Xtremes User Manual

R.-D. Reiss
M. Thomas

Xtremes Group
Zur Alten Burg 29
57076 Siegen
Germany

Inquiries

Please e-mail all inquiries and bug reports to
info@xtremes.de

Check the Xtremes homepage for the latest information:
http://www.xtremes.de

Copyright Notice

Disclaimer

Contents

1 Introduction **1**
1.1 Hierarchy of the Menu System 2
1.2 The Risktec Computing Environment 5
1.2.1 StatPascal: Textual Programming in Statistics 5
1.2.2 RiskTec: A Client/Server Architecture 5
1.2.3 XGPL: Graphical Programming in Statistics 7
1.3 Further Documentation 7

2 Installation and First Steps **9**
2.1 Installation of Xtremes 9
2.2 Example Session 10
2.2.1 Plotting Curves 10
2.2.2 Selecting a Mouse Mode 11

3 Plotting Curves **13**
3.1 Plotting Histograms of Discrete Distributions 13
3.2 Plotting Parametric, Continuous Curves 14
3.3 Using Sliders in Plot Windows 15
3.4 Documenting Illustrations 15
3.5 The Toolbar 18
3.6 Microsoft Windows Commands 21

4 Handling and Visualizing Data **23**
4.1 Overview of Data Handling Options 23
4.1.1 Visualization of Data 24
4.1.2 Applying Estimators to the Active Data 25
4.2 Discrete Data 25
4.2.1 Generating Discrete Data 26
4.2.2 Loading and Editing Discrete Data 26
4.2.3 Plotting a Sample Histogram 27
4.2.4 Fitting a Parametric to a Sample Histogram 27
4.3 Continuous Data 28

4.3.1 Generating Continuous Data . . . 28
4.3.2 Visualization of Continuous Data . . . 29
4.4 Time Series . . . 30
4.4.1 Clustering of Data . . . 32
4.5 Format of Data Sets . . . 33
4.6 Generating Data Sets from StatPascal . . . 35

5 Univariate Estimators and Test Procedures **37**
5.1 Gaussian and Related Models . . . 37
5.1.1 Estimators . . . 37
5.1.2 Test Procedures . . . 38
5.2 Extreme Value Models . . . 39
5.2.1 Simulations . . . 39
5.2.2 Units Exposed to Corrosion . . . 40
5.2.3 Gompertz model . . . 41
5.3 Generalized Pareto Distributions . . . 41
5.3.1 Estimators . . . 41
5.3.2 Simulations . . . 42
5.3.3 Censoring . . . 43

6 Multivariate Estimators **45**
6.1 Plot Options . . . 45
6.2 Multivariate Extreme Value Models . . . 47

7 The HYD, ACT and Finance Menus **49**
7.1 Hydrological Menu Options . . . 49
7.1.1 The Hydrograph . . . 49
7.1.2 Calculating T−Year Levels Based on Maxima . . . 49
7.1.3 Partial Duration Series . . . 50
7.2 Actuarial Menu Options . . . 51
7.2.1 Analyzing PML Groups . . . 51
7.2.2 Claim Number Process . . . 52
7.2.3 Risk Process . . . 52
7.3 Analyzing Financial Data . . . 53

8 UserFormula Facilities **55**
8.1 Overview . . . 55
8.2 Calculator . . . 56
8.3 Plotting Curves . . . 57
8.4 Generating Data . . . 57
8.5 Transforming Data . . . 57

Chapter 1

Introduction

Xtremes is a statistical software system that focuses on extreme value analysis. It possesses graphics facilities, a facility to generate and load data, an arsenal of diagnostic tools and statistical procedures, and a numerical part for Monte–Carlo simulations. The pull–down menu structure, in conjunction with dialog boxes and help facilities, makes the program simple and straightforward to use.

An important step towards a computerized extreme value analysis was already done with the MS–DOS version of Xtremes included in the first edition of *Laws of Small Number: Extremes and Rare Events* by M. Falk, J. Hüsler and R.-D. Reiss in 1994. A substantial improvement was achieved by the Windows version of Xtremes in the first edition of *Statistical Analysis.* Xtremes is now applicable on platforms such as Windows, Linux and other Unix platforms.

The UserFOrmula facility may be utilized when a single formula is required to plot a curve or to generate or transform a data set. The applicability of the system can be enhanced by the integrated statistical programming language StatPascal. A separate overview is given for StatPascal and the graphical programming environment XGPL (within RiskTec).

Students may perform their own statistical experiments during a laboratory course. One may start with working through this manual and the appendix of *Statistical Analysis,* and, afterwards, analyze the included real data. The statistical difficulties, inherent in extreme values, may lead to the following advice[1]: "Extreme value analysis should never be performed automatically, without the intervention of the data analyst, because failure to spot unusual features of the data could have serious consequences." An application of this book in conjunction with the included software system can partially serve as an alternative to consulting analysts, yet it is not scheduled as an expert system pretending to lead automatically to reasonable answers.

[1]Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds. J. R. Statist. Soc. B 52, 393–442.

The development of the Xtremes menu system and of the RiskTec software environment has been supported by Patrick Bornhütter (multivariate Student estimator) Simon Budig (UserFormula facility), Martin Elsner (HTML help), Andreas Gaumann (help system), Sylvia Haßmann (censored data), Andreas Heimel (help system, ARMA estimators), Jens Olejak (minimum distance estimators), Wolfgang Merzenich (consultation on the StatPascal compiler), Reinhard Pfau (previous Linux port), Torsten Spillmann (copulas, early version of XGPL plots), Karsten Tambor (early version of the multivatiate mode), and Arthur Böshans, Carsten Wehn, Lars Fischer, Ralf Pollnow (MS Excel frontend) whose help is gratefully acknowledged.

Special thanks are due to John P. Nolan for contributing the DLL–version of his STABLE package, now available within the univariate SUM domain of the Xtremes menu system.

1.1 Hierarchy of the Menu System

We start with some general remarks about statistical computing and programming: "A fundamental design problem is how to balance flexibility and the resulting complexity of options against ease–of–use and ease–of–learning.[2]"

The latter requirements can be satisfied by graphical user interfaces (GUIs) as provided by the Xtremes menu system. The advantage of GUIs is apparent[3]:

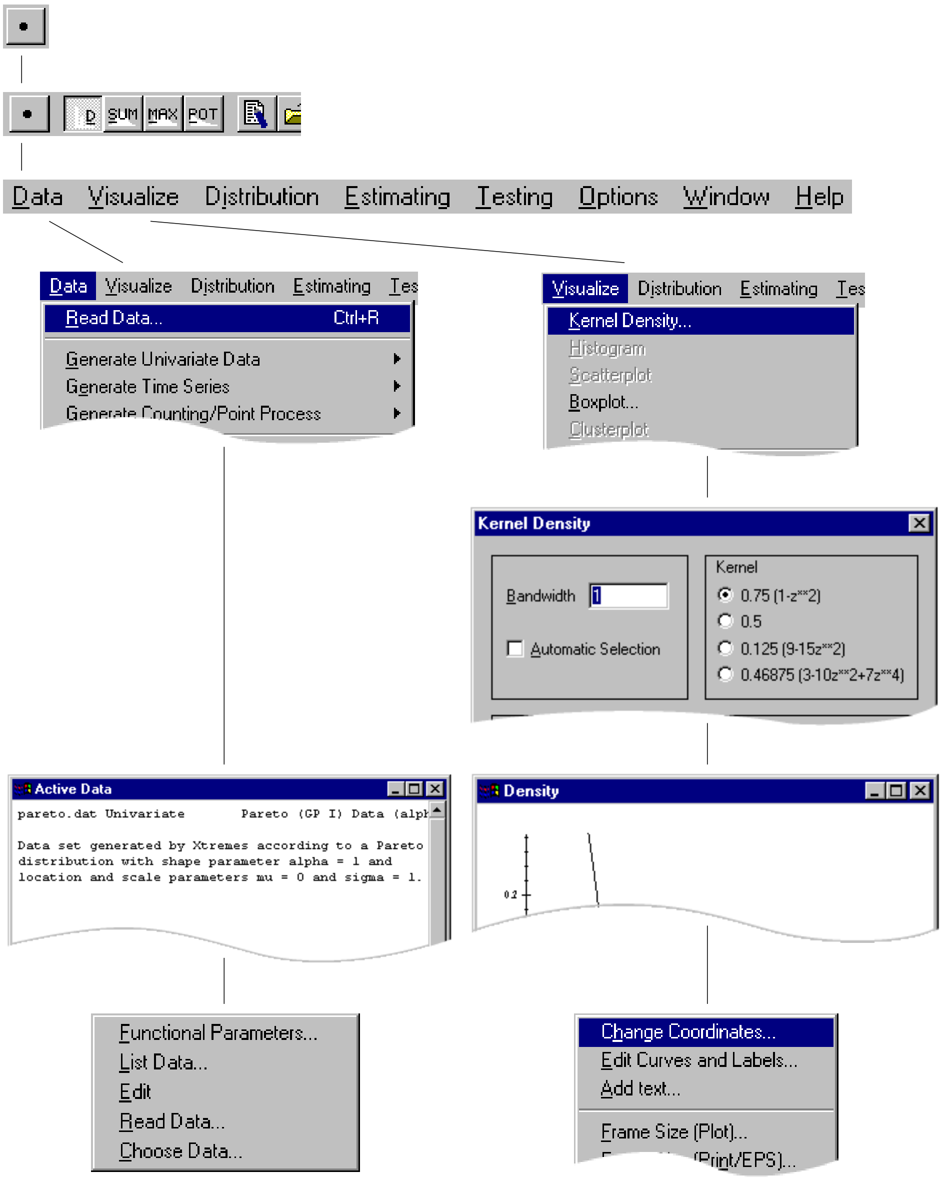
- "all those buttons, sliders, menus, flying icons, waving elves and other visual aids to computing ...
- The point–and–click, drag–and–drop, and generally non–verbal approaches to computing must be accepted as increasingly dominant, particularly for non–technical audiences."

These ideas of efficient GUIs are further developed and integrated in the present Xtremes menu system.

Users having some experience with statistical software systems will be able to run larger parts of the menu system without reading much of the advice and information provided by this text. Yet, it can be helpful to be aware of certain special features of Xtremes as noted on the following two pages.

[2]Biehler, R. (1997). Software for learning and for doing statistics. Int. Stat. Reviews 65, 167–189.

[3]Chambers, J.M. (1998). Computing with data: Concepts and challenges. Technical report, Bell Labs.


D SUM MAX POT
Data Visualize Distribution Estimating Testing Options Window Help
Data Visualize Distribution Estimating Tes
Read Data... Ctrl+R
Generate Univariate Data
Generate Time Series
Generate Counting/Point Process
Visualize Distribution Estimating Tes
Kernel Density...
Histogram
Scatterplot
Boxplot...
Clusterplot
Kernel Density
Bandwidth
Automatic Selection
Kernel
0.75 (1-z**2)
0.5
0.125 (9-15z**2)
0.46875 (3-10z**2+7z**4)
Active Data
pareto.dat Univariate Pareto (GP I) Data (alph
Data set generated by Xtremes according to a Pareto distribution with shape parameter alpha = 1 and location and scale parameters mu = 0 and sigma = 1.
Density
0.2
Functional Parameters...
List Data...
Edit
Read Data...
Choose Data...
Change Coordinates...
Edit Curves and Labels...
Add text...
Frame Size (Plot)...

The illustration on the left–hand side shows the hierarchy of the Xtremes system:

- Univariate and Multivariate Modes: the system is partitioned into a univariate and a multivariate mode that can be selected in the toolbar.
- Domains D(iscrete), SUM, MAX and POT: select certain domains in the toolbar which correspond to different parametric models (discrete, Gaussian, extreme value (EV) or generalized Pareto (GP)) built for discrete data and data that are sums, maxima or exceedances (peaks–over–threshold).
- The Menu–Bar: select menus for handling and visualizing data, plotting distributions by analytical curves, estimating parameters etc. The *Visualize* menu and larger parts of the *Data* menu are independent of the different domains.
- Menus: commands of the system are available in different menus.
- Dialog Boxes: if a menu command requires further parameters, a dialog box is displayed.
- Graphics Windows: the result of a menu command is usually displayed in a graphics window.
- Local Menus: options and commands that are specific to a particular window or dialog box are provided by means of a local menu (available with a right–click somewhere in the window or the dialog box).
- Special Facilities: in the toolbar, select tools to manipulate a graphic. These tools change the action taking place when one drags the mouse in a graphics window. For example, to change the coordinate system, you may click on the coordinate tool in the toolbar and, then, pull up a rectangle.
- Help System: a context–sensitive online help is available by pressing the F1–key at any time besides the general help facility (option *Index* in the *Help* menu).
- Act, $, Hyd Supplements: special options for insurance, finance and hydrology (Chapters 14–16 in Statistical Analysis) are available.
- UFO: select the UserFormula facility to enter your own formulas to plot curves, generate or to transform data (see Chapter 8).
- StatPascal: the integrated programming language StatPascal is activated by means of the SP button.

Keep in mind that the options in the *Visualize* and *Estimator* menus must be applied to an active data set (read from a file or generated in the *Data* menu). Instructions about the installation of the system are given in section 2.1.

1.2 The Risktec Computing Environment

Besides the menu system of Xtremes there is the integrated textual programming language StatPascal. The combination of both provides a solution to the fundamental design problem. As a third element there is the XGPL system for graphical programming in statistics and integrating statistical components (cf. the end of this introduction).

1.2.1 StatPascal: Textual Programming in Statistics

The integrated textual programming language StatPascal enables the user to program extensions to the menu system beyond the possibilities of the UFO facility. Menu options of Xtremes (e.g., names of estimators and distributions) are available as predefined functions and procedures. StatPascal is based on the Pascal language, so users with some experience in Pascal will find that the language is easy to use. Moreover, existing Pascal programs can be ported to StatPascal easily.

StatPascal supports vector and matrix operations like other common statistical languages. A notable difference is the strong type system: each identifier within a StatPascal program must be declared and is associated with a fixed data type. StatPascal programs can therefore be compiled easily. We generate a code for an abstract stack machine which has been designed to handle vectorized operations efficiently. This approach results in a higher execution speed of StatPascal, compared to interpreted systems. Moreover, thorough checks of the static semantics of a program can be performed.

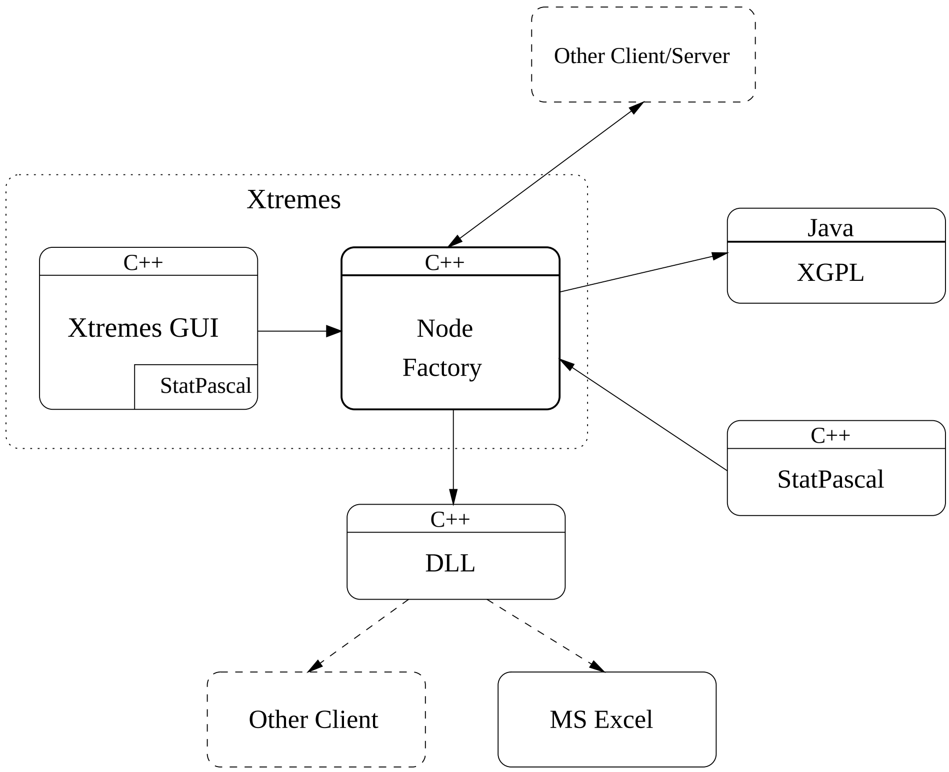
As a consequence of the declarative nature of StatPascal, the language cannot be used as an interactive command interpreter. We believe that interactive tasks are performed more easily using a menu system or a graphical programming language. Quick transformations and plots—requiring the user to enter a single formula—can be performed using the UserFormula facility. Thus, employing a compiled language does not impose serious restrictions on the usability of the system.

A separate command line version of StatPascal is available for Windows and Unix systems. It provides the same functionality (with the exception of graphical output) and is well suited to perform batch jobs with simulations. The command line version is also used to program new components for the RiskTec environment.

For an introduction to StatPascal, the reader is referred to the StatPascal manual, which is provided in the file `spdoc.ps` in the `sp` directory of the Xtremes CD-ROM.

1.2.2 RiskTec: A Client/Server Architecture

Xtremes is embedded in the CORBA–based client/server architecture RiskTec that is shown in the subsequent figure. The central part of the architecture is a component factory (implemented as a part of the Xtremes executable) that produces statistical components to be used by clients.



RiskTec Architecture: arrows point from server to client, whereby solid lines indicate CORBA–based connections.

One client is the XGPL system which allows a visual combination of the components by using a graphical programming language (see the next section for more information). The second client is a DLL that encapsulates the CORBA–based components with a procedural interface. Such an interface can be accessed from MS Excel or other systems capable of including DLLs.

There are two predefined servers in the statistical environment. The first one is the Xtremes system, which exports estimators, data generation routines and plotting facilities. The second one is the command line version of StatPascal with which new components can be written easily by means of *component* programs.

Because the interface definitions of the components and the component factory are open to the user, it is possible to implement further CORBA–based clients and servers. The RiskTec environment is documented in the *RiskTec User Manual.*

1.2.3 XGPL: Graphical Programming in Statistics

XGPL[4] is a graphical programming language that allows the visual combination of statistical components by building a dependence graph. Components are provided by the Risktec environment, which include,

- estimators and visualizations from the Xtremes menu system,
- user written StatPascal component programs,
- external programs (implemented e.g. in Java or C++) that utilize the Risk-Tec CORBA interface.

XGPL provides sliders, access to database systems and the implementation of iterations by encapsulating a subgraph in a meta node. Because XGPL immediately recalculates the affected parts of a graph when the user changes an input value, the system is well suited to develop interactive simulations.

Fig. 1.1 shows a typical application of XGPL. We simulate the distribution of the Hill estimator and the MLE(GP) under varying generalised Pareto distributions. The sliders are provided by XGPL, while the plot routines are exported by Xtremes. The simulation and the calculation of a kernel density is accomplished by means of StatPascal programs, which use predefined functions like *mlegp* for the m.l. estimator in the GP model.

XGPL is implemented in the Java language and therefore runs on a large number of platforms.

1.3 Further Documentation

In addition to this manual, there is further documentation available.

- Online help facility: within the Xtremes menu system, there is context sensitive online help available by pressing the F1 key.
- StatPascal manual: the StatPascal language and its runtime library (giving access to the estimators, distributions and plotting facilities of Xtremes) are described in the StatPascal manual. It also contains a tutorial on Pascal for readers who are unfamiliar with that language.
- Risktec manual: the Risktec environment is documented in the Risktec manual. It covers the following topics:
  - MS Excel plugin
  - DLL interface
  - Access from the R system
  - CORBA interface and Java examples

[4] See http://www.xtremes.de/xgpl/

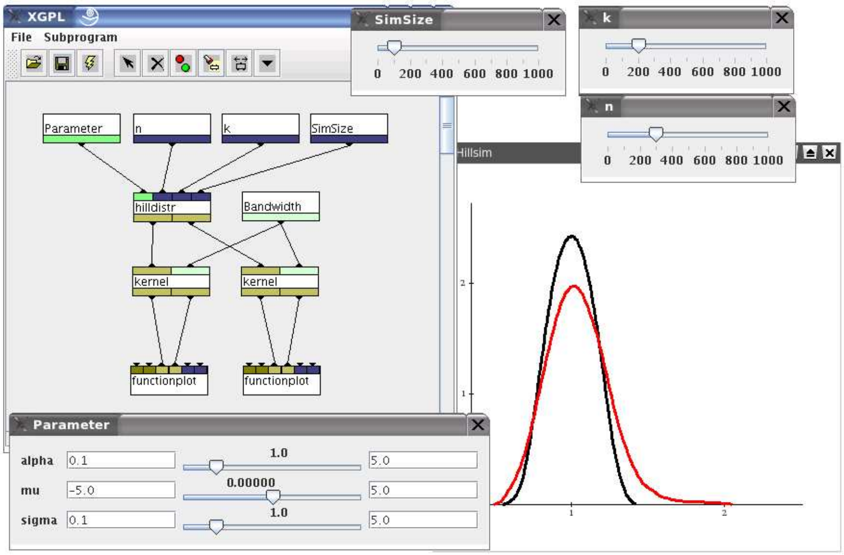


FIG. 1.1. XGPL program simulating distribution of different estimators.

Chapter 2

Installation and First Steps

This chapter describes the installation of the statistical software system Xtremes and provides a first step into the menu system. More details are given in the following chapters.

2.1 Installation of Xtremes

We describe the basic requirements and explain the steps for installing the software system Xtremes on your computer under Windows. Xtremes is a genuine Windows application so that any system running Windows 98/ME/NT/2000/XP/2003 is also capable of executing Xtremes. To facilitate the installation, the CD–ROM contains an installation program that copies all required files to your disk. The installation program should start automatically when you insert the CD–ROM (omit steps 1 and 2 in that case).

1. If the installation program does not start when you insert the CD–ROM, select the *Run* option in the Windows *Start* menu. Under Windows NT 3.51, select the option *File... Execute* in the Program Manager instead.

2. Next, type `d:\install` in the edit line (replace `d` with the drive letter of your CD–ROM).

3. The installation program displays a greeting message. Click the *Continue* button to move on to the installation options.

4. The *Installation Options* dialogue asks if optional parts of Xtremes should be installed. If you are unsure, it is recommended to accept the default options which will install the entire package. Moreover, one can specify a drive and directory where Xtremes should be installed. If a previous version of Xtremes was installed in the suggested directory `c:\Program Files\xtremes`, one may select a different name as, e.g., `c:\Program Files\xtr3`.

Click the *Continue* button to move on to the next dialogue.

5. The *Summary of Installation Options* dialogue displays the available disk space and the space required according to the selection of installation options in the previous dialogue. Click the *Back* button to change your installation options or the *Install* button to begin copying files to your system.

The installation program creates a deinstallation option in the *Software* section of the Windows *Control Panel*. To remove Xtremes from your computer, we recommend to apply that option instead of just deleting the installation directory.

If you do not want to use the installation program, you can also execute Xtremes directly from the CD-ROM. Change to the subdirectory `xtremes` and execute `xtremes.exe` or the DLL-version `xtremes1.exe`. To install Xtremes to your hard disk manually, we suggest to copy the complete `xtremes` folder.

2.2 Example Session

A basic understanding of the structure of Xtremes can already be gained from the Overview given at the beginning of the manual. In the following lines, we describe an example session in which we plot the density of Pareto distribution. More details on the plot options are provided in the following chapter. Anotherr possibility is to work closely with the online–help facility (press the F1–key or enter the *Help... Index*).

Start Xtremes by selecting the Xtremes entry from the *Start* menu. After starting the program, music or a spoken text may be output. For that purpose, install a wave file called rose.wav in the working directory. It is understood that the necessary technical equipment is installed.

2.2.1 Plotting Curves

Xtremes provides easy–to–use plotting facilities. The user can quickly plot a curve, adjust the coordinate system, get a list of the curves displayed in a plot window or export the picture via the clipboard. A special facility is available for producing EPS–files (see Documenting Illustrations on page 15).

To start, let us display a Pareto density on the screen. Make sure that the univariate POT domain is active, that is, the first button in the toolbar shows a single bullet and the button labeled POT is pressed. Select the menu option *Distribution... Pareto* to open a dialog box asking for the parameters of the distribution (see Demo 3.2). The box provides buttons to plot densities, qfs, dfs, mean and median excess functions. Click the density button. A window opens displaying the graph of your curve. Xtremes selects the coordinate system automatically. Another density selected from the menu will be displayed in the same window yet with a different color.

The coordinate system can be adjusted by using the local menu of the window. Click inside the window with the right mouse button and select the option *Change coordinates*. Xtremes stores the previous coordinates which may be restored by executing the option *Restore coordinates* in the toolbar (click the  button) or by pressing the Backspace key.

A more direct, interactive facility to modify the coordinate system is available by selecting a certain mouse mode from the tool bar (explained in the next subsection).

One can get an overview of the curves plotted in a window utilizing the menu option *Edit Curves and Labels*. A list with all curves is displayed, and one can delete some curves or change their options like color, line style, brushes used to hatch histograms, etc. If one wants to know the parameters of a plotted curve, select the information mouse mode tool  and click onto the curve.

Pictures may be exported. Click the option *Print* in the local menu to send your window to the printer. You can also copy it to the clipboard or create EPS–files. Advanced options to format a plot are described in section 3.4 on page 15.

2.2.2 Selecting a Mouse Mode

The mouse mode determines what happens if you click into a window. The default mode just brings a window into the foreground.

Other mouse modes are employed to move or delete curves, change colors and plot options, add text, etc. Detailed explanations of all modes are given in Section 3.5.

As an example, we describe the change of the coordinate system using a specific mouse mode. Activate the coordinate changing mouse mode by clicking the  tool in the toolbar. In this mode, a new coordinate system is selected by pulling up a rectangle: click into the graphics window (left mouse key), hold down the mouse key and move the cursor to the opposite corner. The rectangle may be pulled outside the window to enlarge the coordinate system.

Chapter 3

Plotting Curves

Note that Xtremes starts in the univariate DISCRETE domain. Step–by–step, select the required mode, domain, menu and option by means of left mouseclicks. Then, a dialog box such as the Pareto GP1 dialog box in Fig. 3.1 opens, where one may execute an option.

The distributions introduced in the foregoing sections are available within the menu system of Xtremes. Certain parametric curves (such as histograms, densities, dfs, qfs, etc.) can be displayed in plot windows. In addition, one may generate data according to these distributions (see page 28).

3.1 Plotting Histograms of Discrete Distributions

We repeat the operations which are required to plot the histograms in Statistical Analysis, Fig. 1.2 on page 9. The binomial and Poisson histograms $B_{n,p}\{k\}$ and $P_{np}\{k\}$, $k = 0, \ldots, 20$, for the parameters $n = 40$ and $p = 0.25$ are displayed.

DEMO 3.1. (Plotting Binomial and Poisson Histograms.) (a) Select the option *Distribution... Poisson* in the univariate D(iscrete) domain and choose the parameter $\lambda = np = 10$. Execute the *Histogram* option.

(b) Next, select the *Binomial* option in the same menu and enter $n = 40$ and $p = 0.25$.

(c) The  tool (activating the option mouse mode) may be employed to adjust the positions of the plotted bars.

Xtremes provides one window for each type of curve (histogram, distribution function, quantile function, density, etc.). There is also a clipboard window which is accessible using the clipboard mouse mode tool : put the curve into the clipboard window with an activated clipboard mouse mode and a click onto the curve. Alternatively, drag the curve to a different open plot window with the left mouse button kept pressed.

Activate the deleting mouse mode tool and delete a histogram (curve) by clicking onto it if this is necessary. Also use the coordinate changing mouse mode tool : adjust the coordinate system by pulling up a rectangle with the left mouse button kept pressed.

3.2 Plotting Parametric, Continuous Curves

Select one of the domains SUM, MAX or POT and execute the *Distribution* option in the menu bar. All distributions belonging to the given domain are listed in a pulldown menu. A dialog box opens after having selected a specific distribution. Choose parameters and select a curve.

We display such a dialog box for the Pareto distribution and discuss more detail in the subsequent Demo 3.2

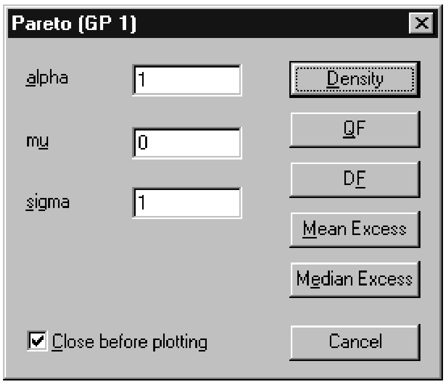


FIG. 3.1. Dialog box *POT... Distribution... Pareto* in the univariate mode. Enter the parameters of the Pareto distribution and plot a curve using the buttons.

DEMO 3.2. (Plotting a Pareto Density.) The dialog box for the Pareto distribution is displayed in Fig. 3.1. Click on the *Density* button to plot the Pareto density for the chosen parameters. A plot window opens showing the graph of the density. The coordinate system is adjusted automatically. Change the coordinate system in the dialog box *Change Coordinates* which is available in the local menu of the plot window and in the *Window* menu. Use the ticks and markers option in the same dialog box to adjust the numbers and position of labels.

It is also possible to draw a straight line into a plot window by pressing the left mouse button. For that purpose activate the line drawing mouse mode with the tool . To create line segments (i.e., lines between start and end point of the mouse interaction) by holding the CTRL–key when creating the line. Note that line segments are not represented as linear functions and may therefore be vertical.

Curves are evaluated at 100 equally–spaced points and, afterwards, a linear interpolation is taken. One may increase the number of points if the visualization is unsatisfactory by means of the options mouse mode tool. A left mouse–click, with activated options mouse mode, opens a dialog box, where one can choose the colors, line styles and plot modes such as linear interpolation (default plot mode), points, bars etc.

3.3 Using Sliders in Plot Windows

The user may change the parameters of a plotted curve by means of sliders which are an indispensable tool to work interactively with data. Select the parameter varying mouse mode  from the toolbar and click onto a curve in a plot window to open a window with sliders for each parameter.

DEMO 3.3. (Varying the Parameter of a Poisson Histogram.) First, plot a Poisson histogram with parameter $\lambda = 1$. Select the parameter varying mouse mode from the toolbar and click on the histogram of the Poisson distribution. Vary the parameter λ, and also change the boundaries for the possible parameter values. In Fig. 3.2 there is a Poisson histogram with the pertaining slider and, in addition, a sample histogram. For a continuation see Demo 4.8.

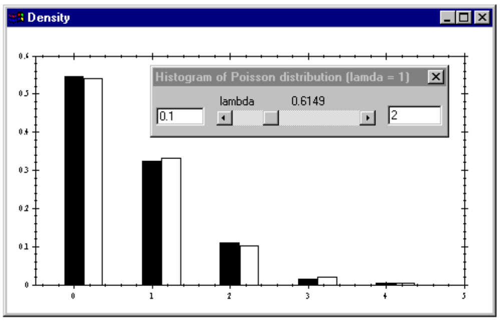


FIG. 3.2. Interactive fitting of a Poisson histogram (right) to sample histogram (left).

The next example concerns a curve which is governed by two parameters, where two sliders are available.

DEMO 3.4. (Changing Parameters and the Bandwidth.) Fig. 3.3 shows a Gumbel density which is fitted to a sample density. The location and scale parameters of the Gumbel density can be varied by the two sliders. The illustration in Fig. 3.3 was produced in this manner. For a continuation see Demo 4.5.

Note that the same technique is utilized in the case of a a kernel (sample) density, where a smoothing parameter can be interactively varied.

3.4 Documenting Illustrations

Xtremes provides various tools to change the outer appearance of a plot and to export it to other systems. One may

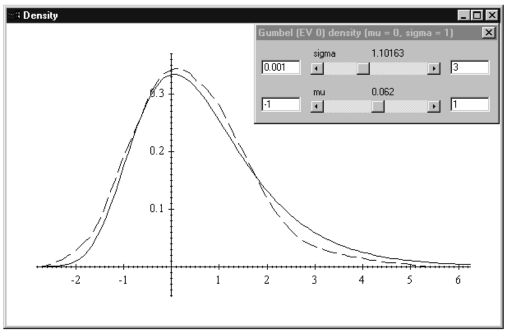


FIG. 3.3. Fitting a Gumbel density (solid) to a sample density (dashed). Use the sliders in the dialog box to change the parameter values.

- copy the contents of a window to the clipboard,
- print a window, or
- create an EPS file.

First change the outer appearance of the plot using the *Change Coordinates* and *Change Curve Options* dialog boxes (for the latter activate the options mouse mode tool  and click onto a curve).

In addition use the Frame Size (Print/EPS) dialog box which is available in the local menu of the plot window. This option sets the size of the exported plot; it also adjusts the size of the window on your screen to give an impression of how the output is going to look like.

We start with a description of advanced plot options (like different colors and line styles) that are used to prepare pictures for exporting. The following options are available:

- Coordinate System: the coordinate system is either displayed within the window or on a rectangle around the actual plot area. These options are controlled in the *Change Coordinates* box of the local menu. The portion of the plot area may be changed to provide space for the attachment of labels outside the frame using the *Frame Size* option.

- Line Styles and Colors: the option mouse mode tool  is used to change the plot options of a curve. A left–click onto the curves opens a dialog box (cf. Fig. 3.4). The user may select

  – predefined line styles,
  – define his own line style by specifying the length of curve segments and gaps

  as well as the thickness of the curve. For example, choose the values

  – 1 and 1 to produce a dotted line,

– 4 and 4 to produce a dashed line.

These procedures lead to a better result on printed pages than the use of predefined line styles (except of the solid line).

Xtremes offers various plot styles for curves. The *Style of Points* option selects a shape for the points that are plotted when the line style *Points* or *Connected Points* is utilized.

Different sizes and hatch styles are provided for histograms. The local menu of a scatterplot window provides the *Options* entry to change the point size.

- Adding Text: select the label mouse mode tool and click at the position where you want to put your label. The font and position of the text may be changed using the parameter varying and option mouse mode tools. It is possible to display vertical text or to move a label to the edge of the window. Labels are treated like curves, so they may be moved to another window or deleted in the same way.

The box for curve options is presented in the following Fig. 3.4. We suggest to use the solid line option or the "line" and "gap" facility.

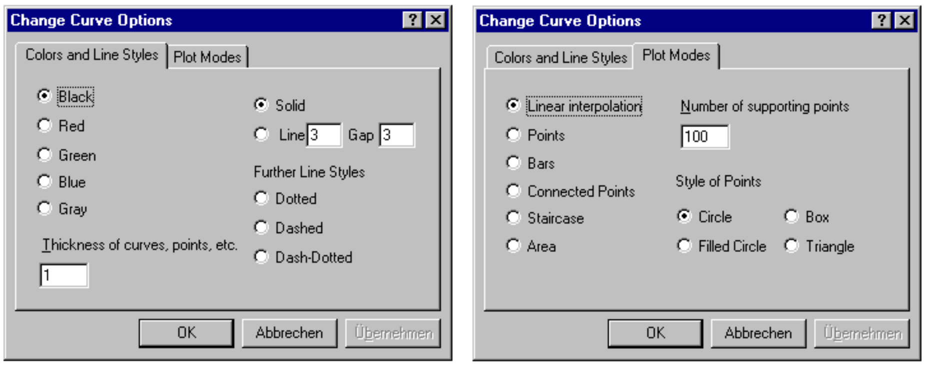


FIG. 3.4. Property sheet to specify plot options.

The contents of an Xtremes plot window may be exported, either by printing a window, saving it as an EPS (Encapsulated Postscript) file or storing it in the clipboard.

- Printing: first, select the option *Frame Size (Print/EPS)* (cf. Fig. 3.5) from the local menu of the active window to define the size of the picture and provide space for the frame (see the next Demo A.1). Then, select *Print* to copy the contents to your printer. *Printer Setup* is utilized to change options of the printer.

- Saving an EPS file: first, set the size of the picture and frame, as in the previous case. Then, select *Save as EPS file* to store the contents in the EPS format. Xtremes asks for a filename.

- Copying to the clipboard (Option *Copy to Clipboard* in the local menu): the contents of the active window are copied to the clipboard in the standard bitmap format. It is possible to insert the contents of the clipboard in other applications like painting programs or word processors.

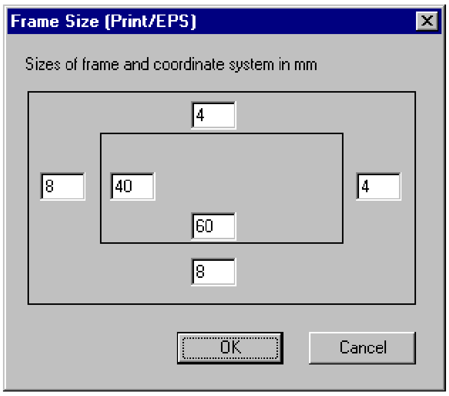


FIG. 3.5. *Frame Size (Print/EPS)* dialog box. The user selects the size of the coordinate system and provides space for text displayed outside the actual picture.

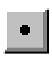
In the following demo further explanations are provided about the dialog box in Fig. 3.5.

DEMO 3.5. (Printing a Graphics Window.) Select the *Frame Size (Print/EPS) option* from the local menu and enter the size of your picture in the dialog box. The values shown in Fig. A.5 entail a picture of the size 72mm $\times$ 52mm, the actual plot area comprises 60mm $\times$ 40mm. After that, proceed with the *Print option* to copy the active window to your printer or select *Save as EPS file* to create an EPS file that can be included in other documents.

LaTeX–users may employ the *epsf* macro package, which provides commands to include postscript files into LaTeX documents (e.g., the commands `\epsfxsize=72mm` `\epsfbox{picture.eps}` load `picture.eps` and scale it to a horizontal size of 72mm).

3.5 The Toolbar

The toolbar below the main menu provides a quick access to frequently used options of Xtremes. The tools enable the user to select different parametric distributions in the main menu. They are also used to select a mouse mode. We start with the tools already described in the Overview.



Switch menu bar from univariate to multivariate mode.

Switch menu bar from multivariate to univariate mode.

Activates pulldown menus for discrete models.

Activates pulldown menus for Gaussian models.

Activates pulldown menus for extreme value (EV) models.

Activates pulldown menus for generalized Pareto (GP) models.

Opens pulldown menu with options for hydrology data (Chapter 14 in Statistical Analysis).

Opens pulldown menu with options for insurances data (Chapter 17 in Statistical Analysis).

Opens pulldown menu with options for finance data (Chapter 16 in Statistical Analysis).

Opens pulldown menu providing UFO facilities.

Opens *StatPascal Editor Window* to enter and run StatPascal programs.

Next the tools are listed that are not described in the Overview.

Opens ASCII–editor window.

Opens the Windows file dialog box and loads a data set. The file dialog box provides options to delete or copy files.

The active data set is displayed in a text window.

Restores coordinate system in active window to the size before the last change.

The toolbar is also used to select a mouse mode. The mouse mode determines the action taking place when the user clicks into a window or onto a curve[1].

Standard mouse mode: no special actions occurs.

[1]Mouse modes, where one must click onto a curve, are marked with *.

Option mouse mode*: changes display options of a curve (e.g. color, line styles, number of supporting points, etc.). The actual dialog box depends on the type of the curve.

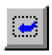
Parameter varying mouse mode*: opens window with sliders for each parameter of a curve. Parameters are changed dynamically while sliders are dragged.

Clipboard mouse mode*:

- moves the curves to the Xtremes clipboard window. When this mode is applied in the Xtremes clipboard window, the systems asks for a destination window;
- a curve can be directly dragged to a different plot window (also to a scatterplot window), if the left mouse button is kept pressed until the cursor is located in the destination window.

Deleting mouse mode*: deletes curves from a plot window.

Information mouse mode*: displays parameters of curve.



Coordinate changing mouse mode: adjust the coordinate system by pulling up a rectangle or, in the trivariate setup, rotate the coordinate system.

Point selection mouse mode: use this mode to cut off points in a bivariate scatterplot. Options of the local menu of a scatterplot do not use the cut points.

Line drawing mouse mode: adds straight lines to a plot window.

Label mouse mode: adds text labels to a plot. See page 15 for details.

Curve tabulating mouse mode*: the supporting points of a curve can be tabulated by storing them into a bivariate data set. For that purpose, adjust the coordinate system and the number of supporting points (enter the *Change Coordinates* box to adjust the range of the supporting points and use the option tool  to select the number of supporting points). We refer to Demo 4.14 for an application. Also see Demo 8.2.

Hotkeys of Xtremes

ALT+.	Activate univariate/multivariate mode.

ALT+D	Open the D(ISCRETE) domain.
ALT+S	Open the SUM domain.
ALT+M	Open the MAX domain.
ALT+P	Open the POT domain.
Backspace	Restore previous coordinate system in plot window.

Some hotkeys provide a shortcut to a menu option, e.g. pressing CTRL+D executes *Data... List Data*. These hotkeys are listed behind the corresponding menu option.

3.6 Microsoft Windows Commands

The following commands are common to all Microsoft Windows applications, yet their use is not very intuitive.

Operation in List Boxes

While a single item is marked by clicking it with the left mouse button, the selection of two or more items is achieved with one of the following options.

1. Tracking the mouse (i.e., pressing the left button and holding it down while moving the mouse cursor) selects all list items the mouse touches. To select two or more non–consecutive items, click onto them and hold down the Ctrl key while clicking.
2. If you utilize the keyboard, press the F8 key. Now it is possible to move the marker in the list box using the arrow keys while the space bar selects an entry.

Editor Fields

Text in an edit field (a single field for the entry of a parameter as well as the integrated data or StatPascal editors) is marked by either tracking the mouse or by using the arrow keys while holding down the Shift key. The following commands deal with marked text.

CTRL+Insert	Copy text to the clipboard.
SHIFT+Insert	Insert text from the clipboard. If some text is marked, then it is replaced by the inserted text.

Edit fields provide a local menu (available with a right click) where these operations are available. We also refer to the online help which can be activated by means of the F1 key or the *Help* menu.

Chapter 4

Handling and Visualizing Data

Xtremes is especially designed for exploratory data analysis within the extreme value setting, including a critical testing of the validity of parametric models by means of nonparametric tools, and provides some interactive highlights in that context.

In this chapter, we start with an overview of the data handling options. The following sections discuss details of discrete, continuous and time series data. At the end of the chapter, the file format of data sets and the StatPascal routines for creating data sets are discussed.

4.1 Overview of Data Handling Options

At the beginning, the user should restrict himself to handling data included in the package or generated by Xtremes. To read a data set from the disk, execute the menu option *Data... Read Data*. A file dialog box appears.

Proceed to the `dat` folder and select any file. Xtremes loads this file and opens a window entitled `Active Sample` displaying information about the data set. Read another data set and notice that the description of the active data set changes.

Now, two data sets are kept in memory. One can choose the active data set from the ones already loaded by executing the menu option *Data... Choose Data*. A list of all data sets used in the current session is displayed and a new active data set can be selected. All visualization and estimation procedures are based on the active data set.

Xtremes also enables the user generating data sets. Use the menu option *Data... Generate Univariate Data* and select a distribution from the menu. A dialog box opens asking for parameters, the sample size and a filename. Files are stored in the active directory.

After clicking OK the data set is generated and a short description appears in the `Active Sample` window.

Unlike most other Windows applications, Xtremes does not change its working directory when a data set (or a StatPascal program) is loaded from another directory. If one needs to load multiple data sets from a different subdirectory, the *Drag and Drop* facility of Xtremes can be utilized. Just select the files in the Explorer window and drop them anywhere on the Xtremes window.

To load your own data sets into Xtremes, several options are provided.

- Create a text file containing the data set (see section 4.5).
- Use the integrated programming language StatPascal to access your data and transfer it to the menu system (see section 4.6.).
- The CORBA server can be used to transfer data between Xtremes and other software systems like R or Microsoft Excel without generating any files. Consult the *RiskTec User Manual* for further information.

4.1.1 Visualization of Data

One must generate or read a data set from a file before employing an option from the *Visualize* menu. Some options may be applied to certain data types only. Xtremes disables the corresponding menu items if such an option is not applicable. One can change the active data set by means of the menu option *Data... Choose Data*.

A simple way to display data is in the form of a text. Load a data set or generate one using *Data... Generate Data* and select the menu option *Data... List Data*. Then, Xtremes opens a text window showing your data set. You can use the scroll bar to browse the data.

The *Visualize* menu contains options to display sample dfs, qfs, histograms, scatterplots, mean and median excess functions, among others. *Kernel Density* also provides options that reflect the data points at the right, left or both ends of the support. The bandwidth can be chosen by the user, an automatic selection (via cross–validation) is available.

The visualization options are also available in the local menu of a *List Data* window. They are applied to the displayed data set (rather than the active one) if selected from the local menu. An easy way to work with more than one data set is therefore to list them, minimize the windows and work with the local menus.

Time series are visualized by means of the scatterplot option. Note that each scatterplot is displayed in a separate window. You can cut points from a scatterplot using the point selection (scissors) mouse mode tool . The option *Least Squares Polynomial* in the local menu of the scatterplot window leads to a dialog box for polynomial regression.

The scatterplot option is also applicable to multivariate data. Depending on the active mode (univariate or multivariate), the user has to select two or three components. In the latter case, the points are displayed using a 3–D dynamic plot.

4.1.2 Applying Estimators to the Active Data

The three chapters in Part II of Statistical Analysis correspond to four different domains of Xtremes called *D(iscrete)*, *SUM*, *MAX* and *POT*. Each domain provides different distributions and estimators in the *Data... Generate Univariate Data*, *Distribution* and *Estimate* menu. One may switch between the different domains by means of the buttons D(ISCRETE), POT, MAX and SUM in the toolbar. In the following example, we focus on estimators in the POT domain because it provides the richest facilities.

DEMO 4.1. (Estimation Using the Hill Estimator.) To start, let us apply the Hill estimator to standard Pareto data.

(a) First, create a data set using *Data... Generate Univariate Data... Pareto(GP1).*

(b) Next, execute the option *Estimating... Hill(GP1/GP2)* . Recall that generalized Pareto models are fitted to the upper tail of the distribution. Therefore, the estimator requires the number k of upper extremes to be used for the estimation. You can change the number of extremes by clicking the up or down arrows in the estimator dialog box.

(c) A plot of $\hat{\alpha}_{n,k}$ or $\hat{\sigma}_{n,k}$ as a function in k is obtained using the diagram option. Choose the desired parameters before clicking the button.

Various parametric curves (plotted with the estimated parameter values) can be selected from the estimator dialog box. Comparing these curves with the corresponding nonparametric ones, the user is able to judge visually the quality of the estimation.

Similar dialog boxes are provided within the MAX and SUM domains. One can work with the other parts of Xtremes while estimator dialogs are open. It is also possible to use two or more estimators at the same time to compare their results.

Chapter 5 discusses the estimator dialog boxes in more detail.

4.2 Discrete Data

Xtremes can handle different data types as, e.g., discrete, univariate or grouped data and time series. The visualization options in the menu system depend on the type of the active data set. Conversions between different data types are available by the *Data... Convert to* facilities. Section 4.5 contains a complete reference list of the different data types. For operations within one data type, use the *Data... Transform to...* facilities.

4.2.1 Generating Discrete Data

To generate a discrete data, use the menu option *Data... Generate Discrete Data* and select a distribution from the menu (make sure that the D(iscrete) domain is active). A dialog box opens asking for parameters, the sample size and a filename. Files are stored in the working directory (usually, `\xtremes`) if no directory is specified.

After clicking OK, Xtremes generates the data set and stores it in the selected file. In the Active Sample window, one may also use a local menu which can be activated by a right mouseclick onto the window.

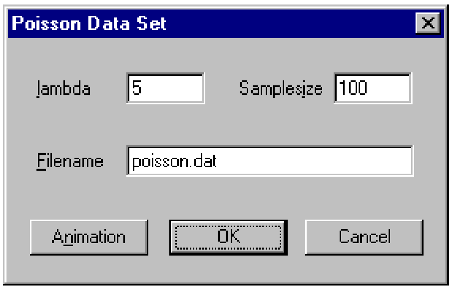


FIG. 4.1. Dialog box *Data... Generate Discrete Data... Poisson.* Select the parameter λ of the Poisson distribution and a filename. Execute OK.

DEMO 4.2. (Generating Discrete Data.) Select, e.g., the menu option *Data... Generate Discrete Data... Poisson* to open the dialog box shown in Fig. 4.1. Click OK to generate the data set. Discrete data sets are stored as pairs (x_i, n_i) of reals x_i and integers n_i, if the data point x_i occurs n_i times.

Facilities of a didactic nature are included which can be applied in elementary statistics courses. Use the *Animation* option shown in Fig. 4.1.

4.2.2 Loading and Editing Discrete Data

To read a data set from the disk, execute the menu option *Data... Read Data.* A file dialog box appears, where one can select any file. A collection of example data sets is stored in the `dat` subdirectory.

DEMO 4.3. (Editing Discrete Data.) Data sets are stored as plain ASCII files. Certain specifications may be given at the top of the file, such as the type of the data set and the sample size. Moreover, one may include a shorter and a more detailed description. For that purpose, activate the editor  in the toolbar and load the generated data set from the file as, e.g., from `poisson.dat` in the working directory `xtremes`. The operation in Demo 4.2 resulted in a data set of the following structure.

```
Xtremes Discrete Data
Type: Poisson Data (lambda = 5)
\begin(description)
Data set generated by Xtremes according to a Poisson
```

```
distribution with parameter 5.
\end(description)
SampleSize: 12
1    3
2    13
3    9
...
```

4.2.3 Plotting a Sample Histogram

Let the active data set be of type Xtremes Discrete Data as described in Demo 4.3. A sample histogram may be plotted by employing the *Visualize* menu.

DEMO 4.4. (Plotting a Sample Histogram for Discrete Data.) Fig. 4.2 shows the sample histogram of the data set generated in Demo 4.2, which can be plotted by executing the menu option *Visualize... Histogram.*

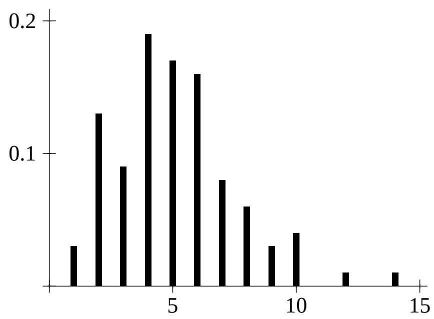


FIG. 4.2. Sample histogram of discrete data set generated under a Poisson distribution with parameter $\lambda = 5$.

The histogram option is also applicable to grouped data (see Demo 4.7).

4.2.4 Fitting a Parametric to a Sample Histogram

One may jointly plot a parametric and a sample histogram into the *Density* window. Subsequently, we use this approach to fit a Poisson distribution to a data set. This can be done by changing the parameter of the plotted Poisson histogram by using a slider.

DEMO 4.5. (Fitting a Poisson Distribution to a Discrete Data Set.) The following table shows the number j of cavalrymen killed in a regiment by horse–kicks within one year in ten particular regiments of the Prussian army over a period of twenty years yielding $n = 200$ observations (due to von Bortkiewicz, 1898).

number of victims j	0	1	2	3	4	≥ 5
frequency $n(j)$	109	65	22	3	1	0

Load the data set (stored in `horse.dat`) and display a sample histogram of values $p(j) = n(j)/n$. Proceed to the menu option *Distribution... Poisson* and add a Poisson histogram with parameter $\lambda = 1$ to the plot.

Use a slider (cf. Demo 3.3) to determine visually a Poisson histogram which fits to the sample histogram. Fig. 3.2 shows a screenshot of the procedure. Estimators for discrete distributions are provided in the Estimate menu when the DIScrete domain is active.

4.3 Continuous Data

Next, we deal with the options provided to generate and visualize continuous data.

4.3.1 Generating Continuous Data

Select the menu option *Data... Generate Univariate Data* within the domains SUM, MAX or POT to generate continuous data. The user is offered the distributions which are also available in the *Distribution* menu. A dialog box opens when an option is executed.

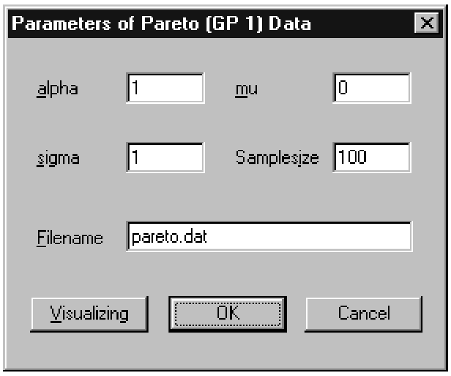


FIG. 4.3. Dialog box *Data... Generate Univariate Data... Pareto*. Select the parameters of the distribution and a filename. Execute OK.

In addition to the parameters of the distribution, one must specify the sample size and a filename under which the data set will be stored. The generated data set becomes the active one which can be analyzed by Xtremes. This data set is of the type Xtremes Univariate Data.

DEMO 4.6. (Inspecting Generated Pareto Data.) First, generate data as it was described in the preceding lines (cf. Fig. 4.3).

(a) A window, entitled *Active Data*, opens providing information about the generated data. To inspect the data, click onto the *Active Data* window with the right mouse button to enter a local menu and choose the *List Data* option. Then, the data are displayed in the order of their outcome.

(b) Activate the editor  in the toolbar and load the generated data set from the file `pareto.dat` in the working directory `xtremes`. You will see headlines of the following form

```
     Xtremes Univariate Data
Type: Pareto (GP 1) Data (alpha = 1, mu = 0, sigma = 1)
\begin(description)
Data set generated by Xtremes according to ...
\end(description)
Samplesize: 100
```

followed by the data in one column.

4.3.2 Visualization of Continuous Data

Nonparametric procedures can be combined with procedures from the *Distribution* menu. For example, plot a parametric and a sample density into the *Density* window.

DEMO 4.7. (Plotting a Histogram for Continuous Data.) (a) Generate 100 standard normal data and plot the pertaining density (cf. Fig. 4.4).

(b) Transform the data to Xtremes Grouped Data with respect to the grid $t_j = j$. Using *Data... Convert to... Grouped Data* one opens a dialog box. An edit field is used to define the bin boundaries. Equally–sized cells may be obtained by means of the *from*, *to* and *step width* fields. Take the values -4, 4 and 1.

(c) Plot the pertaining histogram employing the option *Visualize... Histogram*.

(d) Employ the options mouse mode from the toolbar to vary the style of the histogram.

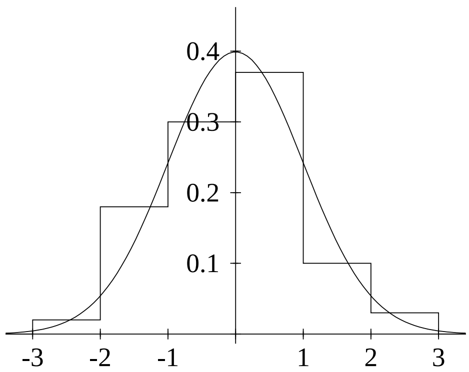


FIG. 4.4. Standard normal density and histogram based on 100 standard normal data.

DEMO 4.8. (Changing the Bandwidth using a Slider.) After executing the previous demo, the active data set contains a grouped version of the standard normal data generated in step (a). Use *Data... Choose Data* to activate the univariate data set and exectute the option *Visualize... Kernel Density.* The technique used in the Demos 3.3 and 3.4 can also be utilized to change the bandwidth of a kernel density. The illustration in Fig. 3.3 was produced in this manner.

One of the most important, also complicated, questions of extreme value analysis is to fit a GP df (qf, density) to the sample df (qf, density) based on a data set that contains

- the original observations,
- the upper order statistics.

The estimated df (qf, density) depends on the respective data set.

DEMO 4.9. (Fitting a GP Distribution to Original Observations and the Upper Order Statistics.) Activate the MAX domain and generate 400 EV data by executing *Data... Generate Univariate Data... EV...* .

(a) Plot a kernel density by means of *Visualize... Kernel Density...* . Activate the POT domain and estimate a GP density by means of the MLE(GP) based on the 60 upper extremes. Plot the pertaining density. One gets a plot as in Statistical Analysis, Fig. 2.9 on page 58, left–hand side.

(b) Order the data according to their magnitude and store the 60 largest data in another file. Repeat the procedures in (a). Now one gets a plot as in Statistical Analysis, Fig. 2.9, right–hand side.

4.4 Time Series

Scatterplots can be produced for time series and multivariate data (apply *Visualize... Scatterplot*), whereby two components must be selected in the latter case.

Options for a time series analysis are available in the local menu of the scatterplot window (right mouseclick into the plot window). These include polynomial regression, moving averages (with user–defined filters), estimation of seasonal components and sample autocovariance and autocorrelation functions.

Each scatterplot window displays only one data set. If a menu command is applied which transforms the data set, the result is displayed in a new window (e.g., by calling a moving average one gets the smoothed data set and the residuals). Apply the *Save Actual Points* option in the local menu to store these values into a file.

Please note that the visualization options for scatterplots are more limited that the ones for other curves. However, you can convert a scatterplot to a normal curve by copying it to the Xtremes Clipboard with the Clipboard Mouse Mode.

The menu system (*Data... Generate Time Series*) provides options to generate time series:

- MA(q): enter the coefficients of the MA polynomial and a UserFormula expression for a qf of the innovations.

- AR(1): Gaussian AR(1) series are available.
- ARMA(p, q): Gaussian ARMA(p, q) series are simulated by using the innovations algorithm. The coefficients of the AR and MA–polynomials must be provided by the user.

Transformations of these processes are obtained by means of the UserFormula facility. Other processes may be generated by using the integrated programming language StatPascal.

Demo 4.10. (Paretian Time Series with Gaussian AR(1) Dependence Structure.) An application of the UserFormula facility is given. Generate a Gaussian AR(1) sequence (i, x_i), $i = 1, \ldots, 1000$, for $d = 0.7$ (*Data... Generate Time Series... Gaussian AR(1)*). Apply

```
3 + 4*paretoqf(2, gaussiandf(x))
```

in *UFO... Transform Data* to the active data set, which will result in

$$(i,\, 3 + 4 * W^{-1}_{1,2,3,4}(\Phi(x_i))), \qquad i = 1, \ldots, 1000.$$

Thus, one obtains a stationary sequence with Pareto $W_{1,2}$ marginals and a dependence structure inherited from the Gaussian AR(1) sequence.

Several estimators for AR and ARMA processes are available in the Xtremes menu.

Demo 4.11. (Estimators in ARMA–Models.) Enter the ARMA(p, q)–dialog box in the *Estimating* submenu within the SUM domain .

(a) (AR(p): Yule–Walker.) In the dialog box, one must specify the order p of the AR polynomial. The order of the MA polynomial must be zero. Select *Yule-Walker AR(p)* and execute *Estimate*.

(b) (ARMA(p, q): Hannan–Rissanen.) Specify the orders p and q of the AR and MA polynomials.

(c) (ARMA(p, q): Innovations Algorithm.) Specify the orders p and q of the AR and MA polynomials.

(d) (ARMA(p, q): MLE.) First execute one of the preceding estimators. If the initially estimated process is not causal, then the MLE is not applicable and an error message appears. The MLE is numerically computed by means of a Newton-Raphson procedure. Press the *MLE ARMA(p, q)* button in the dialog box to execute the option.

Demo 4.12. (Generating a Poisson Process with GP Marks.) Activate in the POT-menu a dialog box by *Data... Generate Counting/Poisson Process... Marked Poisson Process* and select the intensity of the Poisson process and the parameters of the GP marks. Store the times and the GP marks in a file. The data may be visualized in a scatterplot.

Demo 4.13. (Time Series of Maximum Daily Temperatures in Death Valley.) (a) Use *Data... Read Data* to load the file em–deval.dat from the subdirectory `xtremes\dat`.

(b) Use *Data... Convert to... Time Series* to extract the first two columns entitled Number and Max. Fig. 2.14 on page 68 in Statistical Analysis shows a scatterplot of this time series.

In the subsequent demo, the missing 31 values for Dec. 1994 of the Death Valley data are filled using a local polynomial regression approach. In this demo, we also learn to adopt the curve tabulating mouse mode tool which saves the supporting points of a curve into a file.

DEMO 4.14. (Tabulating a Curve and Filling a Gap.) (a) Display a scatterplot (*Visualize... Scatterplot*) of the time series in Demo 4.13 and use the scissors tool (point selection mouse mode) to cut off all points except about 30–40 around the gap. Fit a quadratic least squares polynomial (local menu in the plot window). Next, adjust the coordinate system to the interval $123 \le x \le 153$ and use the option mouse mode tool to select 31 supporting points for the polynomial.

(b) The curve tabulating mouse mode tool saves the supporting points of the least squares curve into a file.

(c) Finally, merge them into the original data set utilizing two editor windows and the clipboard.

4.4.1 Clustering of Data

The special option *Visualize... Clusterplot* is implemented which enables the visualization of clustered Xtremes Time Series. First a scatterplot window opens. A SHIFT + left mouseclick into the scatterplot window creates a threshold at the position of the cursor. The exceedances and exceedance times i/n become visible (as in Fig. 2.21 in Statistical Analysis on page 78). In addition, three separate windows open for

- the cluster size distribution with respect to the given threshold u,
- a plot of the mcsizes for thresholds $v \ge u$, and
- a plot of the reciprocal mcsizes (extremal indices) for $v \ge u$.

DEMO 4.15. (Using the Xtremes Clipboard Window.) We are going to generate an illustration as given in Fig. 2.23 in Statistical Analysis on page 79.

(a) Generate 4000 Gaussian AR(1) data for the correlation coefficient $d = 0.8$ in the SUM domain (*Data... Generate Time Series... Gaussian AR(1)*). Execute *Visualize... Clusterplot*. Use options in the local menu of the clusterplot window. Create a threshold at $u = 1$ by a SHIFT + left mouseclick at the position of the cursor.

(b) Activate the clipboard mouse mode and click onto the plot of the reciprocal sample mcsize. This plot appears in the clipboard window. Export this plot to another window by the same procedure.

4.5 Format of Data Sets

To load your own data into Xtremes, it suices to create a plain ASCII ile with the rows and columns of your data. Certain specifications can be given at the top of the file, such as the type of the data set and the sample size. Moreover, one may include a shorter and a more detailed description. While these provisions make a data set more accessible within the menu system, one should note that they are optional (see also *Data Sets without Headers* at the end of this section).

Data sets can be entered by utilizing any text editor available on your operating system. It is possible to use the integrated editor, yet one should be aware of the fact that some older versions of Windows (in particular, 95 and 98) limit the text size of the editor to 64 KBytes. Under Linux and Windows NT/2000/XP/2003, text files of arbitrary size can be handled.

We start with an example showing the data entry using the integrated editor. Suppose you want to create a univariate data set with the following values: 1, 3.5, 7, -4.

Start the editor by selecting the editor button in the toolbar and click on the *Header* button in the toolbar of the editor window. A dialog box asking for the type of the data set opens. Select *Univariate Data* to create a template of a univariate data set and fill in the following fields:

```
Xtremes Univariate Data
Type: Artificial example
\begin(description)
This is an artificial data set. It
was entered using the integrated editor.
\end(description)
Sample Size: 4
1
3.5
7
-4
17
```

The first line defines the type of the data set—in the present case Xtremes Univariate Data. A list of all types is given below. The second line starts with `Type:` and provides a short description which will be shown in the list of loaded data sets (*Data... Choose Data*). It is also added to the description of curves based on this data set. The description must be restricted to one line.

Between the lines `\begin(description)` and `\end(description)`, a longer description may be added. It is displayed in the `Active Sample` window. The next line determines the size of the data set.

Then, the data are listed, one value for each line. After having typed the text, save it to a file (e.g., in the subdirectory `\dat`). Afterwards, your data set becomes the active one. One may also simulate a data set of the desired type using the option *Generate...* .

Xtremes supports the following data types.

- **Xtremes Univariate Data.** Real data $x_1, \ldots, x_n$ in any order, as presented above. Execute *Data... Transform Data... Sort* to sort these data according to their magnitude.
- **Xtremes Grouped Data.** Pairs (t_j, n_j) of edges t_j and frequencies n_j of data in cells $[t_j, t_{j+1})$ as, e.g.,

4.3	17
5	2
.	.
.	.
17.3	0

  if 17 data points belong to the cell $[4.3, 5)$ and the largest edge is 17.3. Omit this last line if the largest cell is unbounded, yet, internally, the edge $t_{n+1} = 2t_n - t_{n-1}$ is added as largest edge. The term `Sample Size` is used here for the number of specified cells.
- **Xtremes Discrete Data.** Pairs (x_i, n_i) of reals x_i and integers n_i, if the data point x_i occurs n_i times. This data type is suitable for samples drawn from discrete distributions as, e.g.,

0	7
1	17
2	25
3	2

  `Sample Size` denotes the number of discrete points included.
- **Xtremes Time Series.** Pairs (i, x_i) of integers i and reals x_i as, e.g.,

1	17.5
2	-2
3	0.34
4	0.001

  The discrete time must be given in increasing order. Some of the pairs (i, x_i) can be omitted (see, e.g., ct–sulco.dat), so that the entry `Sample Size` is not necessarily the number of data points within the file. It may be larger than the time of the last point if values were omitted at the end of the file.
- **Xtremes Censored Data.** Pairs (z_i, δ_i) of reals z_i and $\delta_i \in \{0, 1\}$ showing whether the data point z_i is censored ($\delta_i = 0$) or uncensored ($\delta_i = 1$), e.g.,

1.7	1
2.1	0
1.5	1
0.2	1

- **Xtremes Multivariate Data.** Multivariate data $(x_{i,1}, \ldots, x_{i,m})$ are stored using m entries on a line.

  Moreover, the line after `Sample Size` contains an entry defining the dimension m of the data set. It is followed by m names surrounded by quotation marks. They define the headers for the corresponding column, e.g.,

  Sample Size: 12
  Dimension: 4

"Month"	"SO2"	"NO"	"O3"
1	75.2	13.4	17.2
2	83.1	17.9	15.4
3	.	12.8	11.3
4	43.9	15.3	11.3

  Missing values are indicated as a dot. It is possible to combine related univariate data sets of different length to one multivariate data set. The rows containing a dot are ignored when the multivariate data set is transformed or converted.

- **Data Sets Without Header.** Xtremes can also load plain ASCII files containing just a matrix of data, without any headers. Such data sets are treated as multivariate. Moreover, one can use decimal points or decimal commas within a data set.

Data can be converted from one type into another by the option *Data... Convert to.* All canonical conversions are available. There are also some special conversions.

One can apply the UserFormula facility (see chapter 8) to perform transformations not covered by the menu system. More sophisticated conversions are accomplished by means of StatPascal programs.

4.6 Generating Data Sets from StatPascal

Data sets can be generated from the included programming language StatPascal. We merely give an overview of the predefined procedures which transfer data from StatPascal to Xtremes. For further information, the reader is referred to the *StatPascal Manual*. One may create data sets of the following types.

Xtremes Univariate Data: data $x_1, \ldots, x_n$ are collected in a real vector given as argument to the call of the predefined procedure *createunivariate*.

```
var x: vector of real;
...
createunivariate (x, 'filename.dat', 'Description');
```

Instead of a vector, a one–dimensional real array may be given as well.

Xtremes Time Series: in addition to the previous case, a vector containing the times t_i of the observations must be provided.

```
var   x : vector of real;
      t : vector of integer;
...
createtimeseries (t, x, 'filename.dat', 'Description');
```

Xtremes Censored Data: besides a real vector containing the censored data, there is an integer vector with the censoring information.

```
var   z : vector of real;
      delta : vector of integer;
...
createcensored (z, delta, 'filename.dat', 'Description');
```

Xtremes Multivariate Data: the data $x_{i,j}$ are collected in a real matrix. In addition, a string with the column names, separated by '|', must be provided.

```
var x: matrix of real;
    h: string;
...
h := 'Day|Month|...';
createmultivariate (x, h, 'filename.dat', 'Description');
```

Note that a two–dimensional array can be provided instead of a matrix type, because the language supports an implicit type conversion from two–dimensional arrays to matrix types.

As a result of such a procedure you will get an active data set of the type as specified by the command `create...` The Active Data window opens showing the name of your data set and the description provided in the last argument.

Chapter 5

Univariate Estimators and Test Procedures

5.1 Gaussian and Related Models

We present the estimator and test dialog boxes in the univariate SUM domain which provides estimation procedures for the normal (Gaussian) and certain discrete models (Poisson, binomial and negative binomial). Similar boxes that are available in the other domains will be described in the subsequent sections. First, activate the SUM option in the tool bar.

5.1.1 Estimators

The *Estimate* menu provides parametric procedures for univariate and discrete data sets. The estimated values are displayed in a dialog box. Options of the dialog box allow for the plotting of parametric curves based on the estimates; the Q–Q plot option provides a quick check of the validity of the model.

The local menu of the estimator dialog box enables the user to calculate a T–year threshold that is the $(1 - 1/T)$–quantile of the estimated distribution in the present context. A parametric bootstrap option may be applied to obtain confidence intervals for the underlying parameter as well as an estimate of the MSE of the selected estimator.

Demo 5.1. (Visualizing the MLE for the Normal Model.) In Example 3.1.1 in Statistical Analysis on page 85, the MLE in the normal model was applied to Michelson's light speed data, and the pertaining Q–Q plot and normal density were plotted in Fig. 3.1 on page 86.

(a) Read the file nu–miche.dat and execute the kernel density option in the *Visualize* menu.

(b) Next, select the *MLE(Gaussian)* option in the univariate SUM domain. Xtremes opens the estimator dialog box shown in Fig. 5.1. Disable *Close before plotting* and select the option *Density* to add the estimated parametric density to the kernel density. Clicking on the Q–Q plot button produces the Q–Q plot.

(c) Open the local menu by means of a righ–click onto the dialog box.

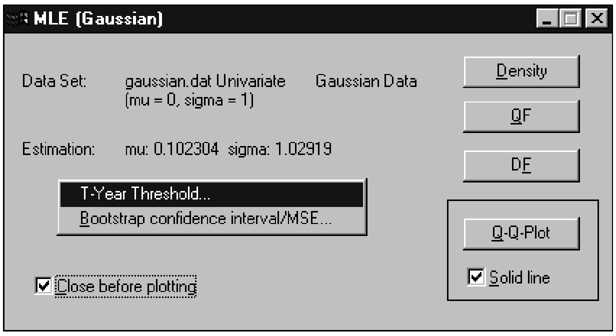


FIG. 5.1. MLE for normal distribution. The box displays the estimated values and offers the plot of parametric curves.

5.1.2 Test Procedures

Test procedures are available in the *Testing* menu. Xtremes offers the tests introduced in this chapter, namely one–sided tests of the mean of the normal distribution with known or unknown variance, as well as the χ^2 and LR–tests for the Poisson distribution. Tests in the Poisson model require the grouping of the data in the test dialog box. Each test displays the p–value.

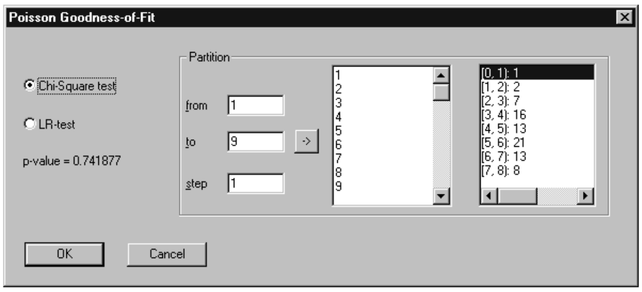


FIG. 5.2. χ^2 and LR–test (Poisson). The p–value for the specified cell boundaries is displayed.

DEMO 5.2. (χ^2–Test.) We apply the χ^2–test for the Poisson distribution to a data set drawn according to a Poisson distribution with parameter $\lambda = 2$. Use the menu option *Data... Generate Discrete Data... Poisson* in the univariate DISCRETE domain to simulate the data set. A test dialog box (see Fig. 5.2) is opened using *Testing... Poisson*.

The dialog box offers the choice between χ^2 and LR–tests. An edit field is used to define cell boundaries; equally–sized cells may be obtained by means of the *From*, *To* and *Step* fields. Xtremes displays the observed frequencies within the cells and the p–value.

5.2 Extreme Value Models

We describe the estimator and simulation options of the menu system for EV distributions.

Estimator dialog boxes in the EV model are similar to those in the Gaussian model. Remember that Fréchet (EV1) and Weibull (EV2) distributions can also be represented using the γ–parameterization. Therefore, each estimator dialog box provides radio buttons to switch between the α and γ–mode.

To some estimators of the shape parameter (like the LRSE), least square estimates for the scale and location parameters are added. One may exchange these least square estimators or define new estimators by applying the integrated programming language StatPascal. Consult the *StatPascal Manual* for further information.

DEMO 5.3. (Applying the MLE(EV) to Maximum Temperatures.) In Example 4.1.2 on page 113 in Statistical Analysis, an EV distribution is fitted to the maximum September temperatures in de Bilt. The file em-dbilt.dat provides a multivariate data set with columns containing the maximum for each month in the year. Use the option *Data... Convert To... Univariate* to select the September data. Apply the MLE(EV) in the MAX domain. The Q–Q plot is available from the estimator dialog box.

A right mouseclick in the estimator dialog box opens a local menu which provides estimates of certain functional parameters, confidence intervals, etc.

5.2.1 Simulations

The *Simulate* menu provides options to simulate the MSE/Bias and the distribution of an estimator.

A data set of estimates $\gamma_{n,1}, \ldots, \gamma_{n,N}$ (respectively $\alpha_{n,1}, \ldots, \alpha_{n,N}$) may be generated under *Simulate... Distribution* in the univariate EV domain. One must specify the parameters n, N and a distribution.

The distribution must be selected by means of the UserFormula facility. Specify a qf (e.g., `paretoqf (2, x)`) to generate standard Pareto data under the shape parameter $\alpha = 2$.

The *Simulate MSE/Bias* dialog box provides options to simulate the MSE and bias of estimators. The estimates are based on a sample of maxima from k blocks of length m.

Note that k or m may be chosen as a running parameter. The result of the simulation is stored as a trivariate data set with columns containing the running parameter, the simulated MSE and the bias.

Fig. 5.3 shows the dialog box *Simulate MSE/Bias*. Create a scatterplot to plot the simulated MSE or Bias against the running parameter.

DEMO 5.4. (Simulation of MSE and Bias of an Estimator.) Table 4.2 on page 112 in Statistical Analysis shows simulated values of the MSE of the MLE(EV) and LRSE(EV).

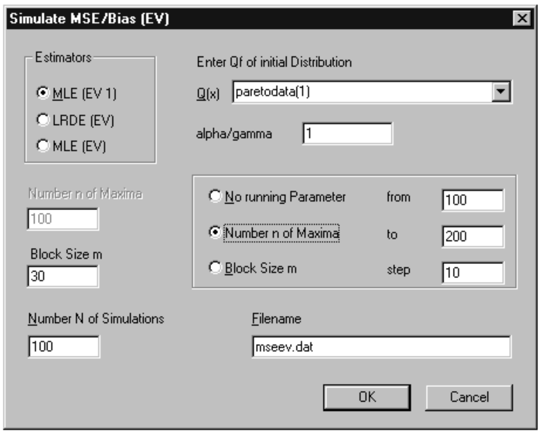


FIG. 5.3. The *Simulate MSE/ Bias* dialog box. A trivariate data set with the running parameter and the simulated MSE/Bias is generated.

To obtain the first entry (MLE(EV) applied to 20 observations generated under Φ^{30}), choose the option *Simulate... MSE/Bias* in the univariate MAX domain and select the parameters $n = 20$ and $m = 30$ in the dialog box. Moreover, specify *alpha/gamma* $= 0$, and enter the expression `gaussianqf(x)` as a qf.

5.2.2 Units Exposed to Corrosion

We shortly indicate the steps which are required to deduce the service life for steel tanks in Example 18.1.2 in Statistical Analysis on page 443.

DEMO 5.5. (a) (Estimating $T = 100$ Unit Pit Depth by Means of the MLE(EV).) Load the data set gu-pit4.dat from the subdirectory `\dat`. This data set becomes the active one. Within the univariate MAX domain execute *Estimating... MLE(EV)*. Within the estimate dialog box open another dialog box by means of a right mouse–click. Determine the $T = 100$ unit value for the estimated parameters. Repeat this procedure for the data sets gu-pit6.dat and gu-pit12.dat.

(b) (Storing the 100–unit Depths in a File.) Activate the data editor in the toolbar and create a data set of type Xtremes Multivariate Data corresponding to the procedure in Demo 4.6. Edit the given data in the following manner

```
Samplesize: 3
Dimension: 2
"year"  "100-unit depth"
4         2.43
6         2.84
12        11.2
```

Store these data in a file. This data set becomes the active one.

(c) (Fitting Least Squares Lines.) Visualize these data by executing *Visualize... Scatterplot*. In the local menu of the Scatterplot Window (activated with a right mouse–click in the plot window) choose *Least Squares Polynomials*. Plot least squares lines of order 1 and 2. Add a constant line with value $l = 5$ and determine the pertaining service life.

5.2.3 Gompertz model

We want to demonstrate in which manner one can visually produce the illustrations in Fig. 19.1 on page 455 in Statistical Analysis. For that purpose we also make use of the UFO facility because Gompertz densities etc. are not included in the menu system. Yet, the Gumbel density is available as a menu option and, therefore, available as a function in UFO (as well as in StatPascal).

DEMO 5.6. (Fitting a Gompertz Density to the Upper Tail of a Histogram.) First load the data set um-lspge.de of type Xtremes Multivariate Data from the subdirectory `\dat`.

(a) (Transforming the Data.) Use *Data... Convert Data... Grouped Data...* to select the first column and another one from the multivariate data set. The result will be a data set of type Xtremes Grouped Data which is now the active data set.

(b) (Plotting a Histogram.) Execute Visualize... Histogram to plot the histogram. If not otherwise selected the plot appears in a plot window entitled Density.

(c) (Plotting and Fitting a Gompertz Density.) Open the UFO facility by clicking on the UFO button in the toolbar and select the option *Plot curve* from the popup menu. Now, type the formula

```
gumbeldensity(-(x-p1)/p2)/p2
```

in the edit field labeled *f(x) or f(p,x)*. Here $p1$ and $p2$ are the location and scale parameters of the Gompertz density which can be varied afterwards by means of sliders. Choose a support from -50 to 120, for example. Open the parameter varying mouse mode and click on the plotted Gompertz density in the Density plot window. A dialog box with sliders for the parameters $p1$ and $p2$ opens. Vary the parameters to fit the Gompertz density to the upper tail of the histogram.

5.3 Generalized Pareto Distributions

The dialog boxes for the GP models are similar to those in the MAX domain. The main difference is the inclusion of the number k of upper extremes as another parameter.

5.3.1 Estimators

Estimator dialog boxes in the POT domain include all the options available in the EV models. Additional dialog elements allow

- to select the number k of upper extremes,
- to plot estimates against the number k of extremes (diagram option).

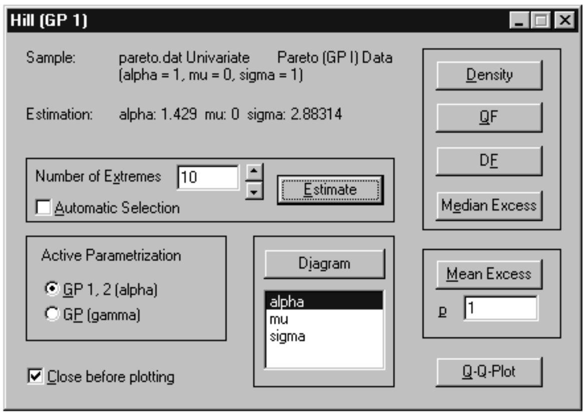


Fig. 5.4. The Hill estimator dialog box. The box provides options to plot curves based on parametric estimates; the selection of the number of upper extremes is a special feature of estimators in the POT domain.

Fig. 5.4 shows the dialog box for the Hill estimator.

Demo 5.7. Read su–tvcon.dat. Apply the option *Bounded Kernel Density* with *Left*. Investigate the performance of the *Moment(GP)* and *Drees–Pickands(GP)* estimators. In each case, plot the *Diagram* for *gamma*. Then, carry out the estimation of γ for $k = 40$. Plot the densities pertaining to the estimated parameters.

The dialog box for the Bayes estimator in the GP1(u) model (see page 132 in Statistical Analysis) corresponds in the upper part to the dialog boxes of the other estimators in that model. In particular, the estimates are displayed in the (α, μ, σ)–parameterization.

Note that the Bayesian calculations are internally carried out in the (α, η)–parameterization. In particular, the priors and posteriors are specified in this parameterization. In the lower part (see Fig. 5.5) of the dialog box one may select the gamma prior for α and the reciprocal gamma prior for η.

Demo 5.8. (Bayes Estimation in the Full Pareto Model.) To carry out the Bayesian estimation in the full Pareto model select *Estimating... Bayes GP1* in the POT domain. Select a prior for α and η. Also plot the priors and posteriors for these parameters by executing the options *Prior* and *Posterior*.

5.3.2 Simulations

In contrast to the simulation facility in the preceding chapter, it is possible to choose the number k of extremes as a running parameter (instead of the block size m). One must be aware that some estimators, especially the MLE(GP), may fail for smaller sample sizes. Such samples are discarded. A table, added to the description of the simulated data set, shows the number of failures for each value of the running parameter.

We now demonstrate the computations which led to to the plots in Fig. 5.1 on page 132 in Statistical Analysis.

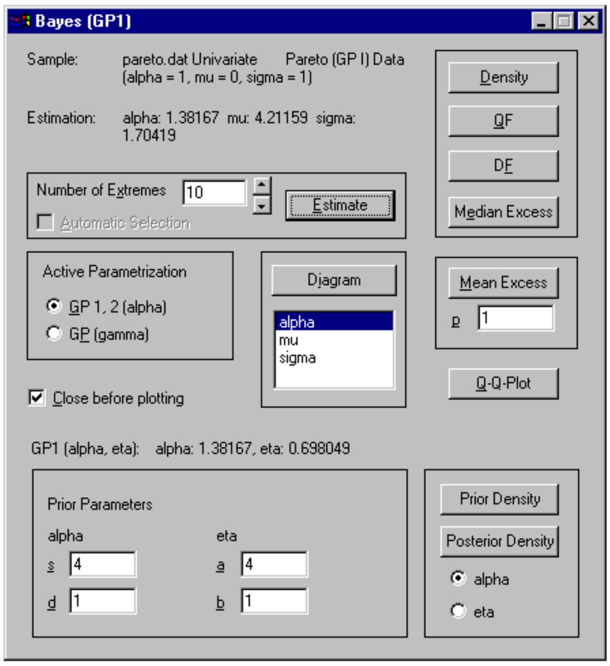


FIG. 5.5. The Bayes GP1 dialog box. The box provides additional options for selecting prior parameters etc.

DEMO 5.9. (Simulating the Density of the Hill Estimator.) Within the POT–domain execute *Simulate... Distribution of Estimator...* and type

```
-1 + paretoqf(10,x)
```

or

```
paretoqf(10,x)
```

into the input field. Choose n, k, and N and the Hill estimator option and execute OK. A data set of N Hill estimates is stored in the selected file. Now use the Visualize menu to generate a density.

5.3.3 Censoring

A special data type is provided for storing censored data. If a nonparametric estimation procedure (i.e. kernel density, sample df or sample qf) is applied to a censored data set, the Kaplan–Meier estimator is utilized. The MLEs for censored data are available within the GP1 and GP models.

One may simulate a censored data set by means of the menu option *Data... Convert To... Censored Data*. The option censors the active univariate data set by using the censoring distribution entered in the dialog box.

DEMO 5.10. (MLE(GP) for Censored Data.) Read the heart transplantation data from mc–heart.dat with *Data... Read Data* and display the data by *Data... List Data*.

(a) Plot the Kaplan–Meier estimate with *Visualize... Sample DF*.

(b) Execute *Estimate... MLE(GP)* and plot the diagram for the parameter γ. Discuss the performance of the MLE(GP) for censored data.

Chapter 6

Multivariate Estimators

6.1 Plot Options

Xtremes also provides some visualization tools for multivariate data and curves. We start with the Demos 6.1 and 6.2 which concern the plotting of a bivariate kernel densities and a scatterplot for trivariate data using the menu system of Xtremes.

DEMO 6.1. (Selection of Direction and Marginal Bandwidths for Bivariate Kernel Density.) (a) Generate 100 bivariate Gaussian data in the multivariate SUM domain (select and execute *Sum... Data... Generate Bivariate Data... Gaussian*) for the correlation coefficient $\rho = 0.8$.

(a) Plot a kernel density for the active data (*Visualize... Kernel Density...*) with the default parameter and bivariate normal kernel. Adjust the coordinate system (right mouseclick into the plot window and select *Change Coordinates*). Activate the coordinate changing mouse mode or press the Ctrl key. Rotate the coordinate system around the z–axis by dragging the mouse with pressed left mouse button.

(c) Adjust the bandwidths and direction of the kernel (select the parameter varying mouse mode and click onto the plot to open the window with sliders).

DEMO 6.2. (Using the 3–D Scatterplots.) (a) Generate 50 trivariate Gaussian data in the multivariate SUM domain (*Data... Generate Multivariate Data... Trivariate Gaussian...*).

(b) Visualize the data in a 3–D scatterplot (*Visualize... Scatterplot*). Several options are available in the local menu (right mouseclick into the plot window). Select an interval for the point magnification (which decreases for increasing distances between points and the viewer) by means of the *Scatterplot Options.* The *Change Coordinates* box offers the option *z–axis locked.*

(c) Deactivating this option, one can rotate the coordinate system not only around the z–axis, but also around any axis which is orthogonal to the direction in which the mouse is moved (for that purpose, activated the coordinate changing mouse mode or press simultaneously the CTRL key).

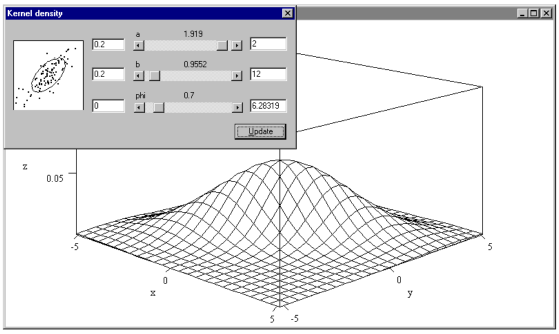


FIG. 6.1. Kernel density for 100 bivariate Gaussian data with correlation coefficient $\rho = 0.8$.

(d) The *Projection Options* control the viewpoint and allow the user to toggle between a perspective and a parallel projection.

Fig. 11.2 and 11.3 on page 288 in Statistical Analysis contain plots of bivariate sum–stable densities and their contour plots which were produced with the help of MVSTABLE[1]. It it evident that Fig. 11.2 can also be produced by applying the UFO facility. We describe MVSTABLE procedures and apply Xtremes to the output of MVSTABLE.

Upon execution, the MVSTABLE package displays a menu, where one can evaluate values of bivariate stable densities on a grid, generate data or fit a bivariate stable distribution to data.

DEMO 6.3. (Plots of Bivariate Stable Densities.) After having selected the first option, MVSTABLE asks for the name of a configuration file

(b) (Configuration File.) containing the following information: (1) the shape parameter α, (2) a specification of the spectral measure Λ, whereby MVSTABLE supports discrete measures defined by listing single points and their mass as well as (scaled) uniforms on the sphere, (3) a shift vector, (4) the definition of a grid, where the density is evaluated. The configuration file for the density in Fig. 11.3 on page 288 in Statistical Analysis is shown in the following lines.

```
0.8
4
3
0   .125
120 .25
300 .25
0    0
0   2  .1
-2   2  .1
```

[1]See `http://www.mathstat.american.edu` and select John Nolan's homepage from the People section of the website.

It uses shape $\alpha = .8$ and employs a discrete non–symmetric spectral measure (indicated by the 4 in the second line) with 3 points. The points carry the masses 0.125 and .250, their positions on the unit sphere are specified in degrees (0, 60, 300). The shift vector is set to zero. Finally, the start and end value as well as the step size for the grid is specified, where the density is evaluated.

(b) (Output.) A trivariate data set with the grid points and the pertaining value of the density is generated in the protocol file mvstable.out. One can utilize, e.g., Xtremes to produce a contour plot or a surface plot from that data set (cf. next demo).

From a trivariate data set, where the first two columns contain the points of a grid (with increasing values of x and y) and the third one the pertaining value of the bivariate function, we compute contour and surface plots.

DEMO 6.4. (Contour and Surface Plots.) Load a file as indicated in the preceding lines (e.g., mvstable.dat in `\dat`) with the *Data... Read Data* option and apply the *Visualize... Surface Plot* or *Visualize... Contour Plot* option in the multivariate mode.

6.2 Multivariate Extreme Value Models

We include two demonstrations which concern the estimation in the Marshall–Olkin and Hüsler–Reiss models.

DEMO 6.5. (Estimation in the Marshall–Olkin Model.) The estimator for the dependence parameter in the standard Marshall–Olkin model is available in the multivariate MAX domain. The estimator will be applied to an artificial data set. (a) Execute *Data... Generate Bivariate Data... Marshall–Olkin* for $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1$ and $\lambda = 0.4$.

(b) Execute *Estimate... Moment(Marshall–Olkin)*; the estimated dependence parameter is displayed. The estimator can also be applied to a multivariate data set with more than two columns yielding a matrix of estimates of the pairwise dependence parameters.

For real data, first estimate the marginal distribution and transform the data to a standard Marshall–Olkin form (cf. Example 12.3.1 on page 306 in Statistical Analysis and Demo 8.3).

DEMO 6.6. (Estimation in the Hüsler–Reiss Model.) The MLE is applied to the ozone data presented in Example 12.3.2 on page 308 in Statistical Analysis. (a) Read the data set cm-ozon1.dat.

(b) Activate the multivariate MAX domain and execute *Estimate... MLE(Huessler–Reiss)*. The estimated parameters (by means of MLE(EV)) of the marginal EV distributions and a matrix of pairwise dependence parameters (estimated by means of the MLE for standard Hüsler–Reiss distributions) is displayed. The transformation of the data to the standard form is done automatically.

Likewise, the moment estimator in the Gumbel–McFadden model can be applied to a data set.

In the multivariate mode of Xtremes select the POT domain. The experienced user may carry out the usual procedures such as

- generating data under parametric distributions;
- visualizing parametric distributions (here, densities, dfs, survivor functions in the form of contour and 3D plots);
- estimating parametric distributions (in the original as well as the canonical parameterization);
- estimating the nonparametric canonical dependence function by means of the estimator in (13.17) on page 319 in Statistical Analysis for different values k,
- employ the general bivariate plotting facilities.

As a first exercise one should reproduce the illustrations in Figures 13.3 and 13.4 on pages 320 and 321 in Statistical Analysis by using Xtremes.

Chapter 7

The HYD, ACT and Finance Menus

7.1 Hydrological Menu Options

The statistical tools for handling an inhomogeneous Poisson process (and, thus, partial duration series with a seasonal variation) are hidden in the hydrological supplement which becomes available by means of the HYDrology button.

A specific scatterplot (called hydrograph) is included. Our approach to the hydrograph is guided by the method of building clusters by runs.

7.1.1 The Hydrograph

A scatterplot with linearly interpolated points (i, y_i) of, e.g., daily measurements y_i of discharges or water levels at time i is a hydrograph. To reduce the huge number of data, one selects relevant upward and downward peaks including the measurements for floods and droughts.

DEMO 7.1. (Hydrograph of Moselle River for 1000 Days.) We assume that the required active data set is already available. Execute the option *Hydrograph* in the HYDrology facility. One must select the reduction coefficient c with $0 \leq c < 1$, the prior run length $r(0) \geq 1$ and the minimum distance $r(1) \geq r(0)$. First, take the default values $c = 1/4$ and $r(1) = r(0) = 1$.

For more details about our special construction of a hydrograph see the explanations in the online–help.

7.1.2 Calculating T–Year Levels Based on Maxima

The following calculations can be carried out within the usual framework. We give the technical details which are required for handling Example 14.1.1 on page 338

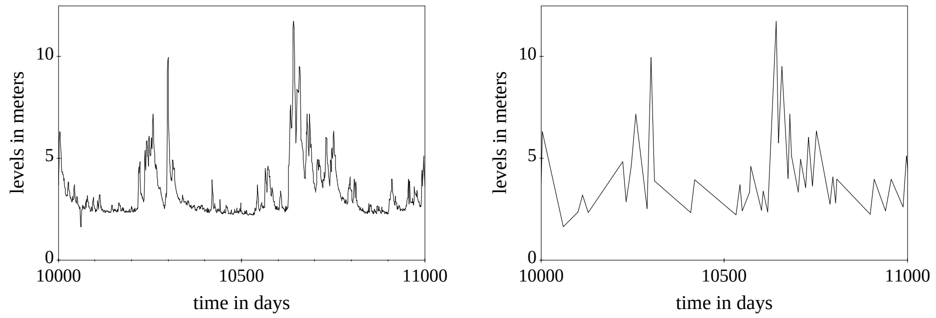


FIG. 7.1. (left.) Original scatterplot. (right.) Reduced scatterplot (hydrograph).

in Statistical Analysis.

DEMO 7.2. (Annual Flood Series of Moselle River.) (a) Read the Moselle levels from ht–mofil.dat. Apply *Data... Convert to... Univariate Data* to get data of the type Xtremes Univariate Data. Extract the annual maxima by executing *Data... Transform Data... Save Blocks Maxima* with *Block Size* equal to 365.

(b) Compare the kernel density estimator and the density of the MLE(EV) (MAX... *Estimating... MLE(EV)... Density*).

(c) After a right mouseclick into the MLE(EV) dialog box, execute the *T–Year Threshold* option.

Seasonal maxima can be dealt with in a similar manner. The T–year threshold must be calculated in the non–identically distributed case.

7.1.3 Partial Duration Series

Partial duration series must be analyzed in the HYDrology supplement. The following demo corresponds to Example 14.2.1 on page 341 in Statistical Analysis.

DEMO 7.3. (Partial Duration Series of Moselle River.) (a) Read again the Moselle levels from ht–mofil.dat. Save the cluster maxima (execute *Data... Transform Data... Save Cluster Maxima...*) with run length $r = 7$ and threshold $u = 5$ to a file.

(b) Briefly analyze the autocorrelation function of the cluster maxima series (*Visualize... Scatterplot* and *Autocorrelation* in the local menu).

(c) Select the *Analyze Partial Duration Series...* facility in the HYDrology supplement. Select *Threshold* for $u = 5$, *Number of neighbors* $k = 35$ and T–year thresholds varying between 10 and 15 with 100 supporting points. Pressing OK one obtains the following output:

1. a scatterplot of the cluster maxima (modulo 365 days),
2. a plot of the weekly frequencies (represented by bars),
3. T–year thresholds $u(T)$ plotted against the time span T, whereby the thresholds run between the bounds specified in the dialog box, and

4. an active data set (with default name pardur.dat) with the day number and the shape and scale parameters of the fitted GP distributions for each day of occurrence of a cluster maxima.

The T–year thresholds and the weekly frequencies can be tabulated by using the curve tabulating mouse mode (activate the  tool in the toolbar and click onto the curve or bars).

7.2 Actuarial Menu Options

Special options for insurance questions are available by means of the ACTuary button in the tool bar of Xtremes. A pop–up menu allows the user

- to analyze the segments in the nonparametric and parametric PML approach,
- to simulate a path of a Poisson or Pólya–Lundberg process, to visualize a path of a risk process, to simulate ruin times and to simulate the ruin probabilities (within a fixed time horizon) for different initial reserves.

The claim arrival times are modeled by a Poisson or Polyá–Lundberg process, while the claim size distribution may be selected from one of generalized Pareto (GP) models (in the α or γ–parameterization).

7.2.1 Analyzing PML Groups

In the following two demos we discuss in detail the steps required for carrying out the analysis of a PML data set. The primary aim is to fit GP models to the claim degrees and claim sizes within PML groups and to calculate the mean degrees and the mean claim sizes.

Demo 7.4. (Creating a PML Data Set.) Given a bivariate data set (of the type Xtremes Multivariate Data) with PMLs and claim sizes in the two columns, one may add boundaries of PML groups and the group number (apply *ACT... PML... Transform to Segments*). In the next step, one may add the claim degrees (apply *Add Degrees*). Further columns may be added to this data set such as, e.g., the number of risks belonging to the PML groups, yet the latter operation is not supported by Xtremes.

Subsequently, we assume that there is a data set with group numbers, lower and upper boundaries of the PML groups, PMLs, claim sizes and claim degrees in the first six columns. The data set im–pmlfi.dat (studied in Section 17.3 in Statistical Analysis) is of that type. The required steps for analyzing such a data set will be outlined in the following demo.

Demo 7.5. (Analyzing a PML Data Set.) (a) Read im–pmlfi.dat from the subdirectory `xtremes\dat`.

(b) Open the relevant dialog box by selecting *ACT... PML... Analyze Segments...*. Choose, e.g., the options *Hill(GP1/GP2)* and *Degrees*. Furthermore, specify the threshold $t = 100$. After having executed OK,

- a new active data set (called para.dat) is generated containing the PML group numbers i, the centers q_i, the parameters of the fitted beta distributions and the estimated mean degrees $(1 + u_i|\alpha_i|)/(1 + |\alpha_i|)$, and
- in addition, there are Xtremes Univariate Data sets seg*.dat containing the claim degrees belonging to the different PML groups.

(c) If the default option *No Plots* was changed to *Plot DF*, then a plot window opens for each PML group displaying the empirical df of the claim degrees within the PML group and the fitted beta (GP2) df (alternatively, one may choose *Plot QF*).

(d) One can also apply this procedure to the original claim sizes by choosing *Claim Sizes* in the dialog box. Then, the mean claim sizes are computed instead of the mean degrees. If the option *Hill(GP1/GP2)* was selected, then the Hill estimator in the Pareto (GP1) model is applied.

Correspondingly, use the dialog box *ACT... PML... Analyze Nearest Neighbors*. Afterwards, for smoothing mean degrees or estimated parameters produce scatterplots and fit polynomial (exp–polynomial) regression curves. For tabulating the required values, use a procedure as explained in Demo 8.2. Notice that this procedure is similar to the one that can be achieved by means of the curve tabulating mouse mode (activate the tool ).

7.2.2 Claim Number Process

The simulated values of the claim arrival process are stored as a univariate data set. One may visualize the path of the process by means of the *Path* option in the *Visualize* menu.

7.2.3 Risk Process

The path of a risk process is simulated within a finite time horizon. The arrival times T_i of the claims and the reserves $U(T_i)$ at that time are stored into a file. The *Visualize* option allows for the interactive generation of the path using the `Up` or `PageUp` keys.

DEMO 7.6. (Visualizing Paths of Risk Processes.) Xtremes will be applied to simulate a path of a risk process as displayed in Fig. 17.5 on page 429 in Statistical Analysis. Enter the pertaining dialog box by *ACT... Risk Process... GP Claim Sizes*. Select the intensity $\lambda = 1.7$, the safety loading $\rho = 0.1$, the safety exponent $\beta = 1$, the time horizon $T = 50$ and the initial reserve $s = 250$. In addition, select the location parameter $\mu = 0$ because we merely take into account the excesses over the threshold $u = 22$ in the reinsurance business. Execute the *Visualize* option and use the `Up` and `PageUp` to generate the arrival of claims.

Ruin Probabilities

Open the dialog box *ACT... Ruin Probabilities...* to simulate ruin probabilities for running initial reserves. For the details we refer to the online help (press the F1 key when the dialog box is open).

Ruin Time Data

The ruin times of independent paths of the risk process are stored in a file. One must fix a time horizon that is written into the data set if no ruin occurs within that time horizon. The pertaining dialog box for choosing the parameters is shown in Fig. 7.2.

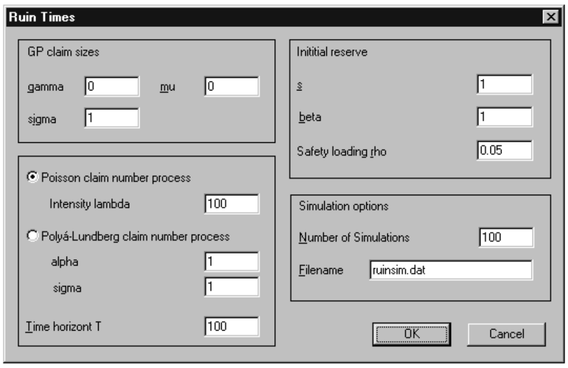


FIG. 7.2. Ruin times box: the options for claim size distribution, claim number process and initial reserve are common to all boxes in the actuarial menu.

Ruin time data may be used to calculate the T–year initial reserve. We demonstrate in which way the plots in Fig. 17.6 on page 431 in Statistical Analysis can be produced by applying Xtremes.

DEMO 7.7. (Generation of Ruin Times and Estimation of the T–Year Initial Reserve.) (a) Enter the dialog box for ruin times by executing *ACT... Ruin Times... GP Claim Sizes.* For the selected parameters simulate $N = 4000$ (Monte Carlo sample size) ruin times which are now available in the active data set. Execute the option *Visualize... Sample QF* to display the simulated ruin time qf H_s^{-1}, where s is the selected initial reserve. By repeating this procedure for several initial reserves, one gets an initial risk contour plot as displayed in Fig. 17.6.

(b) Click the StatPascal (SP) button in the tool bar and load the file `initres.sp` from the StatPascal subdirectory. Choose *Run* to estimate the initial reserve for the preceding setting. The choice of parameters and the output runs in the StatPascal window.

(c) Carry out the bootstrapping (for constructing an upper confidence bound) by the modified program stored in `bootres.sp`.

7.3 Analyzing Financial Data

Financial data can be analyzed to some extent by means of the usual visualization techniques and estimation procedures provided by Xtremes. We always assume that the prices are stored as Xtremes Multivariate Data. In the first three columns the date (in the order day, month and year) is listed, in the fourth column there is the speculative price of the asset for the given date. If there is a different ordering, employ the *Data... Transform...* facility which is available in the Data menu.

DEMO 7.8. (Plotting Log–Prices of Stock Market Index.) Load the file fm–poors.dat from `xtremes\dat`. Inspect the data (right mouseclick into the ACTIVE DATA window and select *List Data*). There is the ordering day/month/year for the date. Execute *Data... Transform... Date Transformation* and, afterwards, *Data... Convert to... Time Series* to convert the data set to a time series. A visualization of the data set is obtained by *Visualize... Scatterplot*. Carry out a log–transformation by *UFO... Transform Data*. Display the result as shown in Fig. 16.1 (left) on page 375 in Statistical Analysis.

A special transformation is needed to distribute Monday returns equally over the weekends or to omit the Monday returns (generally, returns after a gap). The *Price Return Transformation* option can be found in the finance supplement under the button $ in the menu bar.

DEMO 7.9. (Analyzing Stock Market Returns.) Activate fm–poors.dat. Open *$... Price Return Transformation* and execute *Distribute*. Inspect the data. In the fourth column you will find the equally distributed returns. Extract the returns by applying the option *Data... Convert to... Univariate Data*.

(a) Visualize the full data set.

(b) Analyze the upper extremes (the upper tail). Repeat the analysis in Example 16.3.1 on page 379 in Statistical Analysis.

(c) Analyze the lower extremes (the lower tail). First execute the option *Data... Transform Data... Converse Sign*.

(d) Analyze the symmetrized data. First execute the option *Data... Transform Data... Symmetrize*.

To compute the capital/loss coefficient, one can use the usual Xtremes facilities to estimate a quantile in a parametric manner.

DEMO 7.10. (Computing the Capital/Loss Coefficient.) We provide some details about the computations in Example 16.5.2 on page 391 in Statistical Analysis. Activate fm–poors.dat. Generate the equally distributed returns as in Demo 7.9. For $N = 1$ proceed as in (c). Apply the Moment(GP) estimator and calculate the q–quantile by using the option *T–Year Threshold* in the local menu of the estimator dialog box (right mouseclick onto the dialog box) with $1 - 1/T = q$.

For $N = 30$, apply *Data... Transform Data... Save Block Sums* for *BlockSize* equal to 30 to the daily returns. As a result you get the 30–days returns. Then proceed as in the case of daily returns.

Finally, we remark that Xtremes also includes a facility to simulate GARCH series (use *$... GARCH(p,q) Series*). It is intended to implement estimation procedures for such series in future releases.

Chapter 8

UserFormula Facilities

A first extension to the menu system is provided by the UserFormula (UFO) facility. With UserFormula, the user can type in formulas that are used

- to evaluate expressions using a calculator;
- to plot univariate or bivariate curves;
- to generate data sets;
- to transform existing data sets.

The formulas are entered by using the notation of common programming languages.

We give an overview of the functions that are available in UserFormula expressions and describe the options of the UserFormula menu, which opens after clicking the UFO button UFO in the toolbar. Operations that are too complicated for UserFormula may be handled by using the integrated programming language StatPascal.

8.1 Overview

One can access all distributions implemented in Xtremes by calls to predefined functions. There are three different groups of predefined functions.

- Standard mathematical functions like `abs(x)` (absolute value), `exp(x)` (exponential function), `log(x)` (natural logarithm) or `sqrt(x)` (square root), among others.
- Function calls—partly including a shape parameter `a`—under which one may generate data, such as `betadata(a)` or `gumbeldata`. The returned values are

independent for successive calls and governed by the respective distribution in its standard form. In addition, $[0, 1)$–uniform data may be called by the function `random`.

- Functions for densities, qfs and dfs (again partially including a shape parameter `a`) such as:
  - `betadensity(a,x)`, `betadf(a,x)`, `betaqf(a,x)`;
  - `gaussiandensity(x)`, `gaussiandf(x)`, `gaussianqf(x)`; etc.

The last curve plotted within an Xtremes plot window is available under the name `actualcurve`. An application of `actualcurve` is given in Demo 8.2.

The installation program copies a postscript version of the *StatPascal Reference Manual* to the file `spdoc.ps` in the SP subdirectory. The chapter *Library Functions* gives a detailed description of all predefined functions that are available within the UserFormula facility.

8.2 Calculator

The calculator allows the user to type in a formula and evaluate it. Fig. B.1 shows the calculator dialog box.

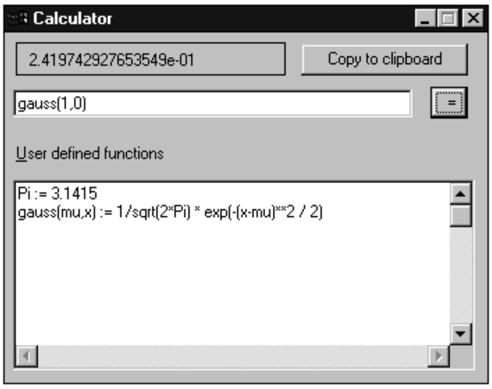


FIG. 8.1. *Calculator* dialog box. Formulas typed in the upper edit field are evaluated. The lower edit field defines variables and functions also available in other parts of Xtremes.

In the lower part of the calculator window, you can define your own functions and variables. Write your definitions in the edit field *User Defined Functions* and click on the =–button. The definitions thus made are available within all dialog boxes providing a UserFormula facility. They can also be used in all edit fields where a real value is expected, e.g., in the dialog boxes used for plots of parametric curves. For example, a Gaussian density including a location parameter is defined in the following way:

```
Pi:=3.1415
gauss(mu,x):=1/sqrt(2*Pi)*exp(-(x-mu)**2/2)
```

The formulas are stored in the file `formula.txt` within the working directory. They are loaded again upon the next start of Xtremes.

8.3 Plotting Curves

The graph of a function $x \to f(x)$ or $x \to f(\boldsymbol{p}, x)$ may be plotted in every graphics window. The optional parameter vector $\boldsymbol{p} = (p1, p2, p3)$ is changed by using the parameter varying mouse mode tool . Instead of $p1$ one may also use p. Within the multivariate mode, a surface plot of a function $(x, y) \to f(x, y)$ is performed.

DEMO 8.1. (Plotting Gaussian Densities with Varying Location Parameter.) Click the UFO button  in the toolbar and select the option *Plot curve* from the popup menu. Now, type the formula `1/sqrt(2*3.1415) * exp(-(x-p1)**2/2)` in the edit field labeled *f(x) or f(p,x)* in the dialog box. If you have entered the definitions shown in the *Calculator* box, you can also write `gauss(p1, x)`.

Especially note the option for the destination window. Xtremes lists all open windows, and you can also enter the name of a new window. Select OK to plot the curve.

8.4 Generating Data

The UserFormula facility may be employed to generate univariate data sets. Click the UFO button and select the option *Generate Data...* A dialog box similar to the one used for plotting curves is utilized.

Now, the user must specify a quantile function (qf) `Q(x)` that is applied to $[0, 1)$–uniform data. For example, use `-log(x)` to generate standard exponential data.

Data distributed according to the distributions implemented in Xtremes is available by means of the predefined functions `*data` (where `*` is replaced by the name of the distribution). For example, one might also write `exponentialdata` in the above example.

8.5 Transforming Data

The UserFormula facility offers the transformation of univariate or multivariate data sets and time series. When you select the option *Transform Data...* in the UFO menu, Xtremes asks for a transformation depending on the type of the active data set.

- Univariate Data x_i: specify a transformation T to generate the data $T(x_i)$.

- Time Series (t_i, x_i): specify two functions $T_1(t, x)$ and $T_2(t, x)$ to obtain the time series values $(T_1(t_i, x_i), T_2(t_i, x_i))$. Note that real–valued times are allowed.
- Multivariate Data $(x_{i,1}, \ldots, x_{i,m})$: specify transformations T_j. The system generates
$$(T_1(x_{i,1}, \ldots, x_{i,m}), \ldots, T_k(x_{i,1}, \ldots, x_{i,m})).$$
In addition to the transformation, one must specify k names for the columns of the transformed data set. See Demo 8.3.

DEMO 8.2. (Smoothing a Data Set Using Polynomial Regression.) Convert the data set to a time series, display a scatterplot and add a regression polynomial. Now, the polynomial is available as `actualcurve`. Therefore, one can apply the transformation $T_1(t, x) = t$ and $T_2(t, x) = \text{actualcurve}(t)$ to store the values of the polynomial, evaluated at the times t of the original time series.

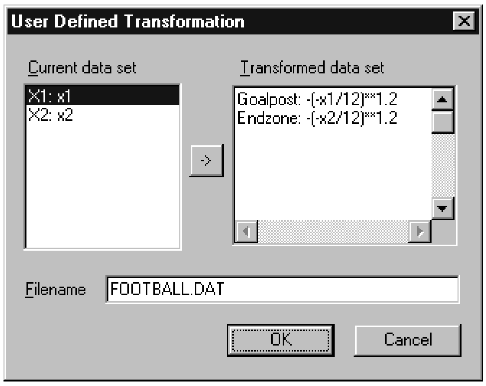


FIG. 8.2. *Transform Data* dialog box for multivariate data. The transformation in (12.39) on page 307 in Statistical Analysis is applied to football.dat (with changed signs).

DEMO 8.3. (UFO Transformation of Multivariate Data.) Read football.dat (cf. Example 12.3.1 on page 306 in Statistical Analysis) and change the signs (*Data... Transform Data... Change Signs*).

Choose UFO and apply *Transform Data*. The dialog box *Transform Data* lists the column names `X1` and `X2` of the current data set on the left–hand side (see Fig. B.2) together with the variable names `x1` and `x2` assigned to the values in the columns. In the edit field on the right–hand side, the user must define the names of the ith column (using the arrow button one may edit a template of the transformation $T(x_1, x_2) = (x_1, x_2)$). In our example, we use the names `Goalpost` and `Endzone` and add the transformed variables `-(-x1/12)**1.2` and `-(-x1/12)**1.2`. Finally, press OK to execute the transformation.