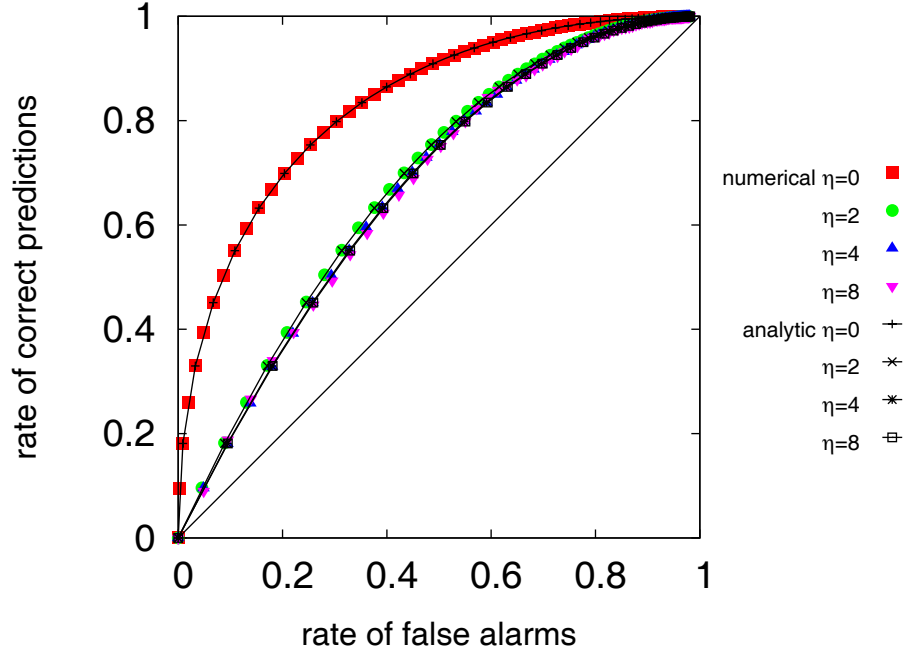


Figure 5.5: The ROCs for exponentially distributed i.i.d. random numbers show no significant dependence on the event magnitude. The ROC curves were made via predicting increments in 10^7 normal i.i.d. random numbers and the predictions were made according to the prediction strategy II as described in Sec. 2.2.3. The black line indicates the analytically evaluated ROC curve for $\eta = 0$.

magnitude of the event d and just switch to relative magnitude of events $\eta = d/\sigma$ at the end. Hence we find the joint PDFs of precursory variable and event to be given by the following expressions

$$p[\mathbf{s}_n, \chi_n(\eta) = 1] = \begin{cases} \frac{\lambda^2}{4} \exp(-\lambda x_n) \int_0^\infty d\gamma \exp(-\lambda(x_n + d + \gamma)), & : x_n > 0, \\ \frac{\lambda^2}{4} \exp(\lambda x_n) \int_0^\infty d\gamma \exp(-\lambda(x_n + d + \gamma)) & : -d \leq x_n < 0, \\ \frac{\lambda^2}{4} \exp(\lambda x_n) \left(\int_0^{-(x_n+d)} d\gamma \exp(\lambda(x_n + d + \gamma)) \right. & \\ \quad \left. + \int_{-(x_n+d)}^\infty d\gamma \exp(-\lambda(x_n + d + \gamma)) \right) & : x_n < -d < 0, \end{cases} \quad (5.22)$$

$$= \begin{cases} \frac{\lambda}{4} \exp(-\lambda(2x_n + d)), & : x_n > 0, \\ \frac{\lambda}{4} \exp(-\lambda d) & : -d \leq x_n < 0, \\ -\frac{\lambda}{4} \exp(\lambda(2x_n + d)) + \frac{\lambda}{2} \exp(\lambda x_n) & : x_n < -d < 0, \end{cases} \quad (5.23)$$

$$= \begin{cases} \frac{\lambda}{4} \exp(-2\lambda x_n - \sqrt{2}\eta), & : x_n > 0, \\ \frac{\lambda}{4} \exp(-\sqrt{2}\eta) & : -\frac{\sqrt{2}}{\lambda}\eta < x_n < 0, \\ -\frac{\lambda}{4} \exp(2\lambda x_n + \sqrt{2}\eta) + \frac{\lambda}{2} \exp(\lambda x_n) & : x_n < -\frac{\sqrt{2}}{\lambda}\eta < 0, \end{cases} \quad (5.24)$$

with $\eta = \lambda d/\sqrt{2}$.

The total probability to find events can be obtained by marginalizing

$$p[\chi_n(\eta) = 1] = \begin{cases} \frac{\lambda}{4} \int_0^\infty dx_n \exp(-2\lambda x_n - \lambda d), & : x_n > 0, \\ \frac{\lambda}{4} \int_{-d}^0 dx_n \exp(-\lambda d) & : -d \leq x_n < 0, \\ -\frac{\lambda}{4} \int_{-\infty}^{-d} dx_n \exp(\lambda(2x_n + d)) \\ \quad + \frac{\lambda}{2} \int_{-\infty}^{-d} dx_n \exp(\lambda x_n) & : x_n < -d < 0. \end{cases} \quad (5.25)$$

$$= \begin{cases} \frac{1}{8} \exp(-\lambda d), & : x_n > 0, \\ \frac{\lambda d}{4} \exp(-\lambda d) & : -d < x_n < 0, \\ -\frac{3}{8} \exp(-\lambda d) & : x_n < -d < 0, \end{cases} \quad (5.26)$$

$$= \begin{cases} \frac{1}{8} \exp(-\sqrt{2}\eta), & : x_n > 0, \\ \frac{\lambda d}{4} \exp(-\sqrt{2}\eta) & : -\frac{\sqrt{2}}{\lambda}\eta \leq x_n < 0, \\ -\frac{3}{8} \exp(-\sqrt{2}\eta) & : x_n < -\frac{\sqrt{2}}{\lambda}\eta < 0. \end{cases} \quad (5.27)$$

If we are not interested in the range of the precursory variable, the total probability to find events is given by the sum of the probabilities for the different regimes, namely

$$p[\chi_n(\eta) = 1|x_n] = \frac{1}{2} \exp(-\sqrt{2}\eta) \left(1 + \frac{\eta}{\sqrt{2}}\right). \quad (5.28)$$

Using Bayes' theorem, we find the posterior probabilities,

$$\rho[x_n|\chi_n(\eta) = 1, \lambda] = \begin{cases} \frac{\lambda}{(2+\sqrt{2}\eta)} \exp(-2\lambda x_n) & : x_n > 0, \\ \frac{\lambda}{(2+\sqrt{2}\eta)} & : -\sigma\eta < x_n < 0, \\ \frac{\lambda}{(2+\sqrt{2}\eta)} (2 \exp(\sqrt{2}\eta + \lambda x_n) - \exp(2\sqrt{2}\eta + 2\lambda x_n)) & : x_n < -\sigma\eta < 0, \end{cases} \quad (5.29)$$

$$\rho[x_n|\chi_n(\eta) = 0, \lambda] = \begin{cases} \frac{\lambda}{2} \exp(-\lambda x_n) \frac{(1-\frac{1}{2} \exp(-\lambda x_n - \sqrt{2}\eta))}{(1-\frac{1}{2}(1+\frac{\eta}{2}) \exp(-\sqrt{2}\eta))} & : x_n > 0, \\ \frac{\lambda}{2} \exp(\lambda x_n) \frac{(1-\frac{1}{2} \exp(-\lambda x_n - \sqrt{2}\eta))}{(1-\frac{1}{2}(1+\frac{\eta}{2}) \exp(-\sqrt{2}\eta))} & : -\sigma\eta < x_n < 0, \\ \frac{\lambda}{4} \frac{\exp(2\lambda x_n + \sqrt{2}\eta)}{(1-\frac{1}{2}(1+\frac{\eta}{2}) \exp(-\sqrt{2}\eta))} & : x_n < -\sigma\eta < 0, \end{cases} \quad (5.30)$$

and the likelihood

$$p[\chi_n(\eta) = 1|x_n, \lambda] = \begin{cases} \frac{1}{2} \exp(-\sqrt{2}\eta - \lambda x_n) & : x_n > 0, \\ \frac{1}{2} \exp(-\sqrt{2}\eta - \lambda x_n) & : -\sigma\eta < x_n < 0, \\ 1 - \frac{1}{2} \exp(\sqrt{2}\eta + \lambda x_n) & : x_n < -\sigma\eta < 0. \end{cases} \quad (5.31)$$

Hence the condition $c(\eta, x_n, \lambda)$ reads

$$c(\eta, x_n, \lambda) = \begin{cases} -\sqrt{2} \left(1 - \frac{(1-\frac{1}{2} \exp(-\sqrt{2}\eta - \lambda x_n))}{(1-\frac{1}{8} \exp(-\sqrt{2}\eta))}\right), & x_n > 0, \quad (I) \\ -\sqrt{2} + \frac{(1-\sqrt{2}\eta)}{\eta} \frac{(1-\frac{1}{2} \exp(-\sqrt{2}\eta - \lambda x_n))}{(1-\frac{\sqrt{2}}{4}\eta \exp(-\sqrt{2}\eta))}, & -\sigma\eta < x_n < 0, \quad (II) \\ -\frac{1}{\sqrt{2}} \exp(\lambda x_n + \sqrt{2}\eta) \left(\frac{1}{1-\frac{1}{2} \exp(\sqrt{2}\eta + \lambda x_n)} + \frac{1}{1-\frac{3}{8} \exp(-\sqrt{2}\eta)}\right), & x_n < -\sigma\eta \quad (III). \end{cases} \quad (5.32)$$

In the limit $\eta \rightarrow \infty$ the expressions in Eq. (5.32) tend either to zero (regime (I) and (III) or to $-\sqrt{2} + 1$ (regime(II)). Figure 5.6 compares the results of the numerical evaluation of the condition and the analytical expression given by Eq. (5.32). Since most precursors of large increments can be found among negative values, the numerical evaluation of $c(\eta, x_n, \lambda)$ becomes worse for positive values of x_n , since in this limit the likelihood is not very well sampled from the

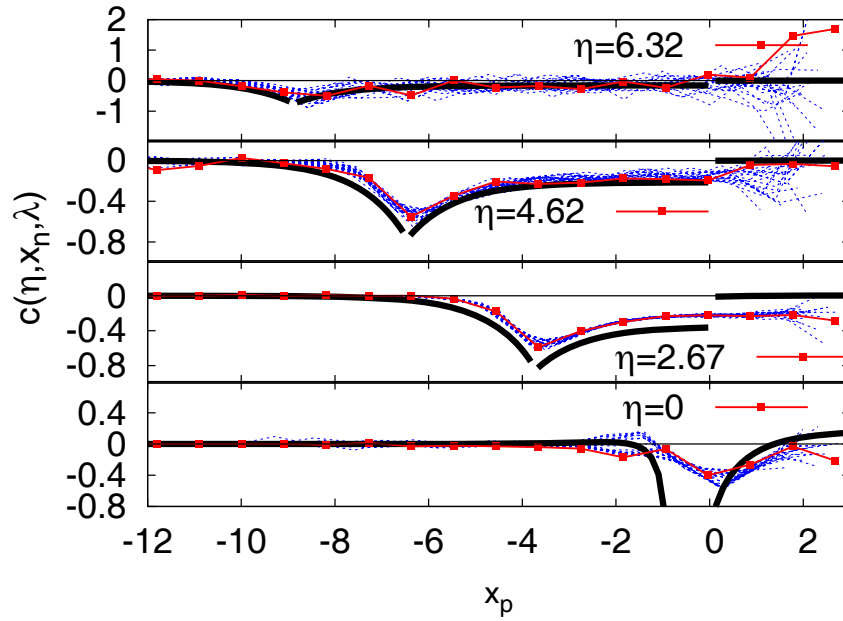


Figure 5.6: The numerically and analytically evaluated condition for the symmetrized exponential. The black line is the result of the analytical evaluation according to Eq. (5.32), the curves plotted with lines and symbols represent the numerical results obtained from the original data set, and the dashed lines represent the results obtained from the corresponding bootstrap samples. Note that for small values of x_n the condition $c(\eta, x_n, \lambda)$ close to zero is for all values of η .

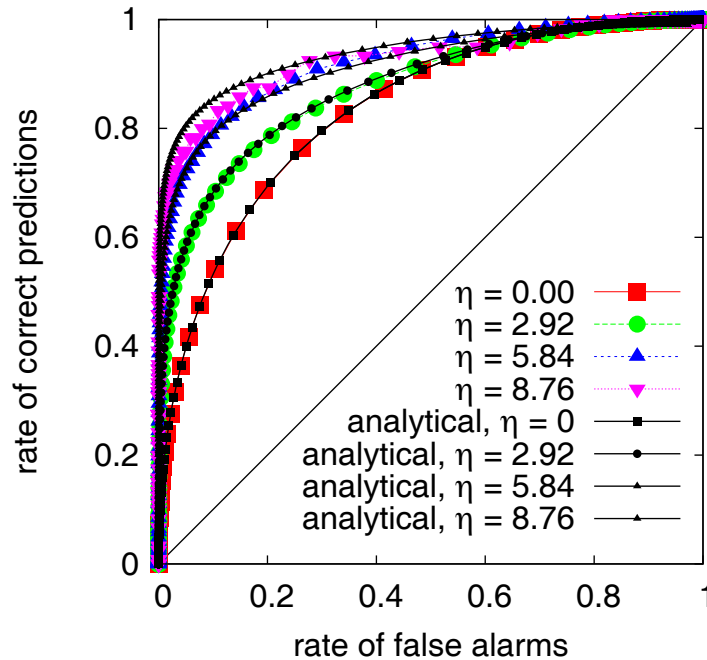


Figure 5.7: The ROCs for symmetrically exponentially distributed i.i.d. random numbers show no significant dependence on the event magnitude. The ROC curves were made via predicting increments in 10^7 normal i.i.d. random numbers and the predictions were made according to the prediction strategy II as described in Sec. 2.2.3. The black line indicates the analytically evaluated ROC-curves.

data. This leads also to the wide spread of the bootstrap samples in this region. Furthermore, Figure 5.6 shows that in the vicinity of the smallest value of the data set, the condition $c(\eta, x_n, \lambda)$ is zero. As we approach larger values of η , $c(\eta, x_n, \lambda)$ approaches zero in the whole range of data values. That is why we would expect to see no influence of the event magnitude on the quality of predictions in the exponential case.

However, the ROC-curves in Fig. 5.7 show a different behavior. Fig. 5.7 shows the numerical evaluated ROC curves made via predicting increments in 10^7 normal i.i.d. random numbers according to strategy II, as described in chapter 2. Additionally the analytical ROCs are calculated by evaluating the integrals over the posterior PDFs for $\eta = 0$. Since the posterior PDF is a non-continuous function we evaluated the integrals in the different regimes of x_n separately (starting from the smallest observed value of x_n) and then added the values.

In the vicinity of the origin the ROC-curves for different η coincide, which corresponds to $c(\eta, x_n) \approx 0$. However, the curves separate with increasing size of the tolerance volume and reveal a positive magnitude dependence. In total the positive magnitude dependence is less pronounced than the one observed for the Gaussian ROC-curves shown in the Sec. 5.2. Although the range of η -values investigated in the exponentially distributed ROC-curves has twice the size than the range of eta values studied in the Gaussian ROC (Sec. 5.2), the exponential ROC-curves are closer to each other, than the corresponding Gaussian ROC-curves.

In total the ROC-curves for the exponentially distributed random numbers provide an examples for the limits of our test condition. Although we are very well able to predict the behavior in the vicinity of the origin, the condition is unable to capture the behavior of the parts of the ROC-curves which corresponds to a larger tolerance volume. It is especially astonishing, that the condition has no positive values for $x_n < 0$, which is in contrast to the ROCs showing a positive magnitude dependence.

5.5 Power-Law distributed random variables

In this section we investigate a sequence of random variables which follow a one-sided power-law distribution

$$\rho(x) = \alpha x_{min}^\alpha x^{-(\alpha+1)} \quad (5.33)$$

for $x \in [x_{min}, \infty)$ the exponent $\alpha \geq 3$, lower endpoint $x_{min} > 0$, and variance $\sigma = \frac{x_{min}}{\alpha-1} \sqrt{\frac{\alpha}{\alpha-2}}$. This distribution has a finite variance for $\alpha \geq 3$. Applying the filter according to App. B, the joint PDF is in the power-law case given by

$$p[x_n, \chi_n(d) = 1, \alpha, x_{min}] = \frac{\alpha x_{min}^{2\alpha}}{x_n^{\alpha+1}} \int_0^\infty d\gamma (x_n + d + \gamma)^{-\alpha-1} \quad (5.34)$$

$$= \frac{\alpha x_{min}^{2\alpha}}{x_n^{\alpha+1}} \frac{1}{(x_n + d)^\alpha} \quad (5.35)$$

$$= p[x_n, \chi_n(\eta) = 1, \alpha, x_{min}] = \frac{\alpha x_{min}^{2\alpha}}{x_n^{\alpha+1}} \frac{1}{\left(x_n + \left(\frac{x_{min}}{\alpha-1}\right) \sqrt{\frac{\alpha}{\alpha-2}} \eta\right)^\alpha}. \quad (5.36)$$

with $\eta = d \left(\frac{\alpha-1}{x_{min}}\right) \sqrt{\frac{\alpha-2}{\alpha}}$. Marginalizing the total probability to find events reads

$$p[\chi_n(d) = 1, \alpha, x_{min}] = \alpha x_{min}^{2\alpha} \int_{x_{min}}^\infty dx_n \frac{1}{x_n^{\alpha+1} (x_n + d)^\alpha}. \quad (5.37)$$

This integral can be solved using the following identity from [95]:

$$\begin{aligned} \int_{x_0}^\infty dx \frac{x^{\mu-1}}{(1 + \beta x)^\nu} &= \frac{u^{\mu-\nu}}{\beta^\nu (\nu - \mu)} {}_2F_1\left(\nu, \nu - \mu; \nu - \mu + 1, -\frac{1}{\beta x_0}\right), \quad [\text{Re } \nu > \text{Re } \mu], \quad (5.38) \\ &= \frac{1}{\beta^\nu} \int_{x_0}^\infty dx \frac{x^{\mu-1}}{\left(\frac{1}{\beta} + x\right)^\nu}. \end{aligned}$$

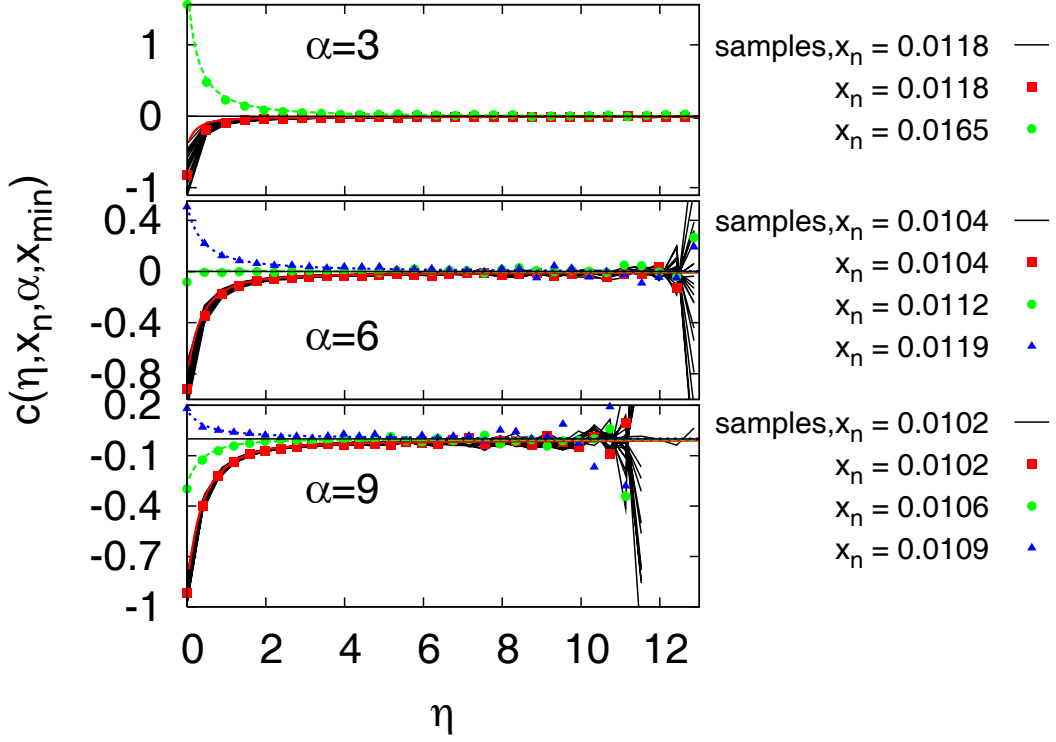


Figure 5.8: The condition $c(\eta, x_n, x_{min}, \alpha)$ for the power-law distribution with lower endpoint $x_{min} = 0.01$ is plotted for constant values of the precursory variable x_n . The symbols represent the results of the numerical evaluation of $c(\eta, x_n, x_{min}, \alpha)$, the colored lines denote the analytic results, and the bundle of black lines denotes the result for the corresponding bootstrap samples evaluated for the optimal precursor. For the “ideal” precursor $x_n = x_{min} = 0.01$ all values of $c(\eta, \alpha, 0.01)$ are negative. Hence one should expect smaller events to be better predictable. One can also see that this effect is sensitive of the choice of the precursor.

Hence setting $x_0 = x_{min}$, $1/\beta = d$, $\mu = -\alpha$ and $\nu = \alpha$ the total probability to find events reads

$$p[\chi_n(d) = 1, \alpha, x_{min}] = \frac{1}{2} {}_2F_1 \left(\alpha, 2\alpha, 2\alpha + 1, -\frac{d}{x_{min}} \right), \quad (5.39)$$

$$= p[\chi_n(\eta) = 1, \alpha] = \frac{1}{2} {}_2F_1 \left(\alpha, 2\alpha, 2\alpha + 1, -\frac{\eta}{(\alpha + 1)} \sqrt{\frac{\alpha}{\alpha - 2}} \right). \quad (5.40)$$

Using Bayes’ theorem we find the following conditional PDFs:

$$\rho[x_n | \chi_n(\eta) = 1, \alpha, x_{min}] = \frac{\alpha x_{min}^{2\alpha}}{x_n^{\alpha+1} \left(x_n + \frac{x_{min}}{\alpha+1} \sqrt{\frac{\alpha}{\alpha-2}} \eta \right)^\alpha P(\eta, \alpha)} \quad (5.41)$$

$$\rho[x_n | \chi_n(\eta) = 0, \alpha, x_{min}] = \frac{\alpha x_{min}^\alpha \left(1 - \left(\frac{x_{min}}{x_n + \frac{x_{min}}{\alpha+1} \sqrt{\frac{\alpha}{\alpha-2}} \eta} \right)^\alpha \right)}{x_n^{\alpha+1} (1 - P(\eta, \alpha))} \quad (5.42)$$

$$p[\chi_n(\eta) = 1 | x_n, \alpha, x_{min}] = \left(\frac{x_{min}}{x_n + \frac{x_{min}}{\alpha+1} \sqrt{\frac{\alpha}{\alpha-2}} \eta} \right)^\alpha. \quad (5.43)$$

Within the range (x_{min}, ∞) the likelihood $p[\chi_n(\eta) = 1 | x_n, \alpha, x_{min}]$ has no well defined maximum. However, since the likelihood is a monotonously decreasing function, we use the lower endpoint x_{min} as a precursor. In order to evaluate the criterion we need to calculate the derivative of

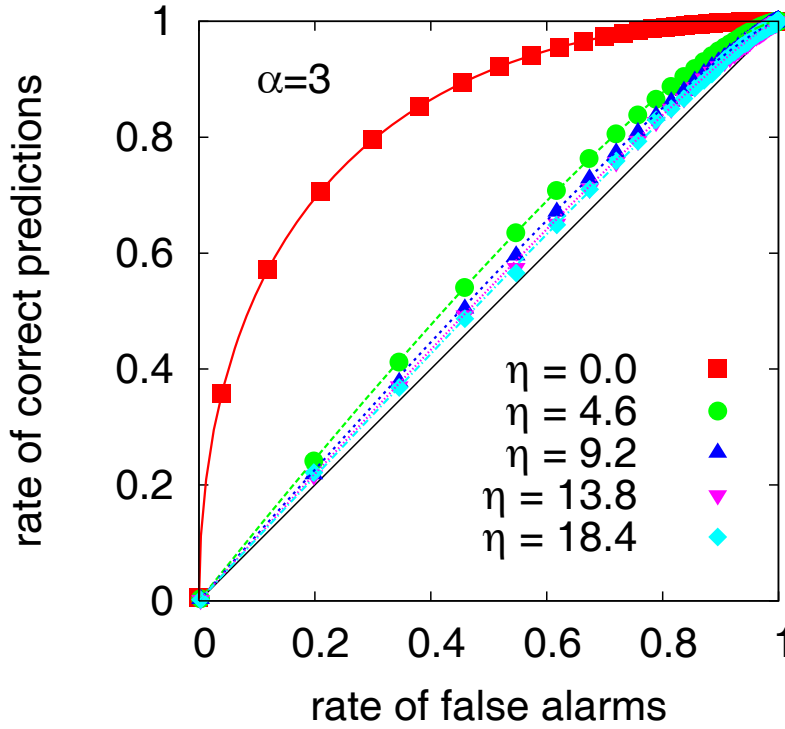


Figure 5.9: ROC-plot for the power-law distribution with $\alpha = 3$ and $x_{min} = 0.01$. The symbols show the numerical results and the lines indicate the analytically calculated ROC curves. The ROC curves were made via predicting increments in 10^7 Pareto distributed i.i.d. random numbers. The predictions were made according to strategy II described in Sec. 2.2.3. Note that we tested only event magnitudes η , for which we found at least 1000 events, so that the effects we observe are not due to a lack of statistics. The ROC curves display that in Pareto distributed i.i.d. random numbers with the lower endpoint $x_{min} = 0.01$ large events are very hard to predict.

the hypergeometric function in $p[\chi_n(\eta) = 1, \alpha, x_{min}]$. The first derivative of a hypergeometric function is according to [89] given by

$$\frac{\partial}{\partial z} {}_2F_1(a, b, c, z) = \frac{ab}{c} {}_2F_1(a+1, b+1, c+1, z) \quad (5.44)$$

In [95] we find the Gauss' recursion relation

$$az {}_2F_1(a+1, b+1, c+1, z) = c {}_2F_1(a, b+1, c, z) - c {}_2F_1(a, b, c, z). \quad (5.45)$$

Hence we can express the derivative of the hypergeometric function as

$$\frac{\partial}{\partial z} {}_2F_1(a, b, c, z) = \frac{b}{z} ({}_2F_1(a, b+1, c, z) - {}_2F_1(a, b, c, z)). \quad (5.46)$$

In our special case, $a = \alpha$, $b = 2\alpha$, $c = 2\alpha + 1$, the first term on the right hand side of Eq.(5.46) is given by ${}_2F_1(\alpha, 2\alpha + 1, 2\alpha + 1, z)$ which can be simplified using

$${}_2F_1(n, b, b, z) = (1-z)^{-n}, \quad (5.47)$$

which can also be found in [95]. Hence the derivative of ${}_2F_1(\alpha, 2\alpha, 2\alpha + 1, z)$ is given by

$$\frac{\partial}{\partial z} {}_2F_1(\alpha, 2\alpha, 2\alpha + 1, f(z)) = \frac{2\alpha}{f(z)} \left(\frac{1}{(1-f(z))^\alpha} - {}_2F_1(\alpha, 2\alpha, 2\alpha + 1, f(z)) \right) \frac{\partial}{\partial z} f(z). \quad (5.48)$$

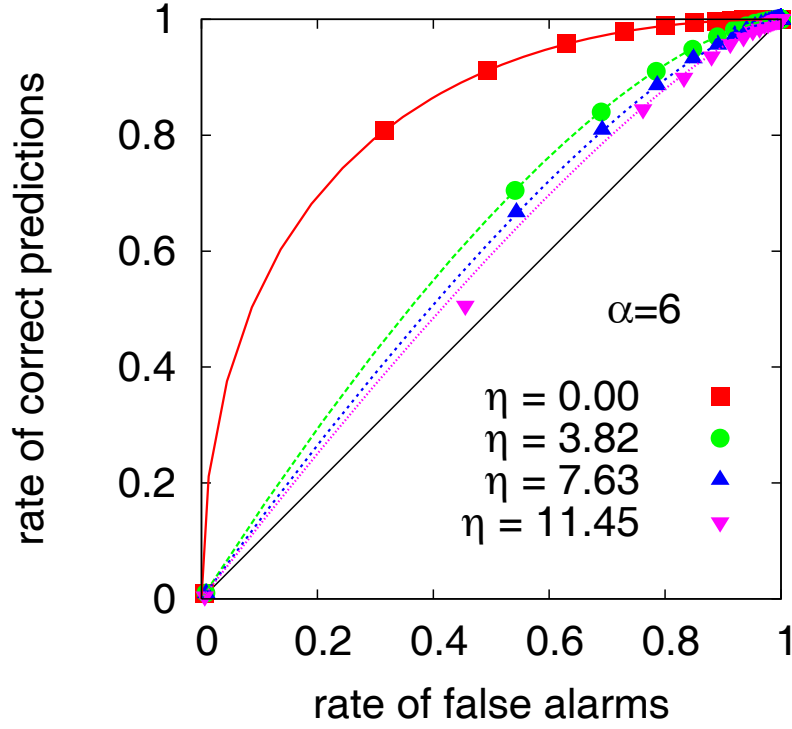


Figure 5.10: (Color online) ROC-plot for the power-law distribution with $\alpha = 6$ and $x_{min} = 0.01$. The symbols show the numerical results, the lines indicate the analytically calculated ROC-curves. The ROC-curves were made via predicting increments in 10^7 Pareto distributed i.i.d. random numbers. The predictions were made according to the strategy II as described in Sec. 2.2.3. The ROC-curves show that in Pareto distributed i.i.d. random numbers with the lower endpoint $x_{min} = 0.01$ large events are especially hard to predict.

Thus the derivative of the total probability reads

$$\frac{\partial P(\eta, \alpha)}{\partial \eta} = \frac{\alpha}{\eta} \left(\frac{1}{\left(1 + \frac{\eta}{\alpha+1} \sqrt{\frac{\alpha}{\alpha-2}}\right)^\alpha} - 2P(\eta, \alpha) \right). \quad (5.49)$$

Using this derivative and inserting the expressions (5.43) and (5.41) for the components of $c(\eta, x_n, x_{min}, \alpha)$ we obtain an explicit analytic expression for the condition, namely

$$c(\eta, x_n, \alpha, x_{min}) = \frac{-\alpha}{\frac{\alpha-1}{x_{min}} \sqrt{\frac{\alpha-2}{\alpha}} x_n + \eta} + \frac{2\alpha (1 - p[\chi_n(\eta) = 1])}{\eta (1 - p[\chi_n(\eta) = 1])} - \frac{2\alpha (1 - p[\chi_n(\eta) = 1])}{\eta (1 - p[\chi_n(\eta) = 1])} \frac{\left(1 + \frac{1}{\alpha-1} \sqrt{\frac{\alpha}{\alpha-2}} \eta\right)^{-\alpha}}{{}_2F_1(\alpha, 2\alpha, 2\alpha + 1, \frac{1}{\alpha-1} \sqrt{\frac{\alpha}{\alpha-2}} \eta)}. \quad (5.50)$$

In Fig. 5.8 we evaluate this expression using *Mathematica* and compare it with the results of an empirical evaluation on the data set of 10^7 i.i.d. random numbers.

Figure (5.8) displays that the value of $c(\eta, x_n, x_{min}, \alpha)$ depends sensitively on the choice of the precursor. For the ideal precursor $u = x_{min}$ all values of $c(\eta, \alpha, x_{min})$ are negative. Hence in this case one should expect smaller events to be better predictable. The corresponding ROC curves in Figs. 5.9, 5.10 and 5.11 verify this statement of $c(\eta, x_n, x_{min}, \alpha)$.

In summary we find that larger events in power-law distributed i.i.d. random numbers are harder to predict the larger they are. This is an admittedly unfortunate result, since extremely

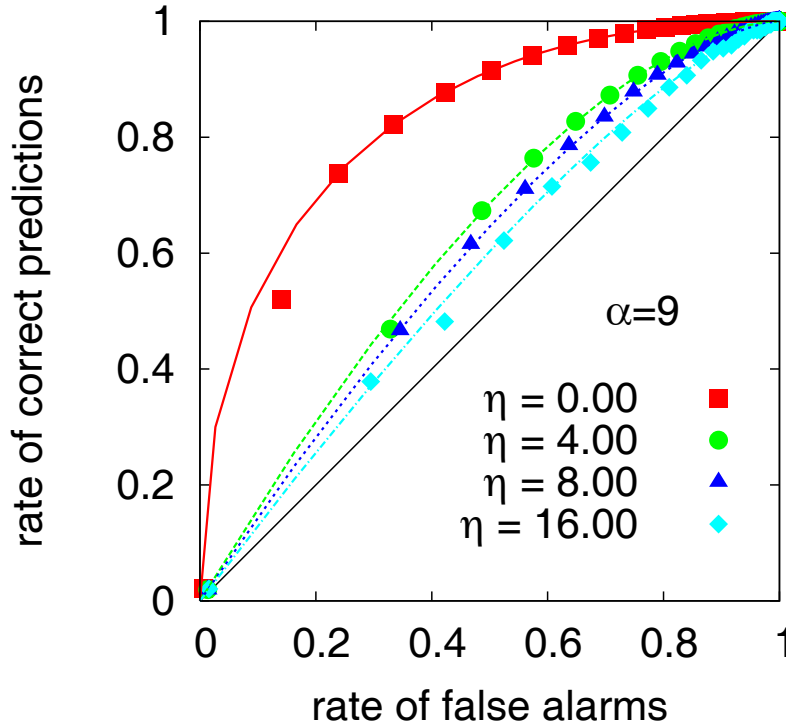


Figure 5.11: ROC-plot for the power-law distribution with $\alpha = 9$ and $x_{min} = 0.01$. The symbols show the numerical results and the lines indicate the analytically calculated ROC curves. The ROC curves were made via predicting increments in 10^7 Pareto distributed i.i.d. random numbers and the predictions were made according strategy II as described in Sec. 2.2.3. The ROC-curves display that in Pareto distributed i.i.d. random numbers smaller events are better predictable and large events are especially hard to predict.

large events occur much more frequently in power-law distributed processes than in Gaussian distributed processes. Hence, their prediction would be highly desirable.

5.6 Symmetrized Power-Law distributed Random Variables

In this section we evaluate the test condition for the magnitude dependence and the corresponding ROC-curves numerically for the prediction of increments in a sequence of symmetrized power-law distributed random variables. The symmetrized power-law distribution is given by

$$\begin{aligned} \rho(x) &= \alpha x_{min}^\alpha x^{-(\alpha+1)}, & x > x_{min} > 0 \\ \rho(x) &= \alpha |x_{max}|^\alpha |x|^{-(\alpha+1)}, & x < x_{max} < 0, \end{aligned} \quad (5.51)$$

with $x_{min} = |x_{max}| = 0.01$, $x_{max} < 0, x_{min} > 0$ and power-law coefficient $\alpha = 3$ were generated by transformation from uniformly distributed random numbers. Since the lower and the upper end-points x_{min} and x_{max} have to be non-zero, there is always an interval $[x_{max}, x_{min}]$ in which our distribution is not defined. Thus trying to evaluate the distributions, the test condition and the ROC-curve analytically does not seem to be promising in this case. However we artificially join both distributions by shifting them to the left (right) by subtracting (adding) x_{min} . The result is a symmetrized power law distribution with mean zero and variance $\sigma = 0.01$.

We can then evaluate the test condition for the magnitude dependence and the corresponding ROC-curves. The results for the test condition, shown in Fig. 5.12, indicate that the condition is approximately zero for most values of η and x_n . This corresponds to the behavior of the ROC-curves in the vicinity of the origin, as shown in Fig. 5.13. Most curves, with the exception of the one for $\eta = 0$, resemble to the vertical axis and are very similar. In the regions of the

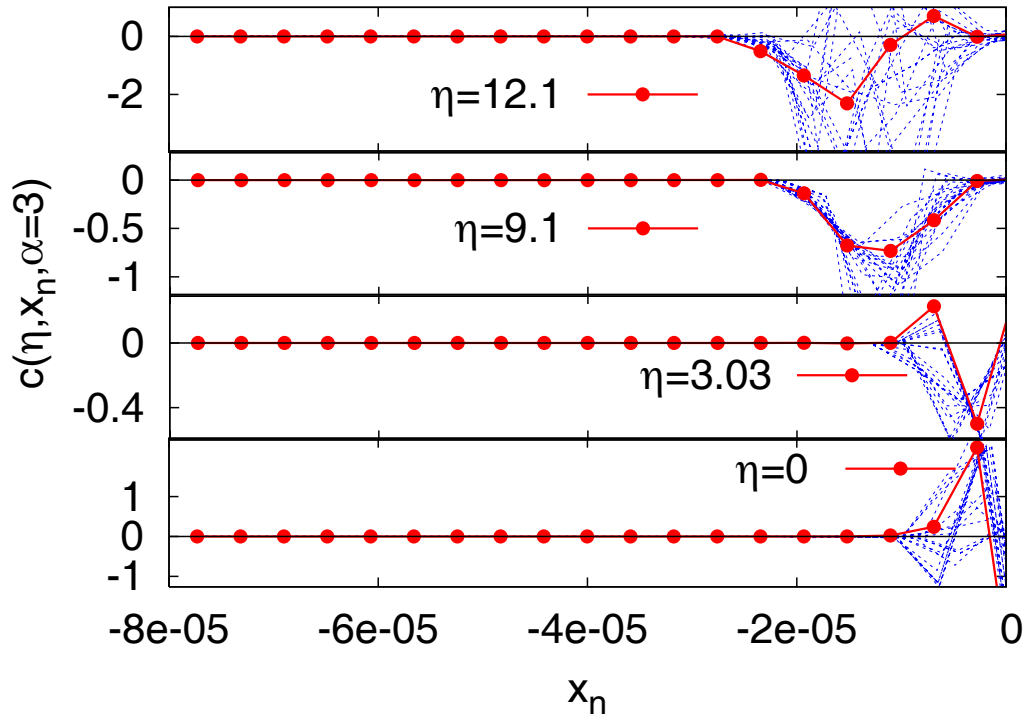


Figure 5.12: The numerically and analytically evaluated condition for the symmetrized power-law distributed random variables with $\alpha = 3$. The curves plotted with lines and symbols represent the numerical results obtained from the original data set, and the dashed lines represent the results obtained from the corresponding bootstrap samples.

ROC, which correspond to an increased tolerance volume we find that large events are harder to predict. This behavior is reflected by the fact, that $c(\eta, x_n)$ is negative for larger η and larger values of x_n .

5.7 Is there a “universal” ROC-Curve for $\eta = 0$?

Comparing the different ROC-curves in this chapter, it is apparent, that the curves for $\eta = 0$ are very similar. They do in fact coincide, as it is shown in Fig. 5.14. Furthermore we observe, that the curves for $\eta = 0$ are symmetric with respect to the diagonal $r_f = 1 - r_c$ and they all pass through the point $(r_c = 0.75, r_f = 0.25)$. These findings do agree with the results of Sornette and Andersen [91] who already found in 2006 that the sign of increments in uncorrelated random numbers can be predicted with a success probability of 75%. Sornette also emphasizes the independence on the underlying distribution function.

In the following we will try to understand these effects in terms of the ROC-curves.

Since the events $x_{n+1} - x_n > 0$ reflect simply an increase of any size, the question whether or not an increment of size $\eta = 0$ occurs, corresponds to the question, whether or not the next value of the time series is larger or smaller than the previous one. In any uncorrelated time series the probability for a larger (smaller) value in the next step should be $1/2$, which the calculations in the previous sections confirm. Using $p[\chi_n(0) = 1] = p[\chi_n(0) = 0] = 0.5$, the rates of correct predictions and the rate of false alarms (as given by Eqs. (2.46) and (2.47)) are given by

$$r_c(u, \delta, 0) = 2 \int_0^\delta d\nu \rho(u + \nu) p[\chi_n(0) = 1 | u + \nu] \quad (5.52)$$

$$r_f(u, \delta, 0) = 2 \int_0^\delta d\nu \rho(u + \nu) (1 - p[\chi_n(0) = 1 | u + \nu]), \quad (5.53)$$

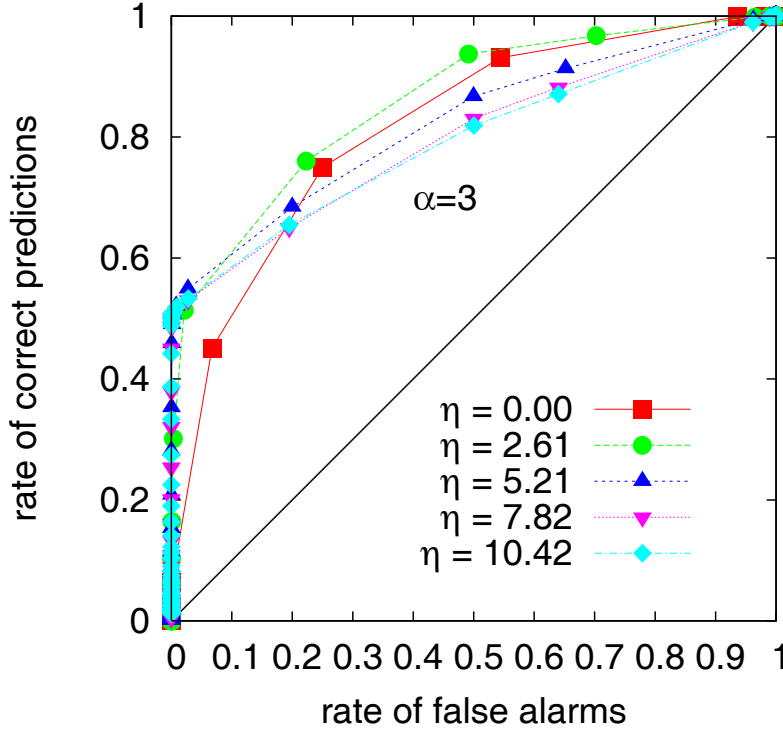


Figure 5.13: The ROCs for symmetrized power-law distributed i.i.d. random numbers show no significant dependence on the event magnitude. The ROC curves were made via predicting increments in 10^7 normal i.i.d. random numbers and the predictions were made according to the prediction strategy II as described in Sec. 2.2.3.

and hence all ROC curves for $\eta = 0$ are given by

$$r_c(u, \delta, 0) = -r_f(u, \delta, 0) + 2 \int_0^\delta d\nu \rho(u + \nu), \quad (5.54)$$

It is easy to see, that the curves given by this expression cross the points $(r_c = 0, r_f = 0)$ in the limit $\delta \rightarrow 0$ and $(r_c = 1.0, r_f = 1.0)$ in the limit $\delta \rightarrow \infty$. However all ROC curves also cross the diagonal from $(0.0, 1.0)$ to $(1.0, 0.0)$. The intersection point, in this case the point $(r_c = 0.75, r_f = 0.25)$, corresponds to the value of δ for which $\int_0^\delta d\nu \rho(u + \nu) = 1/2$ holds, i.e., $(u + \delta)$ represents the median of the distribution $\rho(u + \nu)$.

Demanding, that the ROC-curves for different distributions coincide, is equivalent to demanding that the slope of the ROC-curve is for every point of independent on the choice of the distribution. It is sufficient to discuss the behavior of the slope, since every ROC-curves crosses the points $(0, 0)$ and $(1, 1)$. From Sec. 2.3.4 we know, that the slope of the ROC curve is given by the likelihood ratio, which reads in case of the alarm interval according to strategy II,

$$\Lambda(u, \delta, 0) = \frac{\rho[u + \delta | \chi_n(0) = 1]}{\rho[u + \delta | \chi_n(0) = 0]}. \quad (5.55)$$

$$(5.56)$$

Expressing the posterior PDFs through Bayes' theorem, we see, that for $p[\chi_n(0) = 1] = p[\chi_n(0) = 1] = 1/2$, the likelihood ratio is indeed the ratio of odds as specified in Sec. 2.2.3

$$\Lambda(u, \delta, 0) = \frac{p[\chi_n(0) = 1 | u + \delta]}{p[\chi_n(0) = 0 | u + \delta]} = \frac{p[\chi_n(0) = 1 | u + \delta]}{1 - p[\chi_n(0) = 1 | u + \delta]}. \quad (5.57)$$

Since, $p[\chi_n(0) = 0 | u + \delta] = 1 - p[\chi_n(0) = 1 | u + \delta]$, there exist always a fixed relation between the nominator and the denominator of $\Lambda(u, \delta, 0)$, independently of the details of the likelihood

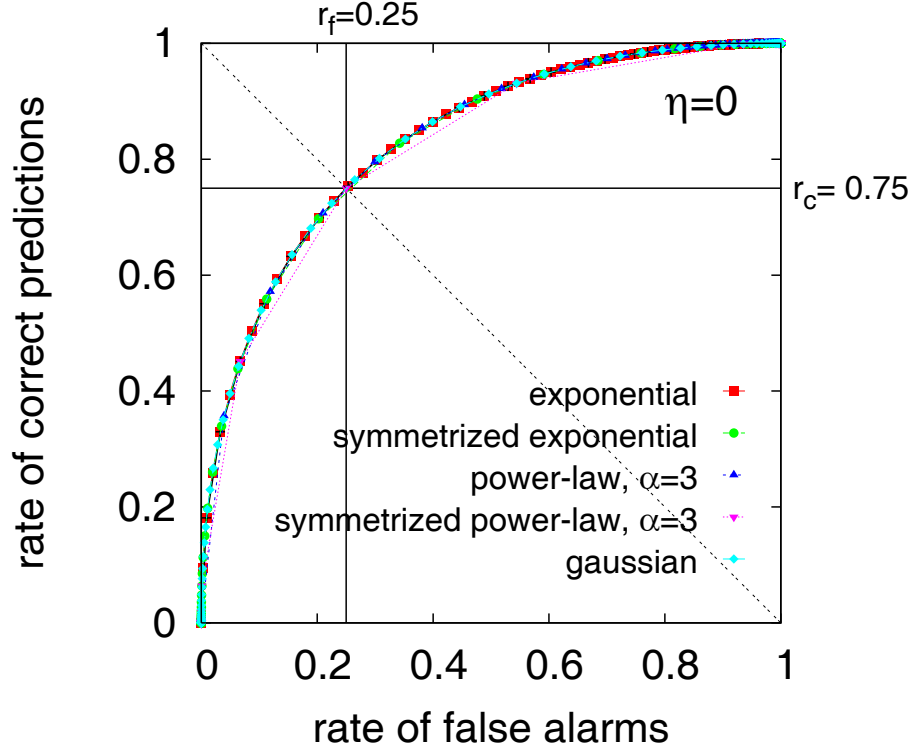


Figure 5.14: The ROCs for different i.i.d. random numbers and $\eta = 0$.

$p[\chi_n(0) = 1|u + \delta]$. In other words, the values of the likelihoods $p[\chi_n(0) = 1|u + \delta]$ at a specific $u + \delta$ are of course different for different marginal PDFs of the processes. However, the fact that we consider the ratio of a probability and its complement insures, that each value $u + \delta$ correspond to a point on the ROC curve in which the slope of the curve is given by Eq. (5.57)

The main difference to predictions for event magnitudes $\eta \neq 0$ is that we find also the term $(1 - p[\chi_n(\eta) = 1])/p[\chi_n(\eta) = 1]$ in the likelihood ratio, i.e., the likelihood ratio is not equal to the ratio of odds

$$\Lambda(u, \delta, \eta) = \frac{(1 - p[\chi_n(\eta) = 1])}{p[\chi_n(\eta) = 1]} \frac{(1 - p[\chi_n(0) = 1|u + \delta])}{p[\chi_n(0) = 1|u + \delta]}. \quad (5.58)$$

5.8 The Influence of the Symmetry of the distribution

Comparing the ROC curves for the symmetric distributions and the one-sided distributions we observe the following features:

- For the one-sided distributions (exponential and power-law) we observe that events with $\eta \neq 0$ are much harder to predict than events with magnitude $\eta = 0$.
- For the one-sided distributions the ROC-curves $\eta \neq 0$ form a bundle of curves, which is well separated from the curve $\eta = 0$. The curves for $\eta \neq 0$ are also very close to each other within this bundle.
- In the symmetric distributions the ROC-curve for $\eta = 0$ is not further apart from the curves $\eta \neq 0$ than the “typical distance” within the curves for $\eta \neq 0$.

5.9 Conclusions

We study the magnitude dependence on the quality of predictions for increments in a time series which consists in sequences of i.i.d. random numbers. Using the first part of the increment x_n

as a precursory variable we predict large increments $x_{n+1} - x_n$ via statistical considerations. In order to measure the quality of the predictions we use ROC curves. Furthermore we evaluate the condition $c(\eta, x_n)$ as introduced in chapter 4 for time series of Gaussian, exponential, symmetrized exponential, power-law and symmetrized power-law i.i.d. random variables. The results obtained from the criterion comply in most cases well with the corresponding ROC-curves. Note that for both, the numerical evaluation of the condition and the ROC-plots, we used only event magnitudes η for which we found at least 1000 events, so that the observed effects are not due to a lack of statistics of the large events.

For the sequence of Gaussian i.i.d. random numbers, we find that large increments are better predictable the larger they are. This feature is reflected by both the ROC-curve and the condition.

For the power-law distributed time series (both, the one-sided and the symmetrized) we observe that events are the harder to predict, the larger they are. However for the sequence of symmetrized power-law distributed random numbers the dependence on the event magnitude becomes only visible for larger alarm volumes. Also in the case of Pareto distributed random numbers, the condition is able to characterize the behavior of the ROC-curve.

Concerning the exponentially distributed random variables we find a qualitatively different behavior for the symmetrized and the one-sided distribution. The ROC-curves for the one-sided distribution display that events with $\eta \neq 0$ are harder to predict, than just the sign of the increment ($\eta = 0$), however the curves for $\eta \neq 0$ almost coincide. Both features are well reflected by the values of $c(\eta, x_n)$ for small x_n . However in the symmetrized exponential case we observe, that ROC-curves for different values of η almost coincide in the vicinity of the origin (i.e., for small alarm volumes), but then later show a positive magnitude dependence for larger alarm volumes. This magnitude dependence is significantly weaker than in the gaussian case, but it nevertheless exists. Our test condition is able to predict the coinciding curves in the vicinity of the origin, but totally misses the better predictability of larger events, which we observe for larger alarm volumes. In fact, the condition rather suggests to find a negative magnitude dependence. Hence the symmetrically exponential distributed random numbers provide us with an example for the limits of our method. Although our test condition is very well able to predict the behavior in the vicinity of the origin, it does in this case not provide information about regions of the ROC which correspond to larger alarm volumes.

While the condition can be easily evaluated analytically, it is not that easy to compute numerically from observed data, since the calculation implies evaluating the derivatives of numerically obtained distributions. Using Savitzky-Golay filters improved the results, but especially in the limit of larger events, where the distributions are difficult to sample, one cannot trust the results of the numerically evaluated criterion. However, for practical applications one could also think about fitting a PDF to the distribution of the underlying process and then evaluate the criterion analytically, starting from the fitted PDF.

Chapter 6

Predicting threshold crossings in AR(1) correlated processes

In this chapter we test the condition $c(\eta, \mathbf{s}_n)$ as given in Eq. (4.19) for threshold crossing in AR(1) processes, which have a Gaussian, power-law and exponential distribution. The most popular example for an extreme event, which consist in a threshold crossing is probably the level of water in a river exceeding the height of a levee. However, one can easily find other examples, in which it would be desirable to predict the exceeding of a threshold. We study the prediction of threshold exceedances in simple short range correlated processes. As already specified in Chap. 2, we define an event, which consist in threshold crossings by a value x_{n+1} of the time series exceeding a given threshold η

$$\chi_n(\eta) = \begin{cases} 1, & x_{n+1} \geq \eta, \\ 0, & x_{n+1} < \eta. \end{cases} \quad (6.1)$$

where the event magnitude η is measured in units of the standard deviation. Due to the correlation of the AR(1) process we use the present value x_n of a time series as a precursory variable for the event happening at time $n + 1$.

6.1 Non-Gaussian AR(1) processes

As already presented in Chap. 3, an autoregressive process of order 1 [AR(1)] (see, e.g., [33]) is given by

$$x_{n+1} = ax_n + \xi_n, \quad (6.2)$$

where ξ_n are uncorrelated random numbers with mean zero. The value and the sign of the coupling strength a determine whether successive values of x_n are clustered or spread. Since we are not in particular interested in the influence of the coupling strength we will set $a = 0.75$ in all following considerations. For $a \neq 0$ the process is exponentially correlated, $\langle x_n x_{n+k} \rangle = a^k$.

Typically the random numbers ξ_n are chosen to be Gaussian distributed. In this case the data generated by the AR(1) model is as well Gaussian distributed. However, due to the summation of random numbers in Eq. (6.2) also non-Gaussian random numbers might lead to a process with an approximately Gaussian distribution. That is why one has to apply other methods in order to obtain an AR(1) correlated process with a non-Gaussian distribution. We create the non-Gaussian distributed AR(1) processes by replacing the data of the Gaussian AR(1) process by random numbers, which follow the desired distribution function. This is done by ordering the data of the Gaussian AR(1) process and the random numbers according to their magnitude and then replacing the n -th largest value of the data set by the n -th largest random number. This procedure leads of course to local fluctuations in the value of the correlations strength a . However, the characteristic behavior of the process is still preserved, as one can see in Fig. 6.1.

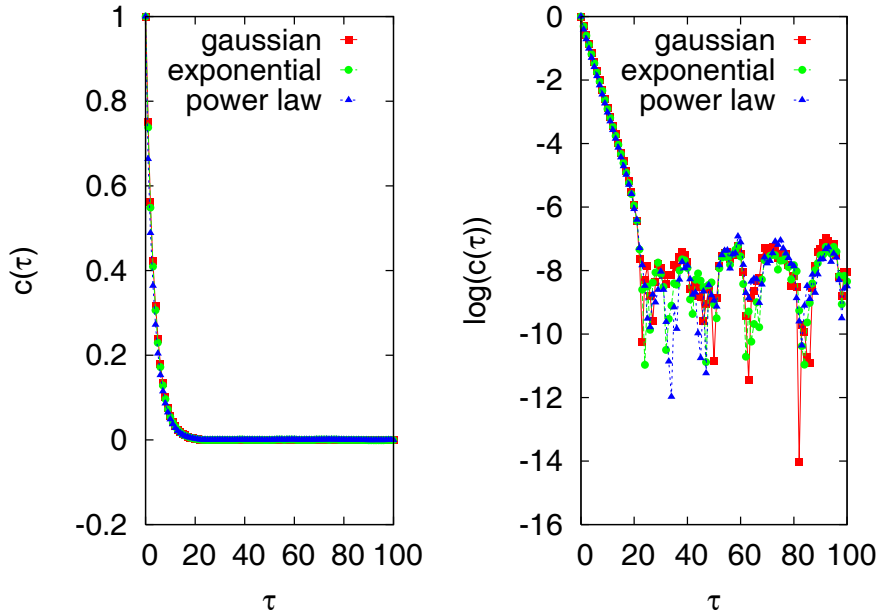


Figure 6.1: The autocorrelation function $c(\tau) = \sum_n (x_n - \mu)(x_{n+\tau} - \mu) / ((n - \tau)\sigma^2)$, with the mean μ and the standard deviation σ is evaluated on the AR(1) correlated data.

For the Gaussian AR(1) process all quantities which enter into the prediction can be evaluated analytically. Since in most cases the structure of the PDF is not known analytically, we evaluate $c(\eta, x_n)$ also numerically. In this case the approximations of the total probability and the likelihood are obtained by "binning and counting" and their numerical derivatives are evaluated via a Savitzky-Golay-filter [93, 94]. The numerical evaluation is done within 10^7 data points. In order to check the stability of this procedure, we evaluate $c(\eta, x_n)$ also on 20 bootstrap samples, which are generated from the original data set. These bootstrap samples consist of 10^7 pairs of event and precursor, which were drawn randomly from the original data set. Thus their PDFs are slightly different in their first and second moment and they contain different numbers of events. Evaluating $c(\eta, x_n)$ on the bootstrap samples thus shows, how sensitive our numerical evaluation procedure is towards changes in the numbers of events. This is especially important for large and therefore rare events.

In order to check the results obtained by the evaluation of $c(\eta, x_n)$, we compute also the corresponding ROC curves analytically and numerically.

Note that for both, the numerical evaluation of the condition and the ROC-plots, we used only event magnitudes η , for which we found at least 100 events, so that the observed effects are not due to a lack of statistics of the large events.

6.2 Results for the Gaussian AR(1) process

We start again from the marginal PDF of the time step x_n of an AR(1) process, which was already used in Chap. 3

$$\rho(x_n, a) = \sqrt{\frac{1-a^2}{2\pi}} \exp\left(-\frac{1-a^2}{2} x_n^2\right). \quad (6.3)$$

Since the magnitude of the events is naturally measured in units of the standard deviation $\sigma(a)$ we introduce a new scaled variable $\eta = \frac{d}{\sigma(a)} = d\sqrt{1-a^2}$. For $a \neq 0$ the process is exponentially correlated $\langle x_n x_{n+k} \rangle = a^k$ and the joint PDF of two successive values $\rho(x_n, x_{n+1})$ is a bivariate Gaussian. From this we derive the joint PDF $p[x_n, \chi_n(\eta) = 1, a]$ by a simple integration using

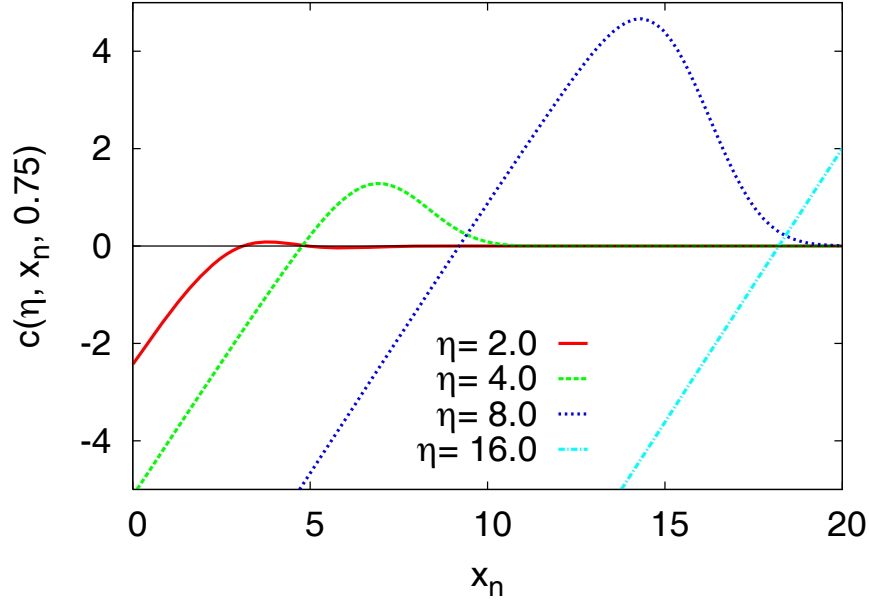


Figure 6.2: The condition $c(\eta, x_n, 0.75)$ for the Gaussian distributed AR(1) process as given by Eq. (6.10).

the Heaviside function Θ as introduced in App. B,

$$p[x_n, \chi_n(\eta) = 1, a] = \int dx_n \Theta(x_n - \eta\sigma) \rho(x_n, x_{n+1}).$$

Since the calculation of the CPDFs and the total probability to find events is very similar to the calculation for the prediction of increments presented in detail in Chap. 3, we will simply list the resulting PDFs for events defined as threshold crossings here. The posterior PDFs to find or not to find events are given by

$$\rho[x_n | \chi_n(\eta) = 1, a] = \frac{\sqrt{1-a^2} \exp\left(-\frac{1-a^2}{2} x_n^2\right)}{2\sqrt{2\pi} \rho^\Theta(a, \eta)} \operatorname{erfc}\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n}{\sqrt{2}}\right), \quad (6.4)$$

$$\rho[x_n | \chi_n(\eta) = 0, a] = \frac{\sqrt{1-a^2} \exp\left(-\frac{1-a^2}{2} x_n^2\right)}{2\sqrt{2\pi} (1 - \rho^\Theta(a, \eta))} \left(1 + \operatorname{erf}\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n}{\sqrt{2}}\right)\right). \quad (6.5)$$

The corresponding likelihood reads

$$p[\chi_n(\eta) = 1 | x_n, a] = \frac{1}{2} \operatorname{erfc}\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n}{\sqrt{2}}\right). \quad (6.6)$$

We recall that the optimal precursor is given by u which maximizes the likelihood and hence $u = \infty$. In a finite data set, this optimal precursor value is $u = \max\{x_n\}$ and the alarm interval becomes $I = [u - \delta, u]$. From the mean value of the posterior PDF $\langle x_n \rangle$ we can obtain the analytic structure of the total PDF to find events

$$p[\chi_n(\eta) = 1, a] = \frac{a}{\sqrt{2(1-a^2)} \langle x_n \rangle} \exp\left(-\frac{\eta^2}{2}\right). \quad (6.7)$$

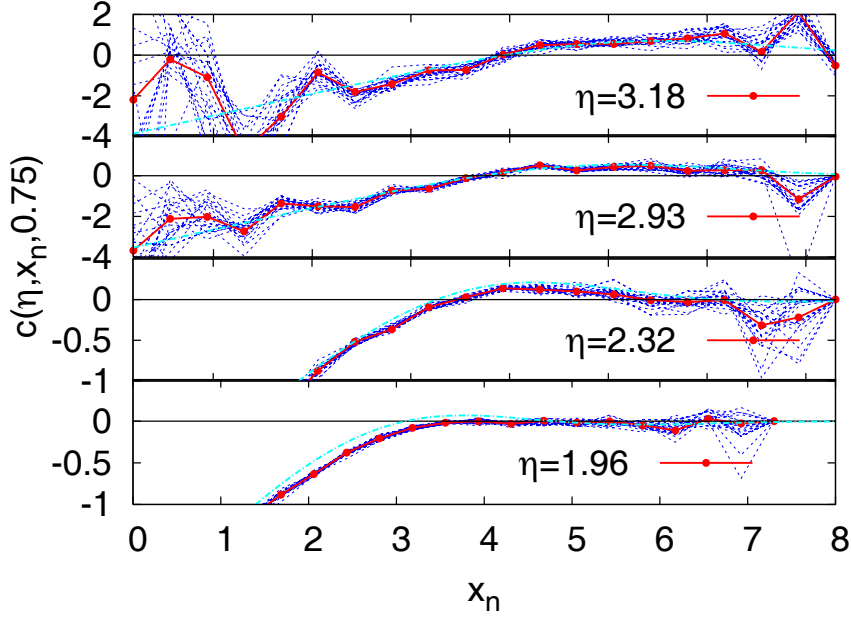


Figure 6.3: The analytical and the numerical results for the evaluation of the condition $c(\eta, x_n, 0.75)$ are compared. The analytical results (dashed light blue line) for the Gaussian distributed AR(1) process are given by Eq. (6.10). The results for the numerical evaluation of $c(\eta, x_n, 0.75)$ are indicated by the symbols connected with lines. The results for the bootstrap samples are plotted in dashed lines.

Using the asymptotic expression of the error function from Eq.(3.19)

$$\operatorname{erfc}(z) \sim \frac{\exp(-z^2)}{\sqrt{\pi}z} \left(1 + \sum_{m=1}^{\infty} (-1)^m \frac{1 \cdot 3 \dots (2m-1)}{(2z^2)^m} \right),$$

$$\left(z \rightarrow \infty, |\arg z| < \frac{3\pi}{4} \right)$$

and approximating the mean value $\langle x_n \rangle$ with the maximum x_n^* of the posterior PDF

$$x_n^* = \sqrt{\frac{2}{\pi}} \frac{a}{1-a^2} \frac{\exp\left(-\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n^*}{\sqrt{2}}\right)^2\right)}{\operatorname{erfc}\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n^*}{\sqrt{2}}\right)}$$

$$\propto \frac{a\eta}{\sqrt{1-a^2} \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)}, \quad \eta \rightarrow \infty, \quad (6.8)$$

we obtain the following approximation of the total probability to find events

$$p[\chi_n(\eta) = 1, a] \propto \frac{\exp\left(-\frac{\eta^2}{2}\right)}{\sqrt{2}\eta} \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right), \quad \eta \rightarrow \infty \quad (6.9)$$

Note that this expression is only valid in the limit of large η . In particular it does not hold for $\eta = 0$. Using Eq. (6.6) and Eq. (6.9) the constraint $c(\eta, x_n, a)$ reads

$$c(\eta, x_n, a) \propto -\sqrt{\frac{2}{\pi(1-a^2)}} \frac{\exp\left(\frac{-1}{2} \left(\frac{\eta}{\sqrt{1-a^2}} - ax_n\right)^2\right)}{\operatorname{erfc}\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n}{\sqrt{2}}\right)}$$

$$+ \left(\eta + \frac{1}{\eta}\right) \frac{\left(1 - \frac{1}{2} \operatorname{erfc}\left(\frac{\eta}{\sqrt{2}\sqrt{1-a^2}} - \frac{ax_n}{\sqrt{2}}\right)\right)}{1 - \frac{\exp(-\eta^2/2)}{\sqrt{2}\eta} \left(1 + \mathcal{O}\left(\frac{1}{\eta^2}\right)\right)} \quad (6.10)$$

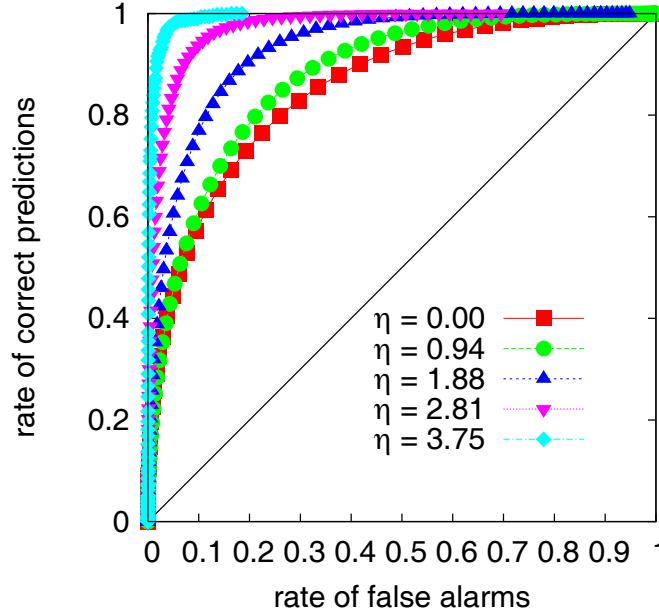


Figure 6.4: ROC curves for the Gaussian distributed AR(1) process with correlation coefficient $a = 0.75$. The ROC curves were made via predicting threshold crossings of magnitude η within 10^7 data points. Note that the quality of the prediction increases with increasing event magnitude.

Using again Eq. 3.19 we obtain the following asymptotic behavior for large values of η

$$c(\eta, x_n, a) \rightarrow \eta \left(\left(\frac{1 - \mathcal{O}(\exp(-\eta^2)/\eta)}{1 - \mathcal{O}(\exp(-\eta^2)/\eta)} \right) - \frac{1}{\sqrt{1-a^2}} \right) + \frac{ax_n}{\sqrt{1-a^2}} \frac{1}{(1 + \mathcal{O}(1/\eta^2))} + \frac{1}{\eta} \left(\frac{1 - \mathcal{O}(\exp(-\eta^2)/\eta)}{1 - \mathcal{O}(\exp(-\eta^2)/\eta)} \right), \quad \eta \rightarrow \infty. \quad (6.11)$$

This expression is larger than zero, if terms of $\mathcal{O}(\exp(-\eta^2)/\eta)$ are negligible. Hence we can conclude, that $c(\eta, x_n, a)$ is positive for large values of η and arbitrary values of x_n .

However for finite values of η we observe a dependence on the precursory variable. Figures 6.2 and 6.3 display that $c(\eta, x_n, a)$ is positive for larger values, i.e., values, which are closer to the ideal precursor $u = \infty$ or respectively $u = \max\{x_n\}$ in any finite data set. Hence we should expect larger events to be better predictable, if our alarm interval is situated in this region, i.e., if the alarm interval $I = [u - \delta, u]$ is small.

The ROC curves in Fig. 6.4 support this result. In the region of low rates of false alarms which corresponds to a small alarm interval we find a strong dependence on the event magnitude in the sense, that larger events are better predictable.

Finally one can discuss the case of the ideal precursor $u = \infty$. Inserting this value of the precursory variable into Eq. 6.10 one obtains $c(\eta, x_n, a) = 0$. This ideal precursor corresponds to the idea of an ideal ROC curve, which is identical to the vertical axis of the ROC-plot. Hence no further improvement of the ROC curve, e.g., by a change of the event magnitude, is possible.

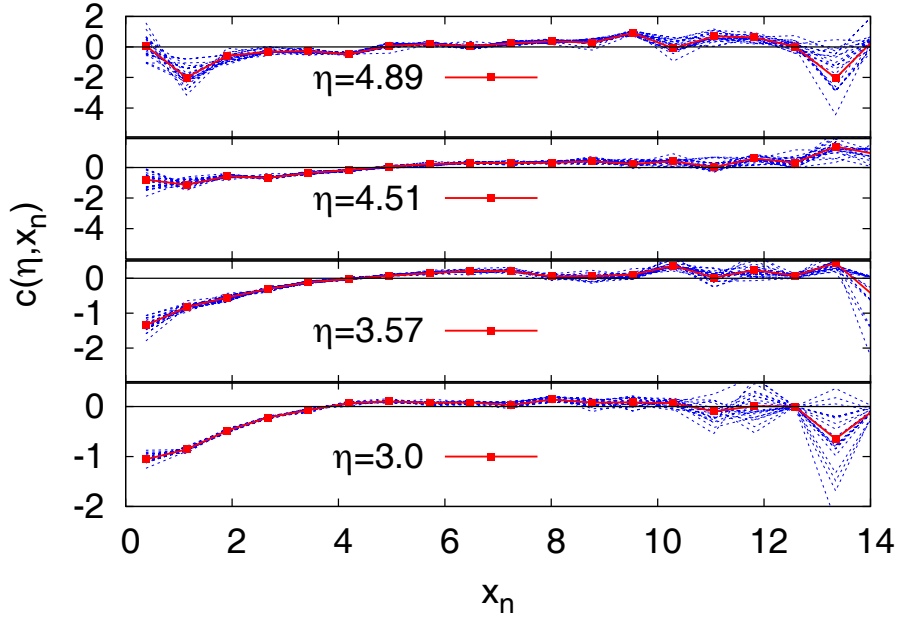


Figure 6.5: The condition according to Eq. (4.19) evaluated on 10^7 exponentially distributed AR(1) correlated data .

6.3 AR(1) Process with Symmetrized Exponential Distribution

The AR(1) data with exponential distribution were created via replacing the values of the Gaussian distributed AR(1) data with exponentially distributed i.i.d. random variables, as explained in Sec. 6.1. The exponential distributed AR(1) process has the following PDF

$$\rho(x) = \frac{\lambda}{2} \exp(-\lambda|x_n|)$$

with $\lambda = 1$ and was generated by transformation from uniformly distributed random numbers. (The uniformly distributed random numbers were generated by using the *Mersenne twister algorithm* [96].) Numerically we find the maximum of the likelihood also in the region of large values of x_n , similar to the Gaussian case with an alarm interval $I = [u - \delta, u]$, where u denotes the largest value of the data set. We compute the condition according to Eq. 4.19 and the ROC curves numerically by using 10^7 exponentially distributed AR(1) correlated data. Fig. 6.5 compares the results of the numerical evaluation of the condition $c(\eta, x_n, \lambda)$. In the vicinity of the larger values of the data set, the condition $c(\eta, x_n, \lambda)$ is positive as in the Gaussian case. The ROC curves in Fig. 6.6 support the positive magnitude dependence indicated by the test condition.

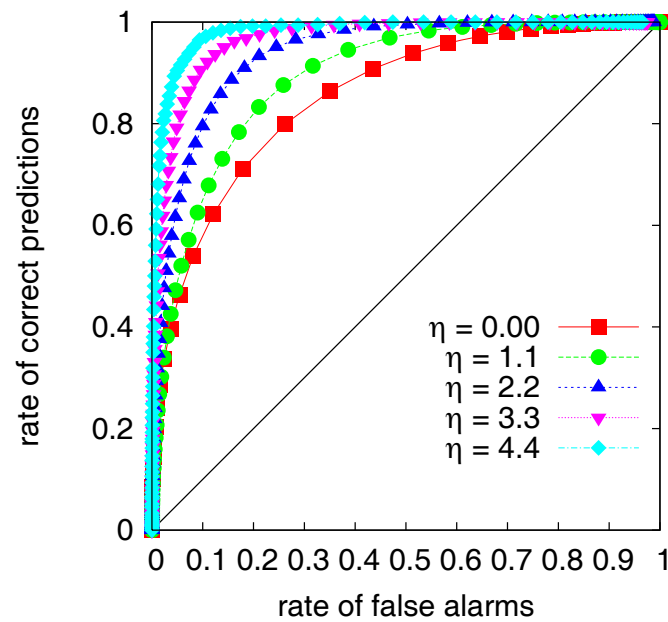


Figure 6.6: The ROC curves where made via predicting threshold crossings in 10^7 exponentially distributed AR(1) correlated data.

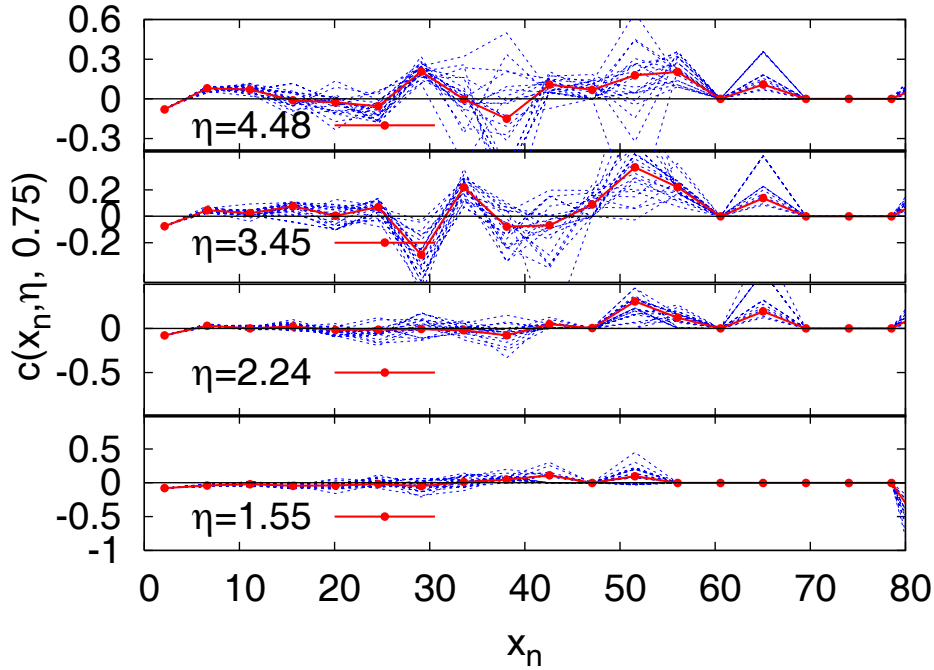


Figure 6.7: The condition $c(\eta, x_n, \alpha, \sigma)$ evaluated on 10^7 power-law distributed AR(1) correlated data with variance $\sigma = 1$, mean zero and power-law coefficient $\alpha = 3$.

6.4 Power-Law Distributed random variables

The AR(1) data with power-law distribution were created via replacing the values of the Gaussian distributed AR(1) data with symmetrized power-law distributed i.i.d. random variables, as explained in Sec. 6.1. The symmetrized power-law distributed i.i.d. random variables follow the following distribution

$$\begin{aligned} \rho(x) &= \alpha x_{min}^\alpha x^{-(\alpha+1)}, & x > x_{min} > 0 \\ \rho(x) &= \alpha |x_{max}|^\alpha |x|^{-(\alpha+1)}, & x < x_{max} < 0, \end{aligned} \quad (6.12)$$

with $x_{min} = |x_{max}| = 0.01$ and power-law coefficient $\alpha = 3$. They were generated by transformation from uniformly distributed random numbers. Since the distributions with $x_{min} = |x_{max}| = 0.01$ would not allow values in the interval $]x_{min}, x_{max}[$, the resulting random numbers were shifted to the left (right) by subtracting (adding) x_{min} . The result is a symmetrized power law distribution with mean zero and variance $\sigma = 0.01$. Finally, the values of the AR(1) process were amplified by multiplication with a constant $c_a = 100$. The resulting power law distributed AR(1) process has a variance of $\sigma = 1$, and a power law exponent of $\alpha = 3$.

Figs. 6.7 and 6.8 show the numerical results for $c(\eta, x_n, \alpha, \sigma)$ and the ROC curves. Again we choose the alarm interval to be $I = [u - \delta, u]$, with u denoting the largest value of the data set. Although $c(\eta, x_n, \alpha, \sigma)$ is less regular than in the Gaussian or the exponential case, its values are mainly above zero, especially in the region of the optimal precursor, i.e., for large values of x_n . This complies to the ROC curves in figure 6.8. Hence large threshold crossings are also within the power-law distributed AR(1) process better predictable than smaller ones. This result for *threshold crossings* in AR(1) correlated data is qualitatively different from the results for the prediction of *increments* in sequences of Pareto distributed random numbers. In Chap. 5 we found that the event magnitude has no influence on the prediction of large *increments* in sequences of exponentially distributed i.i.d. random numbers. This difference can probably be understood by the fact that the condition $c(\eta, x_n)$ is not only a function of the event magnitude η , but also a function of the event class and of the precursor values x_n .

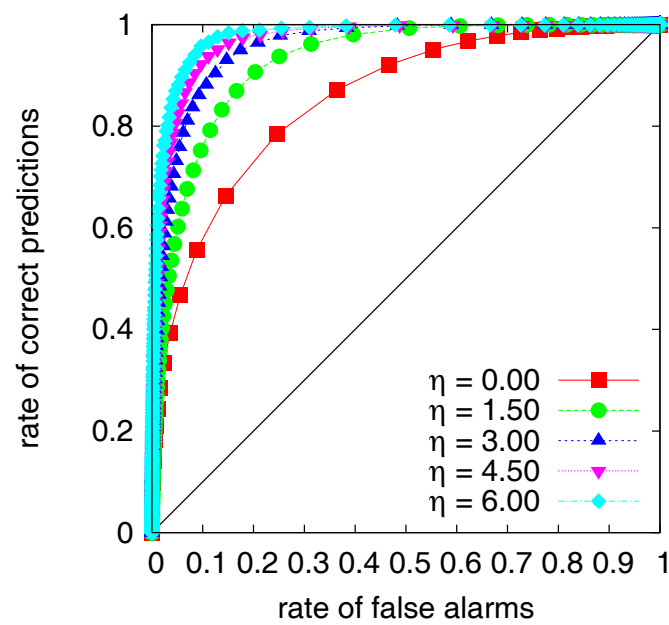


Figure 6.8: ROC-plot for the power-law distribution. The ROC curves were made via predicting increments in 10^7 data points of the AR(1) process with power-law distribution.

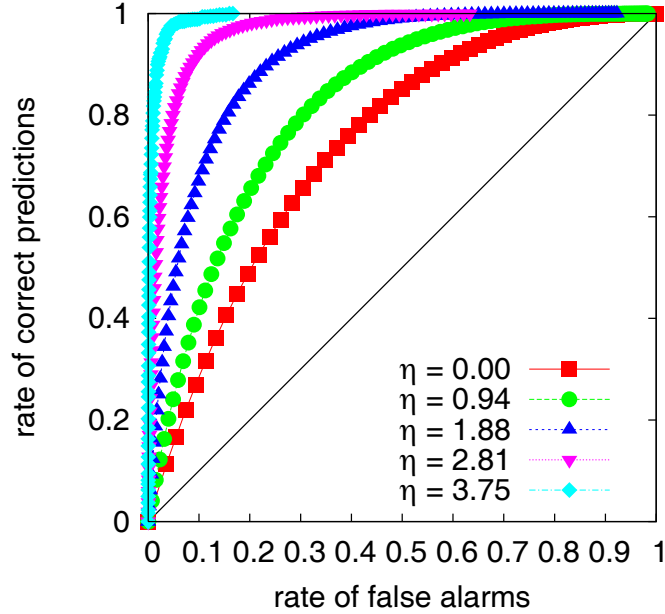


Figure 6.9: ROC curves for the Gaussian AR(1) process made according to the more realistic prediction procedure described in Sec. 6.5

6.5 A more realistic prediction procedure

Predicting an above-threshold event when the current observation itself is already above the threshold might not be really relevant in most applications. Therefore we modify here the sample on which predictions are to be made: We define as events the subset of previous events, where not only the future value is above threshold, but simultaneously the current value is below threshold.

$$\chi_{n+1} = \begin{cases} 1 & : x_{n+1} \geq \eta, \quad x_n < \eta \\ 0 & : x_{n+1} < \eta, \quad x_n < \eta \end{cases} \quad (6.13)$$

The events $\chi_n = 1$ according to this definition are a subset of the previously discussed events and the alarm interval reduces to $I = [\eta - \delta, \eta]$. Hence, this modification reduces the number of events in the time series and might render the prediction task more difficult. The corresponding ROC-curves in Figs. 6.9-6.11 show qualitatively the same dependence on the event magnitude as the ROC curves obtained in the previous section: Threshold crossings in Gaussian, approximately exponentially distributed, and approximately power-law distributed AR(1) processes are better predictable, the larger they are.

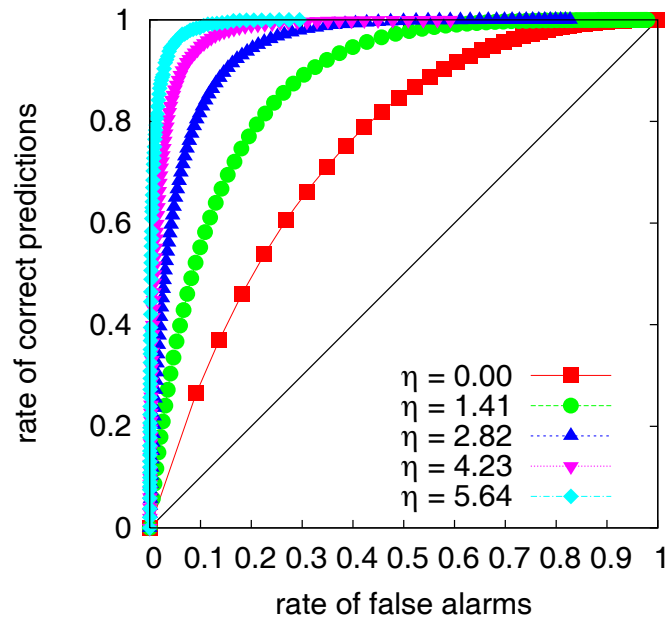


Figure 6.10: ROC curves for the exponentially distributed AR(1) process made according to the more realistic prediction procedure described in Sec. 6.5

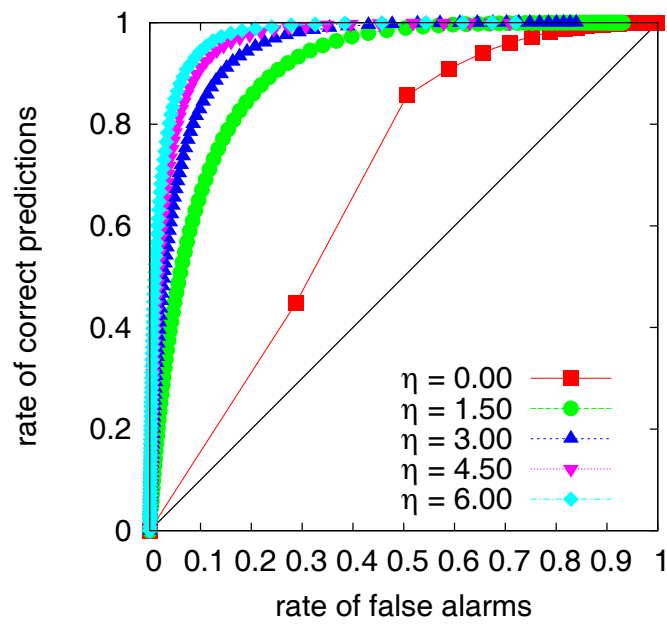


Figure 6.11: ROC curves for the power law distributed AR(1) process made according to the more realistic prediction procedure described in Sec. 6.5

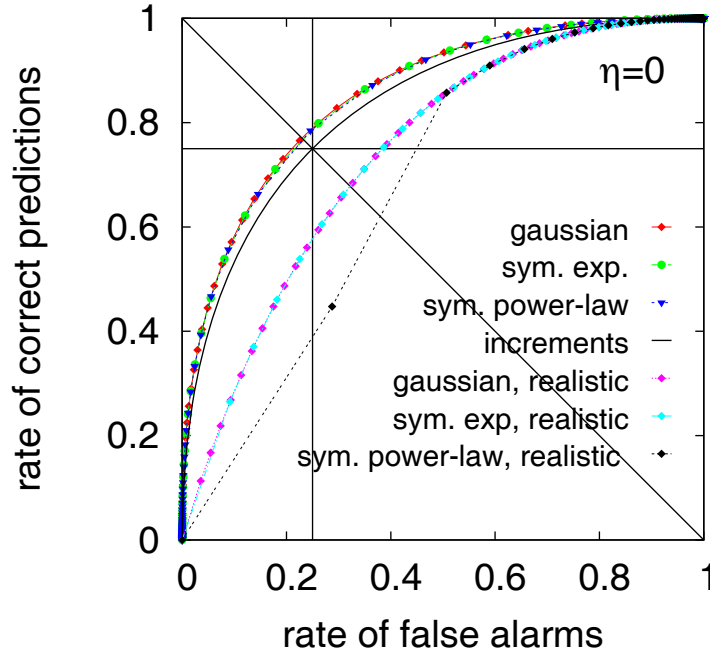


Figure 6.12: ROC curves for $\eta = 0$ for different distributions, the annotation “realistic” refers to the ROC curves in described in Sec. 6.5.

6.6 ROC curves for $\eta = 0$

Analogous to the considerations for the predictions of increments of magnitude zero (see. Sec. 5.7), we observe, that the ROC curves indicating the success of the prediction of threshold crossings of size zero, i.e. the sign of x_{n+1} coincide as it is displayed in Fig. 6.12.

We can understand this coinciding curves by considerations analogous to Sec. 5.7. Since $p[\chi_n(0) = 1] = 1/2$, the likelihood ratio equals to the ratio of odds, which reads for an alarm interval $I = [u - \delta, u]$ with u being the largest value of the data set

$$\Lambda(u, \delta, 0) = \frac{p[\chi_n(0) = 1|u - \delta]}{p[\chi_n(0) = 0|u - \delta]} = \frac{p[\chi_n(0) = 1|u - \delta]}{1 - p[\chi_n(0) = 1|u - \delta]}. \quad (6.14)$$

Hence the slope of all ROC curves for $\eta = 0$ follow the behavior indicated by this ratio and since all curves cross the origin and the point (1,1), we observe coinciding ROC curves.

However, one can now ask, why we obtain different “universal” ROC curves for the prediction of increments $\eta = 0$, threshold crossings $\eta = 0$ and threshold crossings in the sense of Sec. 6.5. In order to understand this, one has to remind oneself, that that the slope $\Lambda(u, \delta, 0)$ as it is given by Eq. (6.14) is a function of the size of the alarm interval δ . However, in terms of the ROC curves we typically think of $\Lambda(u, \delta, 0)$ as a function of the rate of false alarms, which can be misleading. Since the mapping which relates δ to the rate of false alarms is dependent on the definition of the event and the corresponding alarm volume, we find different ROC curves for different types of events.

6.7 Summary

We study the magnitude dependence of the quality of predictions for threshold crossings in autocorrelated processes of order one and in measured accelerations in a free jet flow. Using the present value x_n as a precursory variable we predict threshold crossings at a future time step x_{n+1} via statistical considerations.

We are especially interested in the influence of the probability distribution of the underlying process on changes in the quality of the predictions, which are evoked by focusing on different

event magnitudes. For Gaussian, exponential and power law distributed AR(1) processes we find, that larger threshold crossings are better predictable, the higher the threshold. In all cases studied the behavior of the ROC curves was reasonably well reflected by the condition $c(\eta, x_n)$, which is an expression that depends on the total probability to find events and the likelihood to observe an event after a given value of x_n . This theoretical results could in principle help to understand the effects reported for avalanches in systems, which display self organized criticality [23].

These results for the prediction of threshold crossings are in contrast to the results for the prediction of *increments* in i.i.d. random numbers presented in Chap. 5. For the prediction of increments we found qualitatively different magnitude dependences in dependence of the different distributions of the data, whereas predictions of threshold crossings show the same magnitude dependence for the distributions under study. This difference can be understood by taking into account the different regimes in which we find the optimal precursors: When predicting increments, the optimal precursors are typically among the smallest values in the data set, while for the prediction of threshold crossings, large values are optimal. Furthermore threshold crossings form a different class of events. Hence both PDFs which contribute to the value of $c(\eta, x_n)$, namely the likelihood and the total probability to find events are different. Hence we should not be surprised to find different results for the predictability of larger events.

In summary we find, that threshold crossings in AR(1) processes are the better predictable, the larger they are.

Chapter 7

Predictions of Events in Fluid Flows

In this chapter we investigate the predictability of increments and threshold crossings in measurements of the flow velocity in a round air into air free jet and of wind speed recorded in the boundary layer. In both fluid flows we observe intermittent behavior, visible in form of non-Gaussian, i.e, exponential distributions of small velocity increments. However, velocity increments, between well separated time steps, corresponding to larger scale structures are still Gaussian distributed. Since, the exponentially and the Gaussian distributed data sets of increments are obtained from the same original data set, we can assume, that many features of the data sets are identical, although they differ in distribution. Hence, the fluid flow velocities provide us with the opportunity to test the influence of the underlying distribution on the magnitude dependence. However, in contrast to the previous studies on synthetic data, we deal here with realistic data sets, containing measurement noise and correlations.

7.1 Distributions of velocity increments in the presence of dissipation range intermittency

Fully developed turbulent flows are characterized by a flux of energy, which is injected into the fluid motion at large scales (*integral scale*) and is then transported to smaller scales (*dissipation range*).¹ This *turbulent cascade* is typically investigated, by studying the longitudinal (in direction of the main flow) velocity increments [97], [98]

$$v(r, t) = \mathbf{e} \cdot [\mathbf{u}(\mathbf{x} + r\mathbf{e}, t) - \mathbf{u}(\mathbf{x}, t)], \quad (7.1)$$

where $\mathbf{u}(\mathbf{x}, t)$ denotes the velocity field and \mathbf{e} a unit vector aligned to the main flow.

It is well known ([99], [100], [101], [102], [97]) and many others, that that in fully developed turbulence the fluid motion at small scales is characterized by *intermittency*. In the context of dynamical systems, intermittency is typically discussed as the alternation between regular phases and short irregular chaotic bursts (see. e.g. [103]). The influence of intermittency on fluid motion is mainly described by its influences on the statistics of the velocity increments, namely as a violation of the scale independence of moments of the velocity $\langle v(r, t)^n \rangle$ ([104], [105], [97], [98]) or as “quasi-exponential tails of the PDF of longitudinal [velocity] increments” ([106], [98]). We will focus on the latter feature, which is more relevant in order to test for the magnitude dependence of predictions. Numerous measurements and simulations ([107], [108], [109], [110] and many others) indicate, that the PDFs of longitudinal increments are approximately Gaussian if r is sufficiently large and they “develop exponential wings” [97] if r is decreased. These exponential distributions can even turn into stretched exponential distributions for even smaller r . The distributions of the data sets under study illustrate this effect (see Figs. 7.1 and 7.6.)

¹Precise definitions of the terms “*integral scale*”, “*inertial range/scale*” and “*dissipation range/scale*” can be found in standard textbooks on turbulence, such as [97]. In the context of this study we simply distinguish between “small scales” and “large scales”.

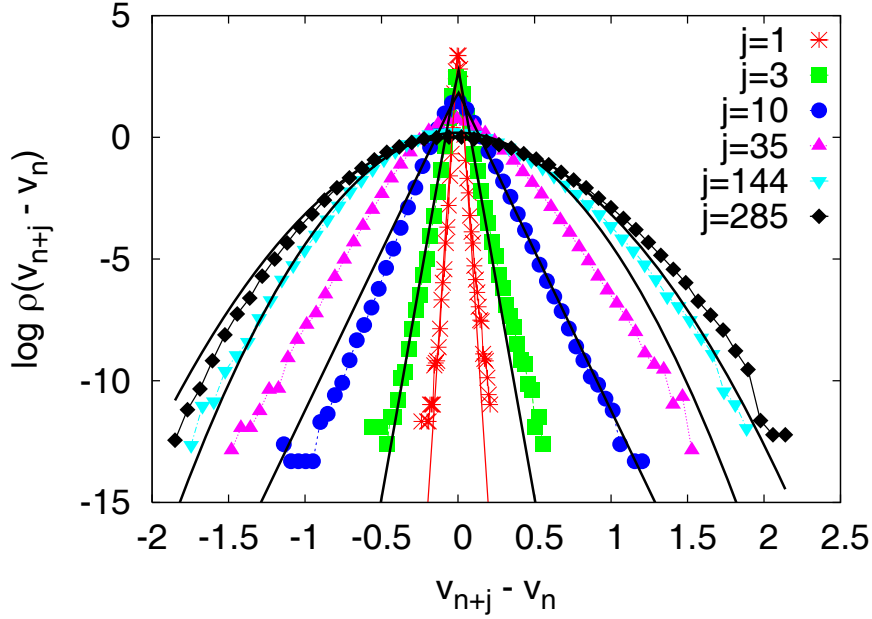


Figure 7.1: PDF of the increments $a_{n,j} = v_{n+j} - v_n$ with $j = 1, 3, 10, 35, 144, 285$. The black lines correspond to Gaussian and exponential PDFs with appropriate values for the standard deviation or the coefficient λ .

A well known parametrization, which allows to describe the evolution of the shape of the PDFs was developed by Castaing, Gagne and Hopfinger [109]. The idea of this description consist in assuming, that the deviation from a Gaussian results from the influence of the energy transfer rate towards smaller scales ϵ_r . This energy transfer rate is according to the hypothesis of Kolmogorov [104] and Obukov [105] (later refined by Kraichnan [111]) assumed to follow a lognormal distribution. Since the variance $\sigma \propto (r\epsilon_r)^{1/3}$ of the velocity increment PDF is essentially given by the scale r and the energy transfer rate, Castaing proposes also a log normal distribution of the variance. Combining the Gaussian distribution of the increments with the lognormal distribution of its variance and taking into account the observed skewness of the distributions yields

$$p[v] = \frac{A(a_s)}{2\pi\lambda} \int d\sigma \frac{1}{\sigma^2} \exp \left[-\frac{v^2}{2\sigma^2} \left(1 + a_s \frac{v/\sigma}{\sqrt{1 + v^2/\sigma^2}} \right) \right] \exp \left(-\frac{\ln \left(\frac{\sigma}{\sigma_0} \right)}{2\lambda^2} \right), \quad (7.2)$$

where σ_0 is the most probable variance of v , $\lambda = \sqrt{\langle (\ln \sigma)^2 \rangle}$ denotes the variance of the lognormal distribution of σ and $A(a_s)$ and $a_s > 0$ are normalization constants, describing the skewness. The success of this description leads to the fact, that the whole phenomena of changing shapes of increment distributions is often called *Castaing distributions*.

Renner, Peinke and Friedrich already showed [98] that the free jet data set under study is indeed well characterized by the Castaing distributions. Thus, we will concentrate on investigating not the full Castaing distribution, but simply on their asymptotics, i.e., Gaussian and exponential distributions.

7.2 Predictions of Increments in a Free Jet Flow

Thus the incremental data sets $a_{n,j}$ provides us with the opportunity to test the results for statistical predictions within Gaussian and exponential distributed i.i.d. random numbers on a data set, which exhibits correlated structures. In this section, we apply the method of statistical inference to predict acceleration increments in free jet data. Therefore we use a data set of

1.25×10^7 samples of the local velocity measured in the turbulent region of a round free jet. The data were measured in the group of Joachim Peinke and all details of the experiment can be found in [98]. That is why we will simply give a brief overview here.

Using hot wire anemometry, the velocity of the air in front of a nozzle is measured at a sampling rate of 8 Hz and at a position where the flow can in good approximation be considered as being isotropically turbulent. Taking increments a_i of such a sequence v_i over short time intervals, $a_i = v_{i+j} - v_i$, for j small, yields approximately a symmetric exponential distribution for a_i , whereas for long time intervals, i. e. large j , the distribution of the increments is approximately Gaussian, see Fig. 7.1.² The Taylor hypothesis $r = u(t_{n+1} - t_n)$, with the mean velocity u , allows to relate the time-resolution to a spatial resolution.

We are now interested in predicting increments of the acceleration $a_{n+\kappa,j} - a_{n,j} \geq \eta$ in the incremental data sets $a_{n,j} = v_{n+j} - v_n$. In the following we concentrate on the incremental data set $a_{n,10}$, which has an asymptotically exponential PDF and the data set $a_{n,144}$, which has an asymptotically Gaussian PDF.

The time horizon κ between the prediction and the event corresponds to the time through which the increment is defined. Since the data is strongly correlated we choose a time horizon $\kappa = 285$, which corresponds to an increment 35.625 s ahead in time, in order to observe sufficiently many large increments. In other words, the short-range persistence of the process could otherwise prevent us from detecting large events.

As in the previous sections we are hence exploiting the statistical properties of the time series to make predictions, rather than the dynamical properties.

We can now use the evaluation algorithm which was tested on the previous examples to evaluate the condition for these data sets. The results are shown in Fig. 7.2. We find that at least for larger values of η the main features of $c(x_n, \eta)$ for the exponential and the Gaussian case as described in Chap. 5 are also present in the free jet data. For larger values of η , $c(a_{n,j}, \eta)$ is either larger than zero in the Gaussian case ($j = 144$) or equal to zero in the exponential case ($j = 10$) in the region of interesting precursory variables, i.e., small values of $a_{n,j}$.

However, the presence of the exponential and the Gaussian distributions is also visible in the corresponding ROC curves. In order to generate the ROC curves we used a threshold of the likelihood instead of a specific precursor. In this setting we give an alarm for an extreme event, whenever the likelihood that an extreme event follows an observation is larger than a given threshold value.

In the exponential regime ($j = 10$) shown in Fig. 7.3(a) the ROC curves for different event magnitude η almost coincide, although the range of η is larger ($\eta \in (0, 6.71)$) than in the Gaussian case shown in Fig. 7.3 (b). For $j = 144$ the ROC curves are further apart, which corresponds to the results of Secs. 5.2 and 5.4.

This corresponds to the previous observation, that the magnitude dependence is much more prominent if the underlying distribution is Gaussian. Hence, the specific dependence of the ROC curve on the event magnitude can also in the case of correlated data sets be characterized by the PDF of the underlying process.

²For longitudinal velocity increments, one wing of the distributions is higher than the other. This effect can be understood via Kolmogorov's four-fifths law, which demands a non-zero skewness of the velocity increment [97].

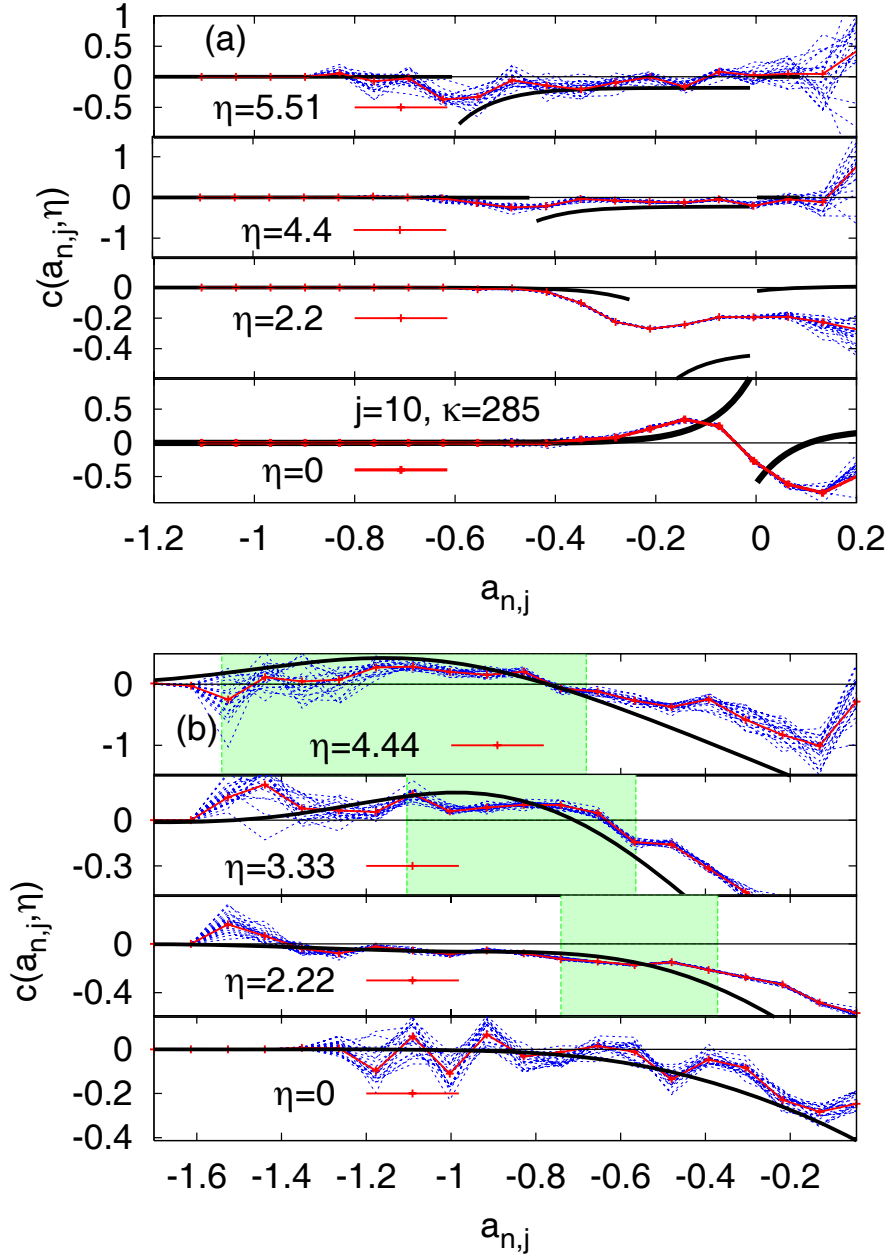


Figure 7.2: Transition from the exponential regime (a) to the Gaussian regime (b) characterized via the numerical evaluation of $c(x_n, \eta)$. The black line corresponds to the analytic results for the Gaussian and the exponential PDF, fitted to the PDFs of the increments, as it is shown in Fig. 7.1. For larger values of η the main features of $c(x_n, \eta)$ for the exponential and the Gaussian case as described in Sec. 5.2 and 5.4 are reproducible. For larger values of η we find that if $-\sigma\eta < a_{n,j} - \sigma\eta/2$ $c(a_n, \eta)$ is either larger than zero in the asymptotically Gaussian case ($j = 144$) or equal to zero in the asymptotically exponential case ($j = 10$).

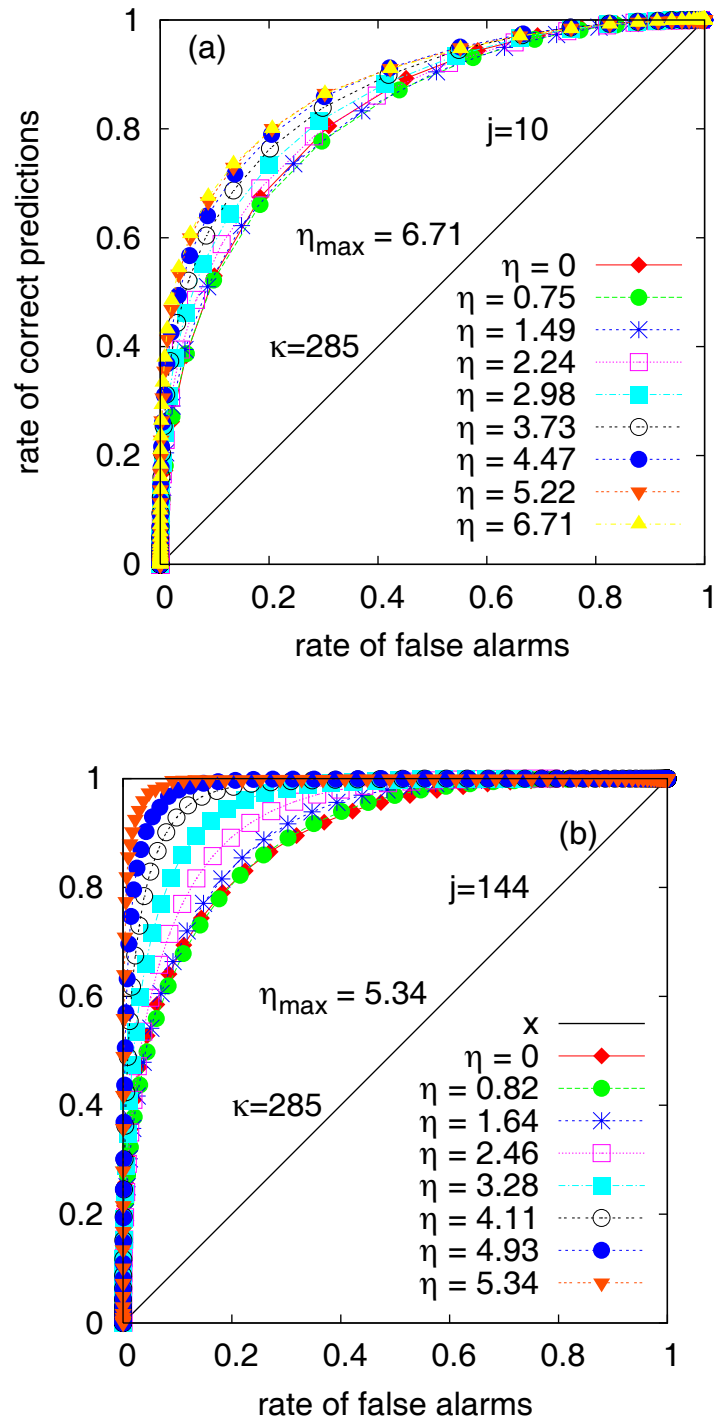


Figure 7.3: Transition from exponential ROC curves (a) to Gaussian ROC curves (b). In the exponential case ($j = 10$), shown in (a) the ROC curves for different event magnitude η are almost the same, although the range of η is larger ($\eta \in (0, 6.71)$) than in the Gaussian case shown in (b). For $j = 144$ the ROC curves are further apart, which corresponds to the results for Gaussian ROC curves (see Sec. 5.2)

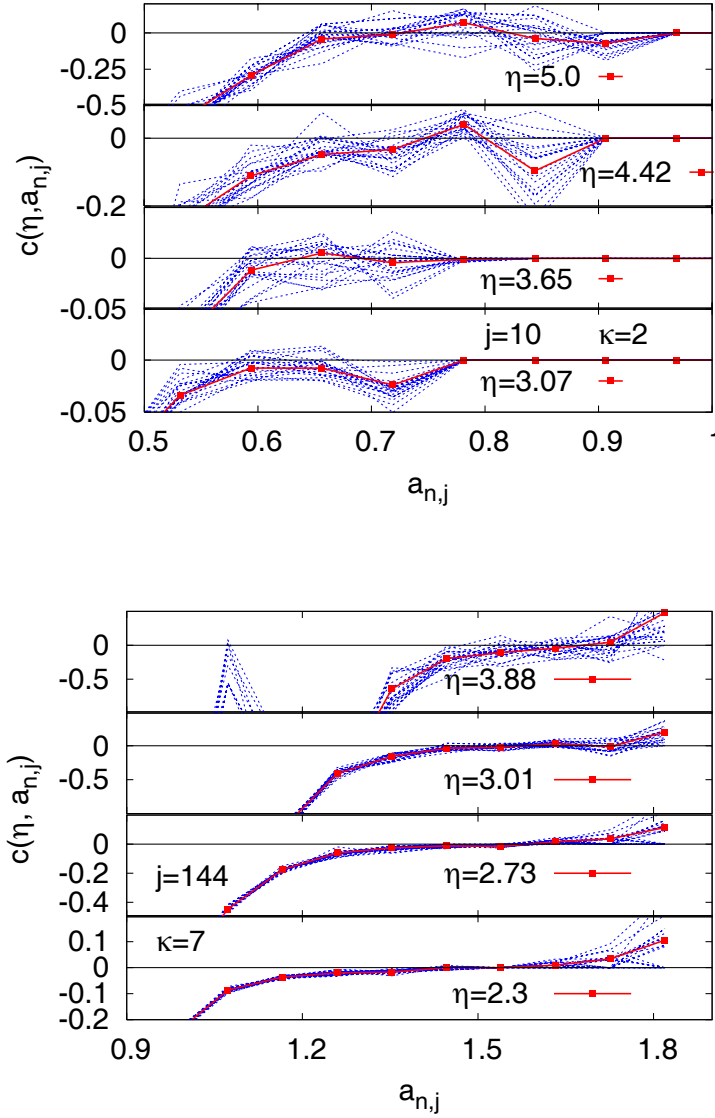


Figure 7.4: The condition $c(\eta, a_{n,j})$ evaluated for the prediction of threshold crossings in the free jet data set. Figure (a) shows the exponential regime (a) and Fig. (b) displays the results in the Gaussian regime (b). Qualitatively both pictures are not very different: In both cases, $c(\eta, a_{n,j})$ is positive for larger values of the precursory variable $a_{n,j}$

7.3 Predictions of Threshold Crossings in a Free Jet Flow

We are now interested in predicting larger values of the acceleration $a_{n+\kappa,j} \geq \eta$ in the incremental data sets $a_{n,j} = v_{n+j} - v_n$. In the following we concentrate on the data set $a_{n,10}$, which has an asymptotically exponential PDF and the data set $a_{n,144}$, which has an asymptotically Gaussian PDF. Since the free jet data are correlated, predicting the next time step would be equivalent to predicting persistence. That is why we choose a prediction horizon of seven time steps, i.e., $\kappa = 7$ in the Gaussian regime ($j=144$) and a smaller time horizon $\kappa = 2$ in the less correlated exponential distributed increments ($j=10$). As in the previous chapters we are hence exploiting the conditional probabilities of the time series to make predictions.

We can now use the algorithm which was tested on the previous examples to evaluate the condition and the corresponding ROC curves for these data sets. The results are shown in Figs. 7.4 and 7.5. In both examples the evaluation of the condition $c(\eta, a_{n,j})$ reflects the behavior of the ROC curves.

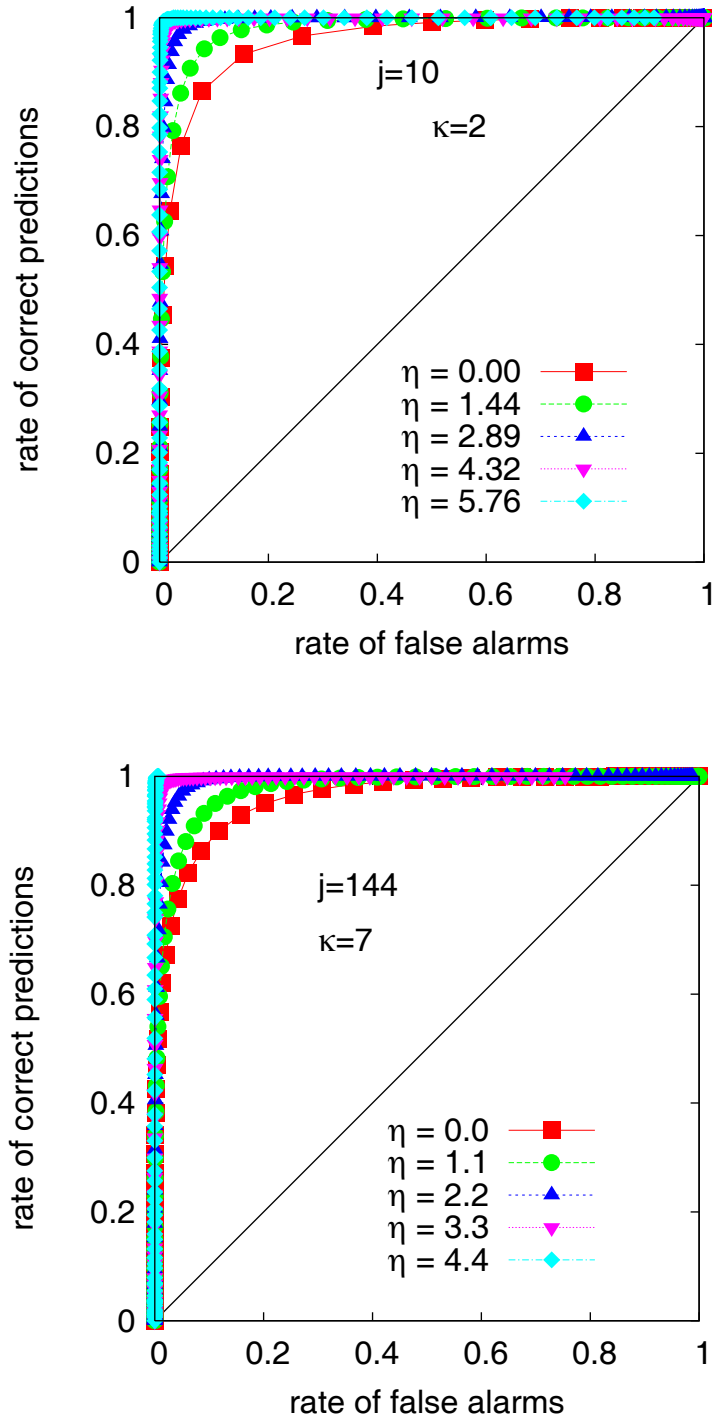


Figure 7.5: ROC curves for the prediction of threshold crossings in the free jet data set. The results for the exponential distributed velocity increments are on the left, the results for the Gaussian regime on the right.

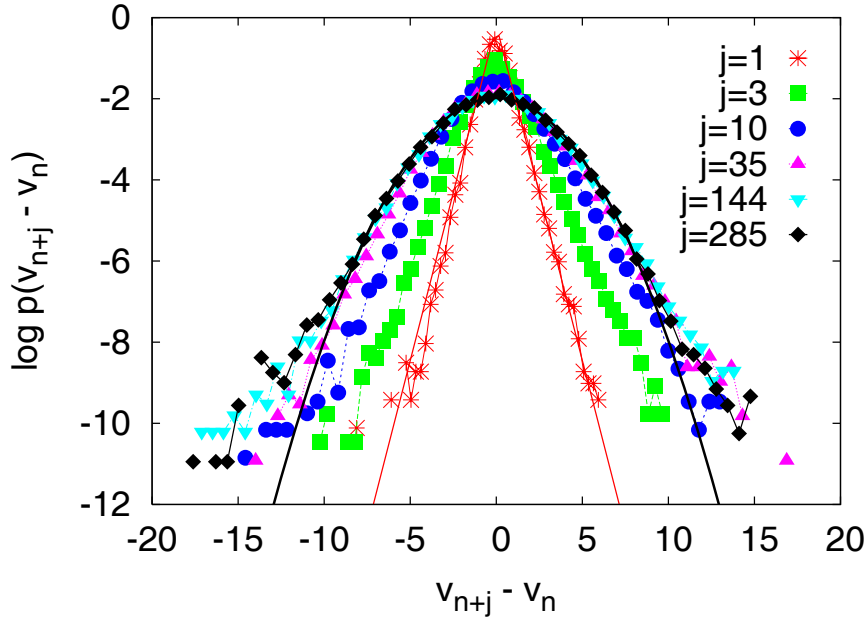


Figure 7.6: Histograms of velocity increments in wind speed. Again, we find that on short scales the pdfs of the increments follow approximately an exponential distribution, whereas the increments are approximately Gaussian distributed for larger j . However, the deviations from the asymptotic distributions are larger than in the laboratory experiment, especially in the tails of the distributions.

7.4 Prediction of Increments in Wind Speeds

As a second example, we study wind speeds recorded at a measurement site in the north of Germany, 66 m above ground at a sampling rate of 1 Hz. These measurements were made by Julia Gotschall in the group of Joachim Peinke at the university of Oldenburg. These data reflect the full complications of field measurements, including non-stationarity and inaccuracies, but also represent a much more complicated turbulent state, namely boundary layer turbulence which is strongly affected by the interaction of the air flow with the earth's surface. Hence the deviations from the asymptotic distributions are larger than in the laboratory experiment, especially in the tails of the distributions.

We predict large increments in the acceleration of the wind, so called turbulent gusts which are of relevance for controlling wind turbines or scheduling aircraft take-off and landing. The time horizon κ between the prediction and the event corresponds to the time through which the increment is defined. Since the data is strongly correlated we choose a time horizon $\kappa = 35$, which corresponds to an increment 35 s ahead in time, in order to observe sufficiently many large increments. The predictions were made via identifying the precursors in the first part of the data set (about 40000 data points) and then predicting in the second part (also 40000 data points). Thus, we made predictions on a part of the data set, which was not used to determine suitable precursors (“training”). The resulting ROC curves (Fig. 7.7) show again a better predictability of larger events if the data set $\{a_i\}$ is asymptotically Gaussian distributed and a much weaker dependence on the event size in the asymptotically exponential case.

7.5 Summary

We investigated the prediction of increments and threshold crossings in time series of longitudinal velocity increments observed in a free jet experiment and in wind speed recordings. If we choose the prediction horizon to be sufficiently large, for the prediction of increments or relatively

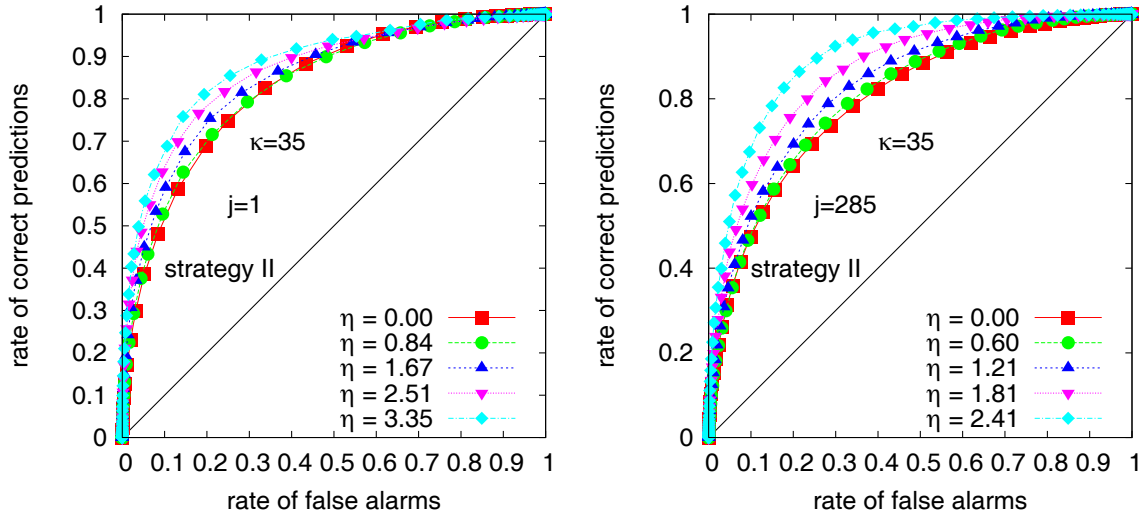


Figure 7.7: ROC statistics for the prediction of large increments of wind speeds for different event magnitudes, predicting $\kappa = 35$ time steps ahead.

small for the prediction of threshold crossings, we can find similar magnitude dependence as it was discussed in previous chapters for predictions in AR(1) processes and sequences of random numbers. Thus, we can apply the test condition also in the presence of correlations and noise of measurement data. Again, we see, that the test condition is only influenced by the correlation via the joint PDF of event and precursor, as it was discussed in Chap. 4.

Chapter 8

Predicting errors in weather forecasts

In this chapter we are interested in studying the magnitude dependence for predictions which are not entirely based on purely statistical considerations. In fact we postprocess the output of model based weather forecasts by using statistical inference and thus try to combine the information provided by both prediction techniques. In contrast to pure statistical forecast verification we apply precursors based on the output of an ensemble forecast system to predict large differences between high resolution forecasts and the actual measurements (verification). Having in mind our previous considerations about the dependence on the event magnitude, we are of course not only interested in learning, whether we can identify large deviations of the high resolution forecast and the verification using the corresponding ensemble forecast, but also in studying the magnitude dependence of these predictions.

8.1 The ECMWF's medium range temperature forecast

The weather forecasts and the corresponding verifications were provided by the European Center for Medium Range Weather Forecasts (ECMWF). The data contain the ECMWF's medium range ensemble forecast and the ECMWF's high resolution forecast and the corresponding verifications.

Very loosely speaking a weather forecast model consists of a system of differential equations, describing the state of the atmosphere. Processes occurring on smaller scales, than the grid resolution of the model are described by stochastic components. The model equations are then fed with initial conditions measured by observation stations on the earth, ships and satellites and evolved up to ten days into the future.

The high resolution forecast consists of the ECMWF's operational forecast model $T_L799L91$. The numbers 799 and 91 indicate the resolution in terms of 799 linear spectral components of a spherical harmonic which determine the horizontal resolution of the grid to be about 25 km and 91 vertical levels of atmospheric layers.

The idea of the ensemble forecast is to provide a measure of the sensitivity on the initial conditions, i.e, for the fact that small errors in the observation can evolve into large scale errors within the 10 day forecast period. Thus, selected small amplitude perturbations are added to the observations in order to create a range of slightly different initial conditions, which are then chosen to be the initial conditions of an ensemble of forecast models. The ECMWF's ensemble forecast consists of 50 ensemble members, fed with perturbed initial conditions plus one member which evolves the unperturbed conditions, the *control*. The resolution of the ensemble forecast models is lower, than the one of the operational forecast, i.e, the model is a $T_L255L40$ model. More details about the ECMWF's forecast models can be found in [24].

The forecasts we used were issued and verified twice a day at 0:00 and 12:00 in the years

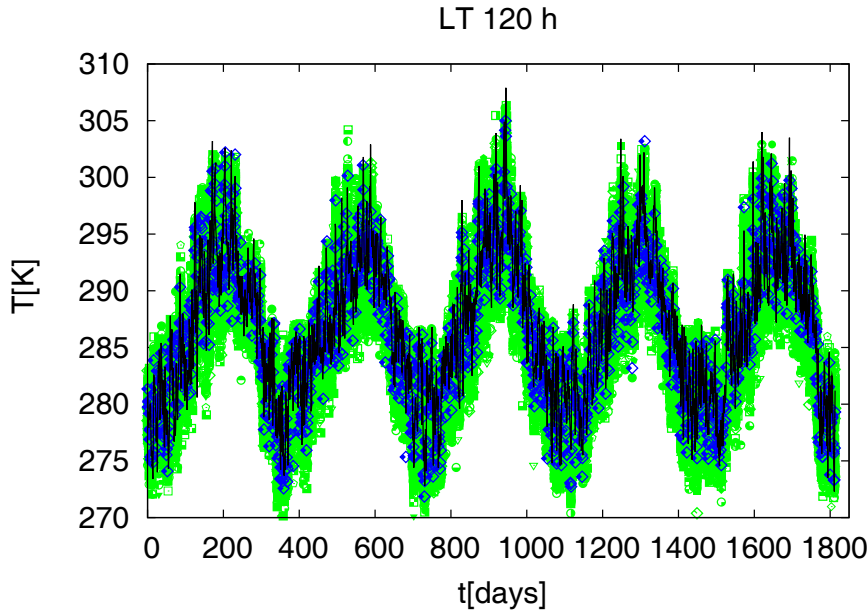


Figure 8.1: The temperature measured at London Heathrow at 12:00 from 2001 to 2005 is indicated by the black line. The corresponding ensemble forecast is represented by the green symbols, the high resolution forecast is represented by the blue symbols.

from 2001 to 2005. Thus, each data set, the one corresponding to the forecast for noon and the one corresponding to the forecast for midnight, contains data corresponding to about 1800 time steps. Furthermore each data set consists of five sub-data sets, which correspond to four neighboring grid points on the circulation model. The fifth data set is obtained by interpolation of the four surrounding data sets. For all following considerations we always use this interpolated data set.

In our investigations we focused on the temperature forecasted for and measured at London Heathrow airport, $51^{\circ}29'N$ $000^{\circ}27'W$ at noon and we interpreted the control as a 51th ensemble member.

In the following sections we present the results for the numerical evaluations of the quantities, which were specified in the previous section, i.e., the likelihood to find events, the ROC-curves and the corresponding area under the ROC curve (AUC, see Sec. 2.3.4), the Brier Score and the Ignorance. We will show the results for a lead time (i.e., the time between the issuing of the forecast and the time, the forecast is made for) of 120h. However all calculations presented here were also made for the lead times 24h, 48h, 96h, 144h, 168h, 192h, 216h, 240h. Since we do not find qualitatively different results for the magnitude dependence we restrict ourselves to the discussion of the results for a lead time of 120h.

The corresponding data sets of high resolution forecast, ensemble forecasts and verification presented in Fig. 8.1 have the following features:

data set	mean	standard variation	maximal value	minimal value
$\{y_n\}$, verification	287.11 K	6.27 K	307.9K	272.3 K
$\{h_n\}$, high res.	286.57 K	6.24 K	271.86 K	305 K
$\{x_n\}$, ensemble	286.02 K	6.05 K	306.33 K	270.1 K
$\{ y_n - h_n \}$	1.98 K	1.61 K	12.35K	0K
$\{ y_n - x_n \}$	1.96 K	1.71 K	14.05 K	0 K

event magnitude	number of events	magnitude in units of σ_y	magnitude in units of $\sigma_{ y-h }$
≥ 0 K	1814	0	0
≥ 0.41 K	1560	0.07	0.26
≥ 0.82 K	1342	0.13	0.51
≥ 1.25 K	1106	0.2	0.78
≥ 1.65 K	901	0.26	1.02
≥ 2.06 K	714	0.33	1.28
≥ 2.47 K	545	0.4	1.53
≥ 2.88 K	414	0.46	1.79
≥ 3.29 K	314	0.52	2.04
≥ 3.71 K	237	0.59	2.30
≥ 4.12 K	175	0.66	2.56
≥ 4.53 K	130	0.72	2.81
≥ 4.94 K	105	0.79	3.07
≥ 5.35 K	72	0.85	3.32
≥ 5.76 K	53	0.92	3.58
≥ 6.17 K	42	0.98	3.83
≥ 6.59 K	30	1.05	4.09
≥ 7.0 K	24	1.12	4.35
≥ 7.40 K	14	1.18	4.6
≥ 7.82 K	11	1.25	4.86
≥ 8.23 K	10	1.31	5.11
≥ 8.63 K	6	1.38	5.36
≥ 9.06 K	6	1.45	5.63
≥ 9.47 K	6	1.51	5.88
≥ 10.29 K	5	1.64	6.39
≥ 10.7 K	4	1.71	6.65
≥ 11.11 K	3	1.77	6.9
≥ 11.53 K	1	1.84	7.16
≥ 11.94 K	1	1.9	7.12

Table 8.1: The number of events $|y_n - h_n|$ for h_n being generated with lead time 120 h.

8.2 Definition of the Events

We define the events under study to be large differences between the issued high resolution forecast h and the verification y . Hence the definition of the events is given by

$$\chi_n(\eta) = \begin{cases} 0 & \text{if } |y_n - h_n| < \eta \\ 1 & \text{if } |y_n - h_n| \geq \eta \end{cases} \quad (8.1)$$

Since it is – in the context of weather forecasting – more relevant to measure the absolute difference between a predicted temperature and the observed temperature, we did **not** scale the event magnitude by the standard deviation of the data set in this section. Thus, **the values for η presented in the figures are absolute values, measured in Kelvin.**

In order to compare the results to the previous chapters, in which we measured the event magnitudes in units of the standard deviations, we present the size measured in units of the standard deviation in Tab. 8.1, which contains also the numbers of events found in the data.

8.3 Identification of the precursor

We search for precursors for differences between the high resolution and the verification, by investigating the differences of the high resolution forecast and the ensemble forecast. More precisely we use the number of ensemble members for which the difference to the high resolution is larger than η as a precursory variable v :

$$v = \#\{i, |h_n - x_n^i| \geq \eta\} = v_n(\eta). \quad (8.2)$$

Thus we construct a time series of the precursory variable $\{v_n(\eta)\}$, using the ensemble forecasts and the high resolution forecast. In analogy to the previous studies, we give an alarm for an event $\chi_n(\eta)$, if we find v_n to be in a tolerance interval, constructed, such that it contains a specified precursor u . In the following we will compare two possible choices of u leading to two different alarm volumes.

Maximum Likelihood Method The first method of identifying u consists in using a maximum likelihood approach, as it was applied in the previous chapters for prediction in one-dimensional time series. Thus, we define our precursor in the following way

$$u_{like}(\eta) = \{v_n(\eta) : p[\chi_n(\eta) = 1 | v_n(\eta)] = \max p[\chi_n(\eta) = 1 | v_n(\eta)]\}. \quad (8.3)$$

Consequently the alarm volume $I_{like}(\eta, \delta) = [u_{like}(\eta) - \frac{\delta}{2}, u_{like}(\eta) + \frac{\delta}{2}]$ is centered around this precursor value. As for the predictions in the previous chapters we can express this procedure by defining a decision variable

$$A[v_n(\eta), u_{like}(\eta), \delta] = \begin{cases} 1 & : \text{ if } v_n \in I_{like}(\eta, \delta) = [u_{like}(\eta) - \frac{\delta}{2}, u_{like}(\eta) + \frac{\delta}{2}], \\ 0 & : \text{ otherwise.} \end{cases} \quad (8.4)$$

Threshold Method The second method of identifying suitable precursory structures is based on “educated guessing” more than on strict mathematical reasoning. We simply give an alarm for an event $\chi_n(\eta)$ if the number of ensemble members, deviating from the high resolution forecast is larger than a specified number. Thus our tolerance volume is given by

$$I_{thresh}(\eta, \delta) = \{v_n(\eta) \geq \delta\}, \quad \text{with } \delta \in (0, 51). \quad (8.5)$$

A decision variable describing the prediction is thus given by

$$A[v_n(\eta), I_{thresh}(\eta, \delta)] = \begin{cases} 1 & : \text{ if } v_n \in I_{thresh}(\eta, \delta) = \{v_n(\eta) \geq \delta\}, \\ 0 & : \text{ otherwise.} \end{cases} \quad (8.6)$$

A possible reasoning of this educated guess goes as follows: In situations, in which we would expect the high resolution forecast to differ significantly from the verification, we assume, that this failure of the forecast is due to an increased sensitivity on small perturbations in the initial conditions. We can then hope, that this sensitivity is not only present in the atmosphere, but also reflected by the ECMWF’s circulation model. Consequently, we would also expect a large number of ensemble members to deviate significantly from the high resolution forecast, which is then turned into an alarm for a deviation.

Note, that the alarm volumes and decision variables are chosen such that the limit $\delta \rightarrow 0$ corresponds to the vicinity of the origin, and $\delta \rightarrow 51$ corresponds to the point (1, 1), as it was done for the ROC curves in the previous chapters.

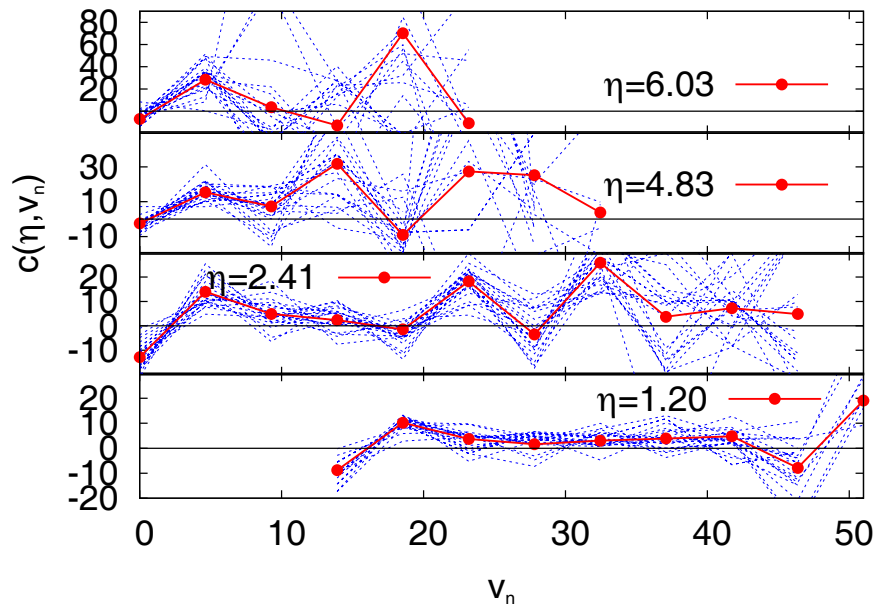


Figure 8.2: The condition $c(\eta, v_n(\eta))$ evaluated for the prediction of differences between the high resolution forecast as defined in Eq. (8.1). The red symbols correspond to the results for the original data set. The dashed blue lines represent the results for the resampled data set (resampling with repetition).

8.3.1 Testing the Magnitude dependence

In this section we apply the developed algorithm which tests the magnitude dependence of ROC curves to time series of precursory variables $\{v_n(\eta)\}$ and events $\{\chi_n(\eta)\}$ as defined in Eqs. (8.1) and (8.2). The tolerance volumes are therefore either defined as indicated in Eq. (8.4) or according to Eq. (8.6). The results presented in Fig. 8.2 indicate, that the condition $c(\eta, v_n)$ is mostly positive and hence one would expect a positive magnitude dependence.

Since the curves $c(\eta, v_n)$ presented in Fig. 8.2 are very noisy, we apply different methods of post processing $c(\eta, v_n)$, which are discussed in the following.

Counting $\#c(\eta, v_n) > 0$: In order to summarize the results of $c(\eta, v_n)$ one could think about counting for all evaluated values v_n the instances in which the condition $c(\eta, v_n)$ is positive or negative.

The results presented in Fig. 8.3 reveal, that for intermediate values of η we find more instances in which $c(\eta, v_n)$ is positive. However, for very small and very large values of η we find the opposite.

Averaging over η : Another method to summarize the behavior of $c(\eta, v_n)$ consist in averaging over different values of η , i.e., we discuss the mean $\langle c(\eta, v_n) \rangle$ instead of the function $c(\eta, v_n)$. The results graphs in Fig. 8.4 are significantly less noisy, than the original results in Fig. 8.2.

Averaging over v_n : Similarly we can also take the mean over all investigated values of v_n . The results displayed in Fig. 8.5 are also less noisy, than the original results in Fig. 8.4. In total the average over v_n suggests also a positive magnitude dependence of the likelihood ratio.

All results obtained by post processing are – from a theoretical point of view – of course less accurate, than evaluating $c(\eta, v_n)$ directly. However, in situations in which only few data-points are available, they can help to display the overall tendencies of $c(\eta, v_n)$. In total the evaluation of the condition $c(\eta, v_n)$ and its averages over η or v_n suggest a positive magnitude dependence.

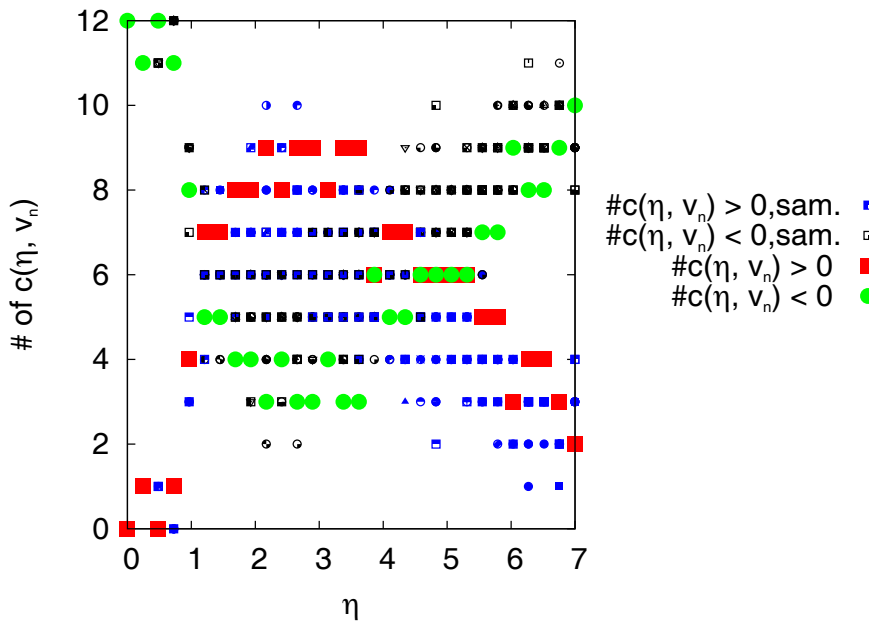


Figure 8.3: This figures shows the instances in which the condition $c(\eta, v_n)$ is positive or negative. The larger symbols in red and green represent the results for the original data set, the smaller symbols in blue and black indicate the results for the resampled data.

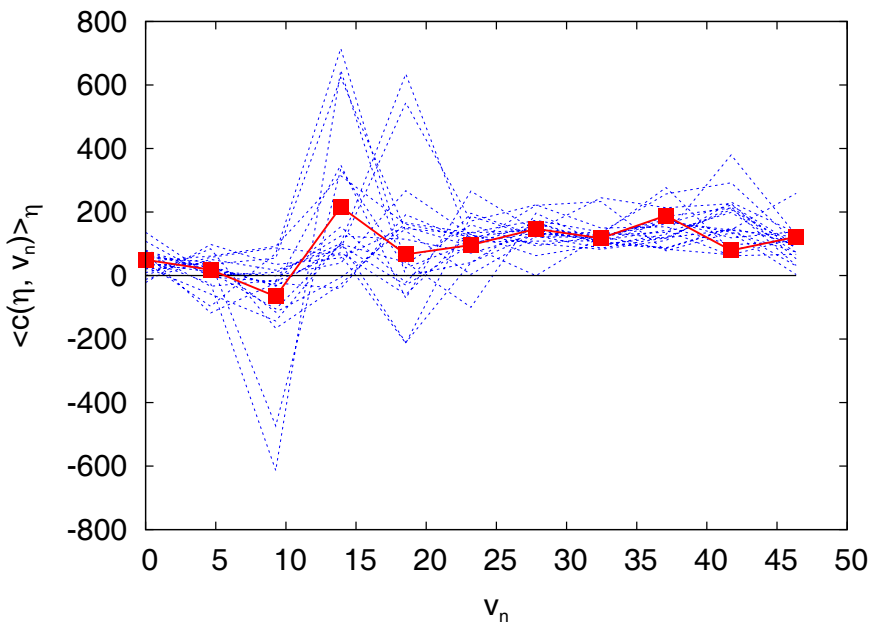


Figure 8.4: The condition $c(\eta, v_n)$ averaged over all 30 investigated values of eta (ranging from 0 K to 6.99 K). Since the mean over all investigated values of η is positive for all values except one, the averaged condition transports the same message than the original condition.

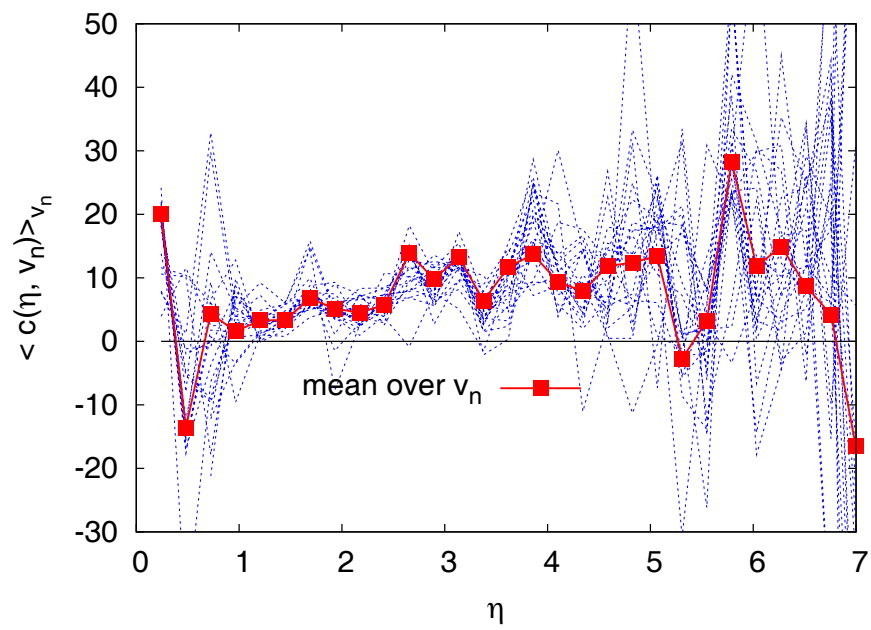


Figure 8.5: The condition $c(\eta, v_n)$, averaged over all investigated values of v_n .

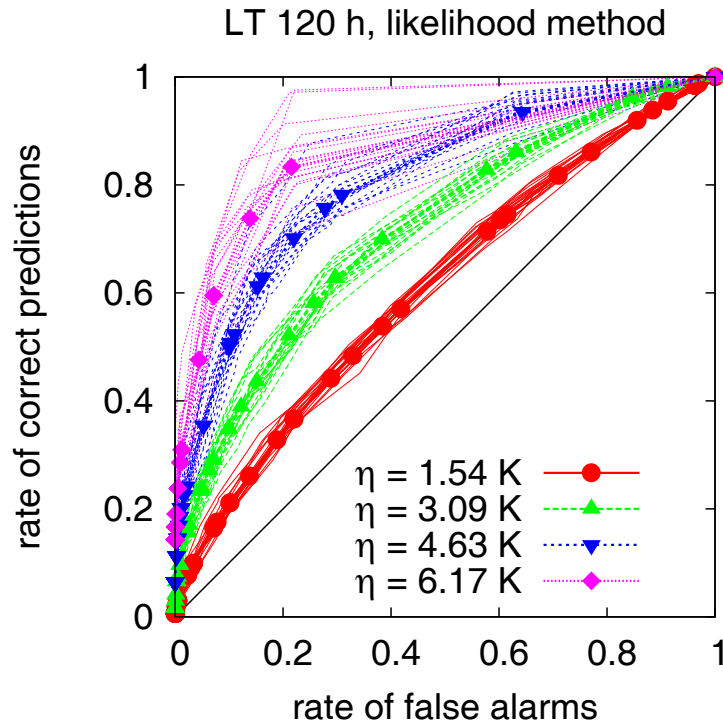


Figure 8.6: ROC curves for the prediction of failures as defined in Eq. (8.1) via the number of ensemble members $v_n(\eta) \in (0, 51)$ for which $|y_n - x_n| \geq \eta$ holds. For these ROC-curves an alarm was given, if we find $v_n(\eta)$ within $I_{like}(\eta, \delta) = [u_{like}(\eta) - \frac{\delta}{2}, u_{like}(\eta) + \frac{\delta}{2}]$, i.e., according to the decision variable defined by Eq. (8.6). The points represent the original data set, the lines denote the results for 20 bootstrap samples, generated by sampling with replacement.

8.3.2 AUC and ROC

In this section we compute ROC curves for the precursors and tolerance intervals defined above. Since our data set is very small (1814 time steps), we cannot concentrate on events which occur more than 1000 times, but we focus on events, which occur more than 20 times. Additionally we generate ROC curves from resampled data sets, in order to get an estimate of the sensitivity of the ROC curves on variations in the distribution of the data. As in the previous studies for the prediction of extreme events in one-dimensional time series, the resampling was done by sampling with repetition from the original data set. The random generator used for the sampling is the *Mersenne twister* [96], as implemented in the *Gnu Scientific Library* [112]. The results are presented in Figs. 8.7-8.8. Surprisingly we find no significant difference in performance between the numerically more expensive maximum likelihood method and our educated guess. This is probably due to the small size of the data set, which leads to a low accuracy in the estimate of the CPDF.

Both sets of ROC-curves and also Fig. 8.8 displaying the area under the ROC curves, reveal a positive magnitude dependence, which is in good agreement with the results for $c(\eta, v_n)$, discussed in the previous section.

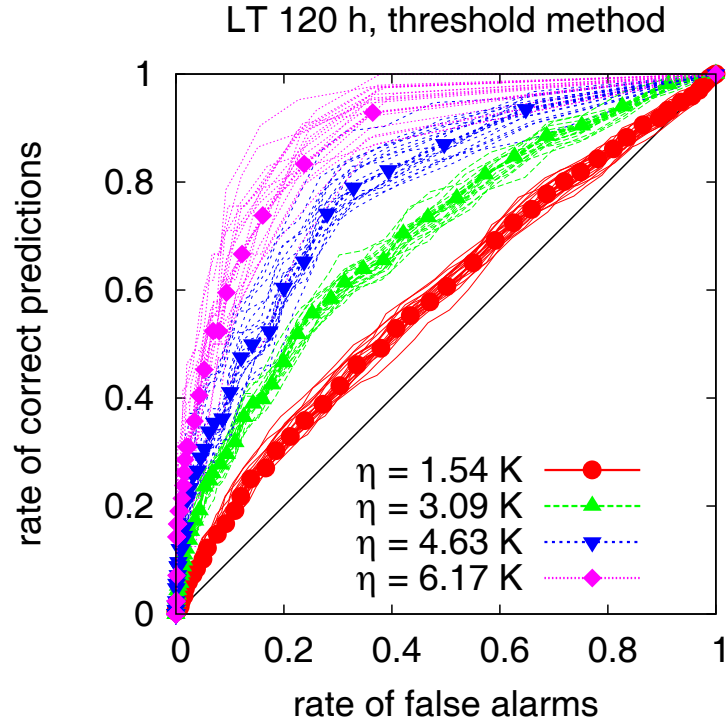


Figure 8.7: ROC curves for the prediction of failures as derived in Eq. (8.1) via the number of ensemble members $v_n(\eta) \in (0, 51)$ for which $|y_n - x_n| \geq \eta$ holds. For these ROC-curves an alarm was given, if we find $v_n(\eta)$ within $I_{thresh}(\eta, \delta) = [v_n(\eta) \geq \delta]$, with $\delta \in (0, 51)$, i.e., according to the decision variable defined by Eq. (8.6). The points represent the original data set, the lines the 20 bootstrap samples.

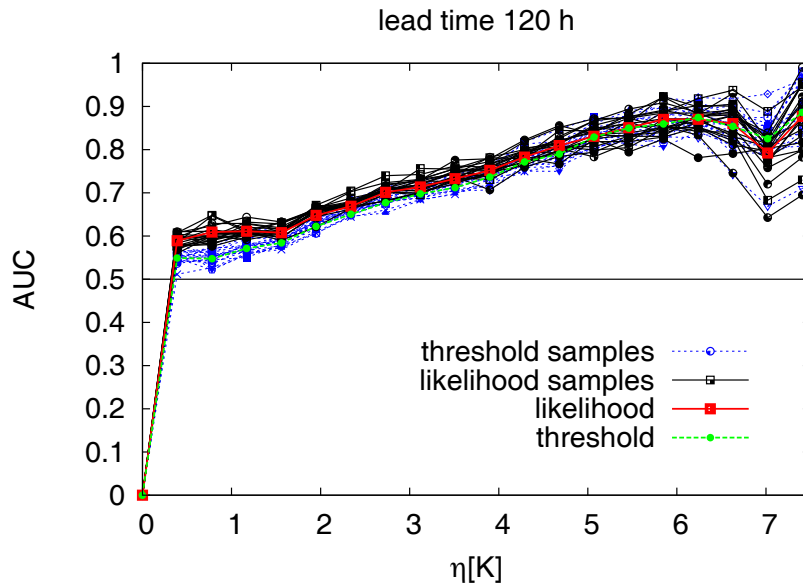


Figure 8.8: The area under the ROC curves for the predictions according to the decision variables in Eqs. (8.4) and (8.6). One can see, that both methods perform similar. Furthermore this figure reveals a positive dependence on the event magnitude.

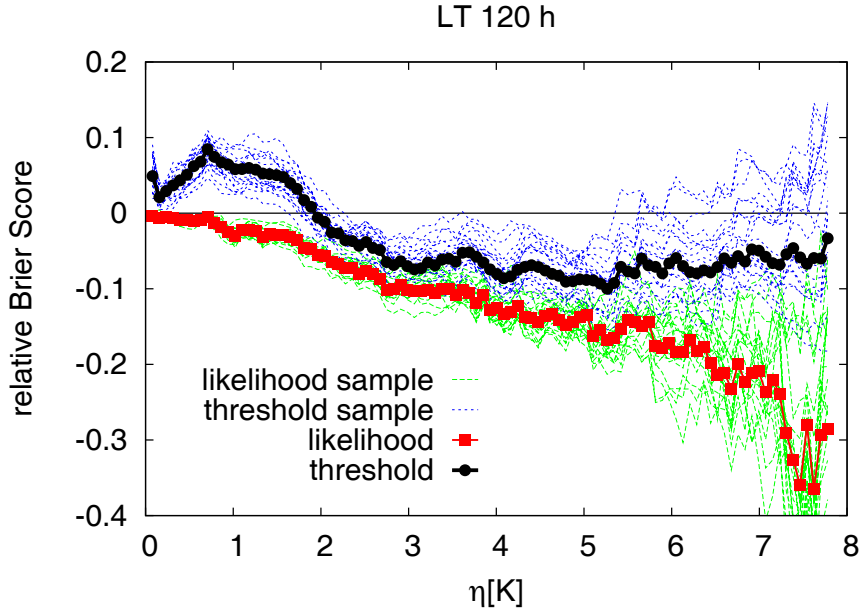


Figure 8.9: The relative Brier Score evaluated for the original data set (results denoted by symbols) and the resampled data sets (results denoted by lines).

8.4 Brier Score and Ignorance

Additionally we evaluate also the relative Brier score and the relative Ignorance, since both scores are common in the framework of weather forecasting. We repeat here the definition of the Brier Score and the ignorance, already presented in Chap. 2. The Brier Score is defined as

$$B(\chi_n(\eta), \hat{p}) = \frac{1}{N} \sum_{i=1}^N (\chi_n(\eta) - \hat{p})^2, \quad (8.7)$$

where \hat{p} denotes a PDF. In our example, \hat{p} is chosen to be the likelihood $p[\chi_n(\eta) = 1|v_n]$ or the relative number of ensemble members $v_n/51$ for which the difference to the high resolution forecast is larger than a given value of η . In order to evaluate the performance of a prediction it is very common to use the relative scores, i.e.,

$$B_{rel}(\chi_n(\eta), \hat{p}, p[\chi_n(\eta) = 1]) = \frac{B[\chi_n(\eta) = 1, p[\chi_n(\eta) = 1]] - B[\chi_n(\eta), \hat{p}]}{B[\chi_n(\eta) = 1, p[\chi_n(\eta) = 1]]}, \quad (8.8)$$

with $B[\chi_n(\eta) = 1, p[\chi_n(\eta) = 1]]$ denoting the Brier score obtained for the total probability $p[\chi_n(\eta) = 1]$ to find events (this probability is often called *climatology* in the context of weather forecasting).

The ignorance score for a binary event is given by

$$Ig[\chi_n(\eta), \hat{p}] = \frac{1}{N} \sum_{i=1}^N -\log(\hat{p}) \cdot \chi_n(\eta) - \log(1 - \hat{p}) \cdot (1 - \chi_n(\eta)), \quad (8.9)$$

where \hat{p} denotes, as above a PDF, which is here either the likelihood $p[\chi_n(\eta) = 1|v_n]$ or the relative number of ensemble members $v_n/51$. Analog to the relative Brier Score, one typically also uses the *relative ignorance*

$$Ig_{rel}(\chi_n(\eta), \hat{p}, p[\chi_n(\eta) = 1]) = \frac{Ig[\chi_n(\eta) = 1, p[\chi_n(\eta) = 1]] - Ig[\chi_n(\eta), \hat{p}]}{Ig[\chi_n(\eta) = 1, p[\chi_n(\eta) = 1]]}, \quad (8.10)$$

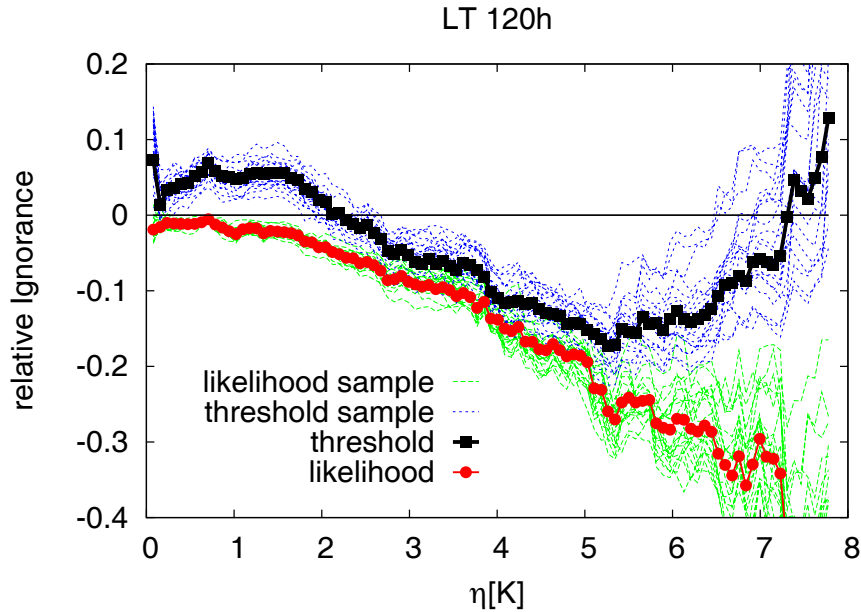


Figure 8.10: The relative Ignorance evaluated for the original data set (results denoted by symbols) and the resampled data sets (results denoted by lines).

in order to measure the quality of a prediction. In order to prevent values of zero or unity obtained for numerical estimates of PDFs, from causing a diverging ignorance, we add two imaginary ensemble members. These two imaginary members have the property that one of them is always counted as a deviation from the high resolution, and the second one is assumed to deviate never from the high resolution forecast. This trick allows us to numerically estimate the ignorance for binary events.

Before we start to discuss the results, it is important to remember that both scores, the Brier Score and the ignorance score indicate a good performance of the prediction method under study, if their values are small.

The results for the relative Brier score and the relative ignorance, evaluated for the likelihood $p[\chi_n(\eta) = 1|v_n]$ and the relative number of deviating ensemble members $v_n(\eta)/51$ are shown in Figs. 8.9 and 8.10. Again we did not only use the original data set, but also resampled data sets in order to obtain an estimate of the sensitivity on small changes in distribution. We can see, that the maximum likelihood approach and the ad hoc construction of the decision variable lead to very similar results.

In total, the results for the ignorance and the Brier score show a better predictability, i.e., smaller values for the prediction of larger events, as the ROC curves.

However, we learned in Chap. 4 that the magnitude dependence of one measure for the quality of a prediction does not necessarily indicate that other measures have the same magnitude dependence. Hence, we should beware of interpreting the results for the ignorance and the Brier score to be in agreement with the magnitude dependence obtained for the ROC curve. Each quality measure has its own magnitude dependence and even if evaluated for the same prediction problem, they do not necessarily agree, although this is the case in this example studied.

8.5 Summary

We use information obtained from the ensemble forecast and the high resolution forecast to successfully predict errors of the high resolution forecast. Comparing the quality of the predictions we find no significant difference in performing between the “ad hoc choice” of the alarm volume and an alarm volume constructed via the numerically more expensive maximum likelihood ap-

proach. This result is probably due to the limited amount of available data, which leads to a suboptimal estimation of the likelihood.

In addition the ROC curves, the Brier score and the likelihood characterizing our prediction experiment show a positive magnitude dependence. The positive magnitude dependence of the ROC curves was also reasonable well reflected by the test condition for the magnitude dependence of ROC curve and likelihood ratio, introduced in Chap. 4. Furthermore, we saw, that in situations in which a limited amount of data does not allow a sufficiently good evaluation of the test condition, we can use the averages of the test condition in order to obtain information about the magnitude dependence.

Chapter 9

Conclusions and open Questions

In this thesis we access the prediction of extreme events from a point of view which does not focus on a specific system generating large impact events, but on the common features of events occurring in very different circumstances as completely different phenomena. We therefore study predictions made by identifying precursory structures in time series, which were identified using a maximum likelihood approach.

The main goal of this thesis is to investigate the dependence of the quality of a prediction on the magnitude of the events under study. Until now, this dependence was only sporadically reported for different phenomena [20, 21, 22, 23] without being understood as a general feature of predictions. We propose the *magnitude dependence* as a property of a prediction, indicating, whether larger events can be better, harder or equally well predicted than smaller events. Furthermore, we specify a condition which can characterize the magnitude dependence of a distinguished measure for the quality of a prediction, namely for the Receiver Operator Characteristic curve (ROC curve). We consider the ROC curve a suitable measure for the performance of predictions of extreme events since it is independent of the relative frequency of the events under study, and does not average, which would hide the contributions of rare events between results for the predictions of smaller events. The specified test condition for the magnitude dependence allows to predict the magnitude dependence of a prediction via studying the joint PDF of events and precursory variables. If we are able to describe the numerical estimate of this joint PDF by an analytic expression, we can not only characterize the magnitude dependence of events observed so far, but infer the magnitude dependence of events larger than the observed ones.

Having the test condition specified, we study the magnitude dependence for the prediction of increments and threshold crossings in sequences of random variables and short- and long-range correlated stochastic processes. Furthermore we investigate the dependence on the event magnitude for the prediction of increments and threshold crossings in velocity increments, measured in a free jet flow and in wind-speed measurements. In dependence of the distributions of the processes under study we obtain positive, negative or zero magnitude dependences.

Additionally we introduce a method of post-processing the output of ensemble weather forecast models in order to identify precursory behavior, which could indicate failures of weather forecasts. We then study not only the success of this method, but also the above referred magnitude dependence.

To summarize, we propose that the magnitude dependence sporadically reported for predictions in different applications is a property inherent to any prediction of a binary event. If the quality of a prediction is measured by ROC curves, the magnitude dependence is fully characterized by the definition of the events under study and the joint PDF of precursor and event.

9.1 List of specific results

The main novel results presented in this thesis are:

- We propose a definition for extreme events in time series, which can serve in order to transfer the term “*extreme event*” into a scientific concept.
- We propose a method of obtaining the joint PDFs of precursor and event analytically via using the Heaviside function as a filter.

In chapter 3:

- Increments in an AR(1) correlated process are the better predictable the larger they are. This result can be obtained both, analytically and numerically.
- Increments in long-range correlated ARMA processes are also the better predictable, the larger they are.
- The magnitude dependence is not explicitly influenced by the correlation of the data, but only by the joint PDF of event and precursor. The influence of the correlation is only of relevance in terms of its effect on this joint PDF.

In chapter 4:

- We define the magnitude dependence of a prediction task as a quantity which is dependent on the process under study, the definition of the events and the measure for the quality of a prediction.
- We derive a test condition for the magnitude dependence of families of likelihood ratios and ROC curves. This condition allows to trace back the magnitude dependence to properties of the relative frequency of events and the likelihood.

In chapter 5:

- For the prediction of increments in a sequence of Gaussian distributed random numbers we obtain a positive magnitude dependence. The corresponding statement of the numerically and analytically evaluated test-condition was verified by the ROC curves.
- For the prediction of increments in a sequence of one-sided exponentially distributed random numbers larger events are harder to predict, i.e, for the optimal precursor we find a negative magnitude dependence. The corresponding statement of the numerically and analytically evaluated test-condition was verified by the respective ROC curves.
- For the prediction of increments in a sequence of symmetrized exponentially distributed random numbers, we obtain an independence of the event magnitude from the test condition. This result is reflected by the first part of the ROC curves, corresponding to a low rate of false alarms. However, the ROC curves show a positive magnitude dependence for larger false alarm rates. This dependence is significantly less pronounced than in the Gaussian case.
- For the prediction of increments in sequences of one-sided power-law distributed random numbers with finite variance, both the test condition and the ROC curve indicate a negative magnitude dependence. This result is independent of the exponent of the power-law, as long as the exponent guarantees a finite variance.

- For the prediction of increments in symmetrized power-law distributed random numbers, the test condition and the ROC curve display a negative magnitude dependence.
- Furthermore we find an “universal”¹ ROC curve for the prediction of increments of magnitude zero, i.e., predictions of the sign of an increment.

In chapter 6:

- For the prediction of threshold crossings in AR(1) correlated processes with Gaussian, asymptotically exponential and power-law distributions we find a positive magnitude dependence.
- If we restrict the set of events to threshold crossings, which are not preceded by a precursor which is an event itself, we find a slightly reduced performance, but also a positive magnitude dependence.
- For predictions of threshold crossings of magnitude zero, i.e, predicting the sign of the next value in the time series, we find again ROC curves which are independent of the distribution of the process under study.

In chapter 7:

- The qualitative results obtained for the predictions of increments and threshold crossings in synthetical time series can be retrieved by predicting in measured time series, containing noise and long-range correlations. The corresponding time series, namely velocity increments measured in a free jet flow experiment and increments of wind speed measurements are in dependence on the definition of the increment either exponentially or Gaussian distributed.
- Specifying the prediction horizon in a suitable way, we find again a positive magnitude dependence in the exponential and the Gaussian case for both increments and threshold crossings.
- For the prediction of increments we can also observe, that the magnitude dependence for the prediction of exponentially distributed increments is significantly weaker than the magnitude dependence of the corresponding Gaussian case.

In chapter 8:

- We successfully predict failures of weather forecasts by using a post-processing method, which transfers deviations within a model based ensemble forecasts system into precursors for failures.
- For relatively short data sets, the maximum likelihood approach to identify suitable precursors performs not better than a numerically less expensive “educated guess” of suitable alarm intervals, which is specified in more detail in Sec. 8.3.
- We observe also for this example a positive magnitude dependence, indicated by the test condition, the ROC curves, the Brier score and the ignorance.
- We find that for short data sets, one can also use averages of the test condition in order to extract the information provided by the test condition itself.

¹The term universal is here used to indicate the independence on the underlying distribution of the process.

9.2 Specific open Questions

We studied the magnitude dependence for predictions within Gaussian, exponentially or power-law distributed data sets and with respect to the ROC curve as a measure for the quality of a prediction. It is important to obtain information of the magnitude dependence in order to judge the quality of a forecast in a reasonable way. Especially, one would like to know if it is possible to extrapolate the performance of a forecast, which is able to detect smaller events, to the performance of larger events of unknown magnitudes.

Starting from the results obtained in this thesis, one could now investigate the magnitude dependence for predictions in data following arbitrary other distributions. It is of particular interest to study the magnitude dependence of data following a Levy stable distribution or a truncated Levy stable distribution. Due to their stability and their heavy tails which enforce the occurrence of extreme events, these distributions are of particular interest and inferring their magnitude dependence is thus highly desirable. Although their tails are also asymptotically power-law distributed, it is not obvious that the results obtained for power-laws with finite variance hold also for power-laws with diverging variance, as they are present in Levy stable distributions.

Furthermore, the concept of the magnitude dependence can be transferred to other measures of predictability. In other words, it would be useful to formulate test conditions for the Brier score, the ignorance, the Kullback Leibler distance and many other measures for the quality of a prediction. It is especially important to relate the results obtained for the ROC curves to predictions evaluated with the so called “error-curve”, which is a well established concept to evaluate the success of earthquake predictions. Since earthquake predictions are made under the assumption of a continuous time process, the error-curve does relate the success of a prediction to the total alarm time. However, when applied to discrete time series and for fixed alarm times, the error-curve can be transformed into a ROC curve. In other words, the ROC curve relates to a special case of the error curve. Hence it would be of special interest to investigate, whether the results for the magnitude dependence of the ROC curve can be generalized.

Leaving the field of statistical inference one can also investigate the magnitude dependence of model based forecasts. It is already empirical knowledge [24] that large scale structures in the atmosphere are more stable and thus better predictable. Using the ensemble spread of the weather forecasts studied in Chap. 8 was a very coarse grained approach in this direction. A more specific approach is discussed under the name *prediction of predictability* [35, 113] which evaluates localized Lyapunov exponents in order to distinguish regions in phase space, which allow a high predictability from those in which infinitesimal uncertainties lead to exponentially diverging trajectories.

9.3 Open Questions on Extreme Events and Outlook

Thinking about extreme events as a general phenomenon occurring in systems possessing complex dynamics, there are very few things which are known and specified yet. It is not even clear, whether one can call the different approaches a field of research yet. Thus, one can start with the naive but reasonable question “What is an extreme event?” and realize that there is no generally accepted definition yet. Although many systems in our surrounding display from time to time behavior that we consider to be extreme, one can always ask whether this classification is not simply the source of our specific reception. Regarding the development of the earth in total, rare disastrous events like hurricanes, droughts or floods are not extreme but perfectly normal. Following this idea, one could also say that events we call extreme, are simply events we are not adapted to, neither by evolution, nor by the way we created our habitats nor by the way our economies work. However, we probably do not want to rely on evolution to enable future generations to deal better with the expected more frequent and more intense [3] extremes

of the future climate.²

If we consequently try to access extreme events by scientific methods, another important question is to ask “Are extreme events like smaller events, only larger?” [114], i.e., can we simply extrapolate from our knowledge about the less rare events observed in a system. In this thesis we implicitly assumed that the answer to the above question is yes, since we always expected the events under study to originate from the same process. Most research on extreme value statistics also uses this assumption. Despite the fact that extreme value statistics for uncorrelated random numbers was already developed in the 1940 and 1950s, there is still no established concept that allows to apply the extremal types theorem to correlated data without making restrictive assumptions about the correlation [115]. However, some authors claim to be able to successfully estimate extreme values distributions on long-range correlated data [116], while others explicitly warn not to do so [117, 118]. Since most of the interesting data sets are indeed correlated, it is highly desirable to explore under which circumstance one can still successfully estimate the generalized extreme value distributions.

In case, that we answer the question “Are extreme events like smaller events, only larger?” [114] with “no”, we can think about extreme events and “normal events” as phenomena ranging from two (or more) different types of dynamics, whose superposition is then perceived as a system displaying rare extreme events. A similar approach is to separate between short range correlations, which determine the part of the dynamics, which is understood and predictable and the influence of long-range correlations (often hidden by the finite size of the data set), which influence the occurrence of extreme events [6].

If one does not demand the events under study to be endogenous, one can also consider all sorts of couplings between different subsystems, one of them being the subsystem in which we observe the extreme event. The generation of the extreme event is then the result of the exogenous driving [30].

Furthermore one could also interpret extreme events and “normal events” as intermittent behavior [119], especially if the extreme events occur in clusters.

The irregular occurrence of extreme events can be related with various model classes possessing complex dynamics, e.g., non-linear maps and differential equations, systems displaying self organized criticality, cellular automata, networks and models for stochastic resonances and others. Much research was already done in the last decades to study properties of these systems. However, it might be worth revisiting them and focus now on their ability to generate extreme events, their statistical properties, e.g., their extreme value distributions, the precise mechanism of the event generation, or precursory behavior which can help to predict them. A promising approach in this direction is the work of Nicolis and Nicolis [18] who studied how different deterministic non-linear dynamics are reflected by the PDFs and CPDFs. Their results indicate that one cannot obtain a simple limiting behavior in the sense of a generalized extreme value distributions for systems displaying deterministic chaos.

Since many of the models listed above are indeed applied in order to describe natural phenomena, such as earthquakes [120], avalanches or forest fires, it will hopefully be possible to relate the knowledge gathered from extreme events occurring in the model systems to real world phenomena.

²Maybe one could also think of research and its application as a part of evolution.

Appendix A

Increments and threshold crossings in ARMA processes

We show how to relate the results obtained using the definition of extreme events as extreme increments ($x_{n+1} - x_n \geq d$, as in Eq. (5.1)) to the case when extreme events are defined as extreme values ($y_{n+1} \geq d$, as in Eq. (5.1)), for ARMA(p,q) processes. An ARMA(p,q) model is defined as [33]

$$\Phi(B)x_n = \theta(B)\xi_n, \quad (\text{A.1})$$

where $\{\xi\}$ correspond to white noise and

$$\begin{aligned} \Phi(B) &= 1 - \Phi_1 B - \Phi_2 B^2 - \dots - \Phi_p B^p, \\ \theta(B) &= 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q, \end{aligned}$$

with $B^j x_n = x_{n-j}$. Searching for extreme increments in a time series $\{x\}$ is equivalent to search for extreme values in the time series $\{y\}$, defined through the transformation

$$y_{n+1} = x_{n+1} - x_n. \quad (\text{A.2})$$

Assuming that $\{x\}$ is described by an ARMA(p,q) process defined by Eq. (A.1), and inserting Eq. (A.2) in Eq. (A.1), one obtains that $\{y\}$ is described by an ARMA(p,q+1) model with the following transformed coefficients

$$\begin{aligned} \Phi_i^\dagger &= \Phi_i \quad i = 1, 2, \dots, p \quad , \\ \theta_i^\dagger &= \theta_i - \theta_{i-1} \quad i = 1, 2, \dots, q \quad , \\ \theta_{q+1}^\dagger &= \theta_q \quad . \end{aligned} \quad (\text{A.3})$$

Due to the transformation (A.2) the precursory structure equivalent to the one used in Sec. 3 is obtained choosing¹

$$y_{pre} = \sum_{j=0}^n y_j - x_0 = x_n. \quad (\text{A.4})$$

With this choice and the corresponding transformation of the process (Eq.(A.2)), the results obtained for extreme increments can be transferred to the case of extreme values. In particular, for the case of AR(1) processes (which corresponds to an ARMA(1,0)) discussed in Chap. 3, all results are also valid for an ARMA(1,1) process with the precursor given by Eq. (A.4) and events defined as threshold crossings.

¹We assume $x_0 = 0$, which is the mean value of $\{x\}$. This assumption is irrelevant for large values of n .

Appendix B

Evaluating the CPDFs of precursors and events

In this section we will discuss how the CPDFs of precursor and event can be obtained analytically. This more technical issue might not fit in the context of a more general background. In order to obtain the CPDFs we are interested in, we use both time series, the original data set $\{x_n\}$ and the tracer time series $\{\chi_n\}$ to construct the joint probabilities $p(\chi_n, x_0, x_{-1}, \dots, x_{1-\tau})$ which contain all dependencies between the sequence of observations down to τ temporal steps into the past. Numerically these joint probabilities can be estimated by simply counting the coincidences of event and precursory variables and then normalizing with the total number of events.

We can additionally obtain the joint PDFs of event and precursors analytically by applying the Heaviside function (see e.g. [121]) $\Theta(x_{n+1} - x_n - d)$ as a filter to the PDF of the stochastic process under investigation.

Prediction of increments in the next time step

The following considerations are made for increments which occur within one time step, i.e. $\kappa = 1$ and a one-dimensional precursory variable, i.e., $\tau = 1$ and $\mathbf{s}_n = x_n$. In order to simplify the notation in this section, we refer to the magnitude of the increment not by the relative magnitude $\eta = (x_{n+1} - x_n)/\sigma$, measured in units of the standard deviation, but in terms of the absolute magnitude $d = x_{n+1} - x_n$.

Since only the time steps (x_n, x_{n+1}) are of relevance for the filtering, we can neglect all previous time steps and apply our filter simply to the joint PDF for (x_n, x_{n+1}) , which has the form $\rho_j(x_n, x_{n+1}) = \rho_n(x_n)\rho_{n+1}(x_{n+1}|x_n)$. This implies that we can regard all previous time-steps x_0, x_1, \dots, x_{n-1} , on which ρ_n and ρ_{n+1} might depend, as parameters. In a broader context, this corresponds to the fact, that due to the conditioning on a finite number of previous time steps, we implicitly treat the process under study as a Markov process, also described in Sec. 2.2.1. The joint PDF of the extreme events $p_n[x_n, \chi_n(d) = 1]$ can then be obtained by multiplication with $\Theta(x_{n+1} - x_n - d)$.

$$p[x_n, \chi_n(d) = 1] = \int_{-\infty}^{\infty} dx_{n+1} \Theta(x_{n+1} - x_n - d) \rho_j(x_n, x_{n+1}) \quad (\text{B.1})$$

$$= \rho(x_n) \int_{-\infty}^{\infty} dx_{n+1} \Theta(x_{n+1} - x_n - d) \rho_j(x_{n+1}|x_n) \quad (\text{B.2})$$

Note, that we integrate over x_{n+1} since our event does not consist in observing a distinguished value of x_{n+1} , but in observing x_{n+1} to be in a certain range of possible values, namely

$$\chi_n = 1 \quad \text{iff} \quad x_{n+1} \geq x_n + d, \quad (\text{B.3})$$

according to the definition of increments in Sec. 2.1.2. This is also the reason, why the joint PDF of precursory variable and event is a probability density in terms of the precursory variable and a probability in terms of the event.

Continuing with our calculation, we substitute $x_{n+1} = x_n + d + \gamma$ and thereby introduce a new random variable $\gamma \in \mathbb{R}$.

$$p[x_n, \chi_n(d) = 1] = \rho(x_n) \int_{-\infty}^{\infty} d\gamma \Theta(\gamma) \rho(x_n + d + \gamma | x_n) \quad (\text{B.4})$$

$$= \rho(x_n) \int_0^{\infty} d\gamma \rho(x_n + d + \gamma | x_n). \quad (\text{B.5})$$

By normalizing with the total probability $p[\chi(d) = 1]$ to find extreme events of size d or larger

$$p[\chi(d) = 1] = \int_{-\infty}^{\infty} dx_n p[x_n, \chi_n(d) = 1], \quad (\text{B.6})$$

we obtain the posterior PDF and the likelihood by using Bayes' theorem:

$$\rho[x_n | \chi_n(d) = 1] = \frac{\rho_n(x_n)}{p[\chi(d) = 1]} \int_0^{\infty} d\gamma \rho(x_n + d + \gamma | x_n); \quad (\text{B.7})$$

$$p[\chi_n(d) = 1 | x_n] = \int_0^{\infty} d\gamma \rho(x_n + d + \gamma | x_n). \quad (\text{B.8})$$

Analogously we can calculate the corresponding PDFs and CPDFs for non-events,

$$p[x_n, \chi_n(d) = 0] = \rho(x_n) \int_{-\infty}^{\infty} dx_{n+1} (1 - \Theta(x_{n+1} - x_n - d)) \rho_j(x_{n+1} | x_n) \quad (\text{B.9})$$

$$= \rho(x_n) \int_{-\infty}^{\infty} d\gamma (1 - \Theta(\gamma)) \rho(x_n + d + \gamma | x_n) \quad (\text{B.10})$$

$$= \rho(x_n) \int_{-\infty}^{\infty} d\gamma \Theta(-\gamma) \rho(x_n + d + \gamma | x_n) \quad (\text{B.11})$$

$$= \rho(x_n) \int_{-\infty}^0 d\gamma \rho(x_n + d + \gamma | x_n), \quad (\text{B.12})$$

where we used the reflection rule $(1 - \Theta(x - a)) = \Theta(a - x)$. By normalizing with the total probability $p[\chi(d) = 0]$ to find non-events

$$p[\chi(d) = 0] = \int_{-\infty}^{\infty} dx_n p[x_n, \chi_n(d) = 0], \quad (\text{B.13})$$

we obtain the posterior PDF and the likelihood by using Bayes theorem.

$$\rho[x_n | \chi_n(d) = 0] = \frac{\rho_n(x_n)}{p[\chi(d) = 0]} \int_{-\infty}^0 d\gamma \rho(x_n + d + \gamma | x_n); \quad (\text{B.14})$$

$$p[\chi_n(d) = 0 | x_n] = \int_{-\infty}^0 d\gamma \rho(x_n + d + \gamma | x_n). \quad (\text{B.15})$$

Prediction of threshold crossings in the next time step

If we are interested in evaluating the corresponding PDFs and CPDFs not for increments but for threshold crossings as defined in Sec. 2.1.2, which occur in the next time step x_{n+1} we simply have to set $x_n = 0$ in the terms which are relevant for the filtering. This means, that we apply the Heaviside function $\Theta(x_{n+1} - d)$ and substitute with $x_{n+1} = d + \gamma$. Hence the joint PDF of event defined as threshold crossings and the precursory variable x_n reads

$$p[x_n, \chi_n(d) = 1] = \rho(x_n) \int_{-\infty}^{\infty} d\gamma \Theta(\gamma) \rho(d + \gamma | x_n) \quad (\text{B.16})$$

$$= \rho(x_n) \int_0^{\infty} d\gamma \rho(d + \gamma | x_n). \quad (\text{B.17})$$

The total PDF to observe /not to observe a threshold crossing of a given size d , the posterior PDFs and the likelihoods can be obtained analogously to the corresponding PDFs and CPDFs for increments (see Eqs. (B.6-B.15)).

Prediction of increments κ time steps into the future

In principle one can also consider predictions with a larger prediction horizon, $\kappa > 1$, if the underlying process is a Markov process, i.e., we require, that the Chapman-Kolmogorov Equation holds,

$$\begin{aligned} \rho(x_n, x_{n+1}, \dots, x_{n+\kappa}) &= \rho(x_{n+\kappa}|x_n)\rho(x_n), \\ \text{with } \rho(x_{n+\kappa}|x_n) &= \int dx_{n+\kappa-1} \dots dx_n \rho(x_{n+\kappa}|x_{n+\kappa-1})\rho(x_{n+\kappa-1}|x_{n+\kappa-2})\dots\rho(x_{n+1}|x_n). \end{aligned} \quad (\text{B.18})$$

The joint PDF of precursor x_n and the event $\chi_{n+\kappa} = 1$ *iff* $x_{n+\kappa} - x_n \geq d$ is then given by

$$\begin{aligned} \rho(x_n, x_{n+1}, \dots, \chi_{n+\kappa}) &= \rho(x_n) \int_0^\infty d\gamma \rho(x_n + d + \gamma|x_n), \\ \text{with } \rho(x_n + d + \gamma|x_n) &= \int dx_{n+\kappa-1} \dots dx_n \rho(x_n + d + \gamma|x_{n+\kappa-1}) \times \\ &\quad \times \rho(x_{n+\kappa-1}|x_{n+\kappa-2})\dots\rho(x_{n+1}|x_n). \end{aligned} \quad (\text{B.19})$$

However, this later example practically useless, since it makes no sense to predict κ time steps ahead in a process, which has only a memory of one time step into the past. In order to expect a certain success of the prediction it is more suitable to consider a generalized Markov process which is at least of order κ . This implies

$$\rho(x_n, x_{n-1}, \dots, x_0) = \rho(x_n, x_{n-1}, \dots, x_{n-\kappa+1}). \quad (\text{B.20})$$

Then we can analogously express the joint pdf of precursor x_n and the event $\chi_{n+\kappa} = 1$ *iff* $x_{n+\kappa} - x_n \geq d$ by

$$\rho(x_n, x_{n+1}, \dots, \chi_{n+\kappa}) = \rho(x_n, x_{n+1}, \dots, x_{n+\kappa-1}) \int_0^\infty d\gamma \rho(x_n + d + \gamma|x_{n+\kappa-1}, \dots, x_n), \quad (\text{B.21})$$

$$= \rho(\mathbf{s}_{n+\kappa}) \int_0^\infty d\gamma \rho(x_n + d + \gamma|\mathbf{s}_{n+\kappa}), \quad (\text{B.22})$$

$$(\text{B.23})$$

if we choose $\tau = \kappa$. The corresponding expressions for the prediction of threshold crossings can be formulated analogously. Note however, that although we can easily formulate Eq. (B.21) it is difficult to obtain for a given problem due to the multiple integration.

Bibliography

- [1] A. B. Davis and D. Sornette, Celebrating the Physics in Geophysics, EOS, Transactions of the American Geophysical Union **86**, 461 (2005).
- [2] D. Sornette, Predictability of catastrophic events: material rupture, earthquakes, turbulence, financial crashes and human birth, in *Proceedings of the National Academy of Sciences USA*, volume 99, pages 2522–2529, 02 2002.
- [3] Intergovernmental Panel on Climate Change, *Climate Change 2007 - The Physical Science Basis*, Cambridge University Press, 2007.
- [4] P. Kaporis, J. Polygiannakis, X. Li, X. Yao, and K. A. Eftaxias, Similarities in precursory features in seismic shocks and epileptic seizures, Europhysics Letters **69**, 657–663 (2005).
- [5] S. Nikolopoulos, P. Kaporis, K. Karamanos, and K. Eftaxias, A unified approach of catastrophic events, Natural Hazards and Earth System Sciences **4**, 615–631 (2004).
- [6] A. Bunde, J. F. Eichner, S. Havlin, and J. W. Kantelhardt, The effect of long-term correlations on the return periods of rare events, Physica A **330**, 1–7 (2003).
- [7] D. L. Turcotte, B. D. Malamud, F. Guzzetti, and P. Reichenbach, Self-organization, the cascade model, and natural hazards, PNAS **99**, 2530–2537 (2002).
- [8] C. Nicolis, V. Balakrishnan, and N. G., Extreme Events in Deterministic Dynamical Systems, Physical Review Letters **97**, 2006 (2006).
- [9] P. Friederichs and A. Hense, Statistical Downscaling of Extreme Precipitation Events Using Censored Quantile Regression, Monthly Weather Review **135**, 2365 (2007).
- [10] E. J. Gumbel, Les valeurs extrêmes des distributions statistiques, Ann. Inst. H. Poincaré **5**, 115–158 (1935).
- [11] B. V. Gnedenko, Sur la distribution limite du terme maximum d’une serie aleatoire, Annals of Mathematics **44**, 423–453 (1943).
- [12] E. J. Gumbel, *Statistics of extremes*, Columbia University Press, 1958.
- [13] L. de Haan, Residual life time at great age, Annals of Probability **2**, 792–804 (1972).
- [14] J. Pickands, Statistical inference using extreme order statistics, Annals of Statistics **3**, 119–131 (1975).
- [15] R. S. Ellis, *Entropy, large deviations, and statistical mechanics*, Springer, 1985.
- [16] U. Frisch and D. Sornette, Extreme deviations and applications, J. Phys. I France **7**, 1155–1171 (1997).
- [17] E. G. Altmann and H. Kantz, Recurrence time analysis, long-term correlations, and extreme events, Phys. Rev. E **71**, 056106 (2005).

- [18] C. Nicolis and S. Nicolis, Return time statistics of extreme events in deterministic dynamical systems, *European Physics Letter* **80**, 40003 (2007).
- [19] V. I. Keilis-Borok and V. G. Kossobokov, Premonitory activation of earthquake flow: algorithm M8, *Physics of the Earth and Planetary Interiors* **61**, 73–83 (1990).
- [20] H. Kantz, D. Holstein, M. Ragwitz, and N. K. Vitanov, Markov chain model for turbulent wind speed data, *Physica A* **342**, 315–321 (2004).
- [21] M. Göber, C. A. Wilson, S. F. Milton, and D. B. Stephenson, Fairplay in the verification of operational quantitative precipitation forecasts, *Journal of Hydrology* **288**, 225–236 (2004).
- [22] D. Lamper, S. D. Howison, and N. F. Johnson, Predictability of Large Future Changes in a Competitive Evolving Population, *Phys. Rev. Lett* **88**, 017902 (2002).
- [23] M. G. Shapoval, A. B. and Shnirman, How size of target avalanches influences prediction efficiency, *International Journal of Modern Physics C* **17**(12), 1777–1790 (2006).
- [24] A. Persson and F. Grazzini, *ECMWF User guide to Forecast Products*, ECMWF, 2005.
- [25] H. Kantz, A. G. Eduardo, S. Hallerberg, D. Holstein, and A. Riegert, *Dynamical Interpretation of Extreme Events: Predictability and Predictions*, chapter 4, pages 69–93, Springer, 2006.
- [26] S. Hallerberg, E. G. Altmann, D. Holstein, and H. Kantz, Precursors of Extreme Increments, *Phys Rev. E* **75**, 016706 (2007).
- [27] S. Hallerberg and H. Kantz, Influence of the Event Magnitude on the Predictability of Extreme Events, *Phys. Rev. E* **77**, 011108 (2008).
- [28] S. Hallerberg and H. Kantz, How Does the Quality of a Prediction depend on the Magnitude of the Events under Study?, *Nonlin. Processes Geophys.* **15**, 321–331 (2008).
- [29] S. Hallerberg, J. Bröcker, and H. Kantz, *Nonlinear Time Series Analysis in the Geosciences*, volume 112 of *Springer Series: Lecture Notes in Earth Sciences*, chapter Prediction of Extreme Events, Springer, 2008, (in Press).
- [30] D. Sornette, *Endogenous versus Exogenous Origins of Crises*, chapter 5, pages 95–116, Springer, 2006.
- [31] J. A. Schumpeter, *Business Cycles: A Theoretical, Historical and Statistical Analysis of the Capitalist Process*, Mc Graw-Hill, New York, 1939.
- [32] D. Sornette and A. Johansen, Significance of log-periodic precursors to financial crashes, *Quant. Finance* **1**, 452–471 (2001).
- [33] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis*, Prentice Hall, 1994.
- [34] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer, 1998.
- [35] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, 1997.
- [36] H. Kantz, D. Holstein, M. Ragwitz, and N. K. Vitanov, Short time prediction of wind speeds from local measurements, in *Wind Energy – Proceedings of the EUROMECH Colloquium*, edited by J. Peinke, P. Schaumann, and S. Barth, Springer, 2006.

- [37] N. Vanderwalle, M. Ausloos, P. Boveroux, and A. Minguet, How the financial crash of October 1997 could have been predicted, *European Physical Journal* **B 4**, 139–141 (1998).
- [38] W. Paul and J. Baschnagel, *Stochastic Processes From Physics to Finance*, Springer, 2000.
- [39] E. Montroll and M. Shlesinger, *Nonequilibrium Phenomena II: From Stochastics to Hydrodynamics*, chapter On the Wonderful World of Random Walks, Elsevier Science Publishers, 1984.
- [40] R. Abel, Die Bedeutung der Ratten und Flöhe für die Verbreitung der Bubonenpest, *Medical Microbiology and Immunology* **48**, 89–119 (1904).
- [41] M. Ikeya, *Earthquakes and Animals: From Folk Legends to Science*, World Scientific Publishing Company, 2005.
- [42] F. Mulargia, Retrospective validation of the time association of precursors, *Geophys. J. Int.* **131**, 500–504 (1997).
- [43] P. Biaggi, R. Piccolo, A. Ermini, F. Y., S. Kingsley, Y. Khatkevich, and E. Gordeev, Hydrogeochemical precursors of strong earthquakes in Kamchatka: further analysis, *Natural Hazards and Earth System Sciences* **1**, 9–14 (2001).
- [44] K. Karamanos, D. Dakopoulos, K. Aloupis, A. Peratzakis, L. Athanasopoulou, S. Nikolopoulos, P. P. Kapiris, and K. Eftaxias, Preseismic electromagnetic signals in terms of complexity, *Phys. Rev. E* **74**, 016104 (2006).
- [45] B. Litt and J. Echauz, Prediction of epileptic seizures, *The Lancett Neurology* **1**, 22–30 (2002).
- [46] F. Mormann, T. Kreuz, R. G. Andrzejak, A. Kraskov, P. David, C. E. Elger, and K. Lehnertz, Epileptic seizures are preceded by a decrease in synchronization, *Epilepsy Res.* **53**, 173–185 (2003).
- [47] R. Aschenbrenner-Scheibe, T. Maiwald, M. Winterhalder, H. U. Voss, J. Timmer, and A. Schulze-Bonhage, How well can epileptic seizures be predicted? An evaluation of a nonlinear method, *Brain* **126**, 2616–2626 (2003).
- [48] R. J. Geller, Earthquake prediction: a critical review, *Geophysical Journal International* **131**, 425–450 (1997).
- [49] F. Mormann, T. Kreuz, C. Rieke, R. G. Andrzejak, A. Kraskov, P. David, C. E. Elger, and K. Lehnertz, On the predictability of epileptic seizures, *Clinical Neurophysiology* **116**, 569–587 (2005).
- [50] B. Schelter, M. Winterhalder, H. g. D. Feldwisch, J. Wohlmuth, J. Nawrath, A. Brandt, A. Schulze-Bonhage, and J. Timmer, Seizure prediction: The impact of long prediction horizons, *Epilepsy Research* **73**, 213–217 (2007).
- [51] R. De and G. Ananthakrishna, Possibility of Prediction of Avalanches in Power Law Systems, in *Proceeding of the first national conference on nonlinear systems and dynamics, India*, 2003, arXiv:cond-mat/0504212v1.
- [52] M. Ragwitz and H. Kantz, Detecting non-linear structure and predicting turbulent gusts in surface wind velocities, *Europhysics Letters* **51**(6), 595–601 (2000).
- [53] H. Kantz, D. Holstein, and M. Ragwitz, Extreme events in surface wind: Predicting turbulent gust, in *Experimental Chaos: 8th Experimental Chaos Conference. AIP Conference Proceedings*, edited by S. Boccaletti, volume 742, pages 315–324, AIP, 2004.

- [54] E. N. Lorenz, Atmospheric predictability as revealed by naturally occurring analogues, *J. Atmosph. Sci.* **26**, 636 (1969).
- [55] J. D. Farmer and J. J. Sidorowich, Predicting Chaotic Time Series, *Phys. Rev. Let.* **59**(8), 845–848 (1987).
- [56] H. Osterhage and K. Lehnertz, Nonlinear Time Series Analysis in Epilepsy, *International Journal of Bifurcation and Chaos* **17**, 3305–3323 (2007).
- [57] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, Elsevier, 2004.
- [58] H. Öttinger, *Stochastic Processes in Polymeric Fluids*, Springer, 1996.
- [59] M. Kijima, *Markov Processes for Stochastic Modeling*, Chapman & Hall, 1997.
- [60] H. Kantz and M. Ragwitz, Phase space reconstruction and nonlinear predictions for stationary and nonstationary Markovian processes, *Int. J. Bifurcation and Chaos* (14), 1935–1945 (2004).
- [61] H. Risken, *The Fokker-Planck Equation*, Springer, 2 edition, 1996.
- [62] T. Bayes, An essay towards solving a problem in the doctrine of chances, *Phil. Trans. Roy. Soc. London* **53**, 370–418 (1763).
- [63] J. M. Bernardo and A. F. Smith, *Bayesian Theory*, John Wiley & Sons, 1994.
- [64] G. A. Young and S. R. L., *Essentials of Statistical Inference*, Cambridge University Press, 2005.
- [65] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Elsevier, 1990.
- [66] J. H. Justice, *Maximum Entropy and Bayesian Methods in Applied Statistics*, Cambridge University Press, 1984.
- [67] R. A. Fisher, On the mathematical foundations of theoretical statistics, *Phi. Trans. Roy. Soc. A* **222**, 309–368 (1922).
- [68] R. A. Fisher, Theory of statistical estimation, *Proc. Camb. Phil. Soc.* **22**, 700–725 (1925).
- [69] Neyman-Pearson, On the Problem of the Most Efficient Tests of Statistical Hypotheses, *Philosophical Transactions of the Royal Society of London Ser. A*, **231**, 289–337 (1933).
- [70] H. Zhang, The Optimality of Naive Bayes, *American Association of Artificial Intelligence* (2004).
- [71] A. Garg and D. Roth, Understanding Probabilistic Classifiers, *ECML'01* (2001).
- [72] A. H. Murphy and R. L. Winkler, A General Framework for Forecast Verification, *Monthly Weather Review* **115**, 1330–1338 (1987).
- [73] I. T. Jolliffe and D. B. Stephenson, *Forecast Verification*, Wiley, 2003.
- [74] E. W. Weisstein, From MathWorld – A Wolfram Web Resource, <http://mathworld.wolfram.com/Predictability.html> .
- [75] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.
- [76] I. J. Good, Rational Decisions, *Journal of the Royal Statistical Society XIV(1) B*, 107–114 (1952).

- [77] J. J. L. Kelly, A new interpretation of information rate, *Bell System Techn. J.* **35**, 917–926 (1956).
- [78] T. A. Brown, Probabilistic Forecasts and Reproducing Scoring Systems, RANS Corporation **RM-6299-ARPA** (1970).
- [79] L. J. Savage, Elicitation of Personal Probabilities and Expectation, *J. Amer. Statist. Ass.* **66**, 783–801 (1971).
- [80] M. Roulston, Evaluating Probabilistic Forecast Using Information Theory, *Monthly Weather Review* **130**, 1653–1660 (2002).
- [81] G. W. Brier, Verification of forecasts expressed in terms of probability, *Mon. Weather Rev.* **78**, 1–3 (1950).
- [82] J. Bröcker and L. A. Smith, Scoring Probabilistic Forecasts: The Importance of Being Proper, *Weather and Forecasting* **22**, 382–388 (2007).
- [83] D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, Wiley, 1966.
- [84] J. P. Egan, *Signal Detection Theory and ROC Analysis*, Academic Press, 1975.
- [85] M. S. Pepe, *The Statistical Evaluation of Medical Tests for Classification and Prediction*, Oxford University Press, 2003.
- [86] I. N. Bronstein, K. A. Semendjajew, G. Musiol, and H. Mühlig, *Taschenbuch der Mathematik*, Verlag Harri Deutsch, 1999.
- [87] S. T. Zhang, Jun and Mueller, A Note on ROC Analysis and non-parametric Estimate of Sensitivity, *Psychometrika* **70**(1), 203–212 (Mar. 2005).
- [88] A. P. Prudnikov and Y. A. Brychkov, *Integrals and Series Vol. II Special Functions*, Gordon and Breach Science Publishers, 1992.
- [89] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, 1972.
- [90] I. Wolfram Research, *Mathematica Edition: Version 5.2*, Wolfram Research, Inc., 2005.
- [91] D. Sornette and J. V. Andersen, Increments of Uncorrelated Time Series Can Be Predicted With a Universal 75% Probability of Success, *Int. J. Mod. Phys. C* **11**, 713–720 (2000).
- [92] S. Prakash, S. Havlin, M. Schwartz, and H. E. Stanley, Structural and Dynamical Properties of Long-Range Correlated Percolation, *Phys. Rev. A* **46**, R1724–R1727 (1992).
- [93] A. Savitzky and J. Golay, Marcel, Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Analytical Chemistry* **36**, 1627–1639 (1964).
- [94] W. H. Press, *Numerical Recipes in C*, Cambridge University Press, 1992.
- [95] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, Acad. Press, 2000.
- [96] M. Matsumoto and T. Nishimura, Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator, *ACM Trans. on Modeling and Computer Simulations* (1998).
- [97] U. Frisch, *Turbulence*, Cambridge University Press, 1995.
- [98] C. Renner, J. Peinke, and R. Friedrich, Experimental indications for Markov properties of small-scale turbulence, *J. Fluid Mech.* **433**, 383–409 (2001).

- [99] G. Batchelor and A. A. Townsend, The nature of turbulent motion at large wave-numbers, Proc. R. Soc. Lond. **A**(199), 238–255 (1949).
- [100] R. H. Kraichnan, Intermittency in the very small scales of turbulence, Phys. Fluids **10**, 2080–2082 (1967).
- [101] A. Kuo and S. Corrsin, Experiments on internal intermittency and fine-structure distribution function in fully turbulent fluid, J. Fluid Mech. **50**, 285–320 (1971).
- [102] U. Frisch and R. Morf, Intermittency in nonlinear dynamics and singularities at complex times, Phys. Rev. A **A**(23), 2673–2705 (1981).
- [103] H. G. Schuster and W. Just, *Deterministic Chaos*, Wiley-VCH, 2005.
- [104] A. N. Kolmogorov, A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number, J. Fluid Mech. (13), 82 (1962).
- [105] A. M. Obukov, Some specific features of atmospheric turbulence, J. Fluid Mech. (13), 77.
- [106] B. Castaing, Y. Gagne, and M. Marchand, Log-similarity for turbulent flows, Physica D **68** (1993).
- [107] C. Van Atta and J. Park, *Statistical Models and Turbulence*, volume 12 of *Lect. Notes in Phys.*, chapter Statistical self-similarity and inertial subrange turbulence, pages 402–426, Springer, 1972.
- [108] Y. Gagne and E. Hopfinger, *New Trends in Nonlinear Dynamics and Pattern-Forming Phenomena*, volume 237 of *NATO ASI*, chapter A new universal scaling for fully developed turbulence: the distribution of velocity increments, pages 315–319, Plenum Press, New York, 1990.
- [109] B. Castaing, Y. Gagne, and E. Hopfinger, Velocity probability density functions of high Reynolds number turbulence, Physica D **46**, 177–200 (1990).
- [110] M. Vincent, A. and Meneguzzi, The spatial structure and statistical properties of homogeneous turbulence, J. Fluid Mech. **225**(1-25) (1991).
- [111] R. H. Kraichnan, On Kolmogorov’s inertial-range theories, J. Fluid Mech. **62**, 305–330 (1974).
- [112] M. Galassi, J. Davies, J. Theiler, B. Gough, G. Jungman, M. Booth, and F. Rossi, *GNU Scientific Library Reference Manual - Revised Second Edition*, Network Theory Ltd., 2006.
- [113] C. Ziehmann, L. A. Smith, and J. Kurths, Localized Lyapunov exponents and the prediction of predictability, Physics Letters A **271**, 237–251 (2000).
- [114] M. Ghil, Are extreme events like smaller events only larger?, A question proposed at the E2C2 Open Conference in Paris (2008).
- [115] S. M. Berman, Limit Theorems for the Maximum Term in Stationary Sequences, Ann. Math. Stat. **35**, 502 (1964).
- [116] J. F. Eichner, J. W. Kantelhardt, A. Bunde, and S. Havlin, Extreme value statistics in records with long-term persistence, Extreme value statistics in records with long-term persistence **73**, 016130 (2006).

- [117] Y. Malevergne, V. Pisanrenko, and D. Sornette, On the power of generalized extreme value (GEV) and generalized Pareto distribution (GDP) estimators for empirical distributions of stock returns, *Applied Financial Economics* **16**, 271–289 (2006).
- [118] J. F. Muzy, E. Bacry, and A. Kozhemyak, Extreme values and fat tails of multifractal fluctuations, *Phys. Rev. E* **73**, 066114 (2006).
- [119] J. Peinke, F. Böttcher, and S. Barth, Anomalous statistics in turbulence, financial markets and other complex systems, *Ann. Phys.* **13**, 450–460 (2004).
- [120] J. Davidsen, P. Grassberger, and M. Paczuski, Earthquake recurrence as a record breaking process, *Geophysical Research Letters* **33**, L11304 (2006).
- [121] J. Spanier and K. B. Oldham, *An Atlas of Functions*, Hemisphere Publ., 1987.