Preview

The Preview presents a tour along the main results of these lecture notes. It introduces concepts and notation that will be used throughout the book. It should help the reader to follow the thrust of the ideas developed in the individual lectures, and to determine which lectures are of sufficient interest to merit a closer look.

A recipe

The question addressed in these lectures is simple: Given a multivariate sample cloud, what can one say about the underlying distribution in a region containing only one or two points of the sample?

Typically the region is a halfspace, and one is concerned about the eventuality of future data points lying far out in the region. We shall use the terminology of financial mathematics and speak of loss and risk. The data cloud could just as well contain data of insurance claims, or data from quality control, biomedical research, or meteorology. In all cases one is interested in the extremal behaviour at the edge of the sample cloud, and one may use the concepts of risk and loss. In a multivariate setting *risk* and loss may be formalized as functions which increase as one moves further out into the halfspace.

In first instance the answer to the question above is: “Nothing”. There are too few points to perform a statistical analysis. However some reflection suggests that one could use the whole sample to fit a distribution, say a Gaussian density, and use the tails of this density to determine the conditional distribution on the given halfspace. In financial mathematics nowadays one is very much aware of the dangers of this approach. The Gaussian distribution gives a good fit for the daily log returns, but not in the tails. So the proper recipe should be: Fit a distribution to the data, and check that the tails fit too. If one can find a distribution, Gaussian say, or elliptic Student, that satisfies these criteria, then this solves the problem, and we are done. In that case there is no need to read further.

What happens if the data cloud looks as if it may derive from a normal distribution, but has heavy tails? There is a convex central black region surrounded by a halo of isolated points. The cloud does not exhibit any striking directional irregularities. Such data sets have been termed *bland* by John Tukey. Only statistical analysis is able to elicit information from bland clouds.

Rather than fitting a distribution to the whole cloud, we shall concentrate on the tails. We assume some regularity at infinity. In finite points regularity is expressed by the existence and continuity of a positive density at those points. Locally the distribution will then look like the *uniform distribution*; under proper scaling the sample cloud will converge vaguely to the standard Poisson point process on $\mathbb{R}^d$ as the number of points in the sample increases. We want to perform a similar analysis at infinity. Of course, in a multivariate setting there are many ways in which halfspaces may diverge. This problem is inherent to multivariate extremes. In order to obtain useful results, we have to introduce some regularity in the model setup.

Ansatz. Conditional distributions on halfspaces with relatively large overlap asymptotically have the same shape.

Let us make the content of the Ansatz more precise.

Definition. Two probability distributions (or random vectors Z and W) have the same *shape* or are of the same *type* if they are *non-degenerate*, and if there exists an affine transformation α such that Z is distributed like $\alpha(W)$. A random vector Z has a *degenerate* distribution if it lives on a hyperplane, equivalently, if there exists a linear functional $\xi \neq 0$ and a real constant c such that $\xi Z = c$ a.s.

For instance, all Gaussian densities have the same shape. Shape (or type) is a geometric concept. Given a non-degenerate Gaussian distribution on $\mathbb{R}^d$ one can find coordinates such that in these coordinates the distribution is standard Gaussian with density

$$e^{-(w_1^2+\cdots+w_d^2)/2}/(2\pi)^{d/2}, \quad w = (w_1, \dots, w_d) \in \mathbb{R}^d.$$

A basic theorem in this setting is the Convergence of Types Theorem (CTT). It allows us to speak of a sequence of vectors as being asymptotically Gaussian. We write $Z_n \Rightarrow Z$ if the distribution functions (dfs) of Z_n converge weakly to the distribution function (df) of Z.

Theorem 1 (Convergence of Types). *If $Z_n \Rightarrow Z$ and $W_n \Rightarrow W$, where W_n and Z_n are of the same type for each n, then the limit vectors, if non-degenerate, are of the same type.*

Proof. See Fisz [1954] or Billingsley [1966]. □

At first sight the CTT may look rather innocuous. In many applied probability questions involving limit theorems it works like a magic hat from which new models may be pulled: In the univariate setting the Central Limit Problem for partial sums yields the stable distributions; the Extreme Value Problem for partial maxima yields the extreme value distributions. See Embrechts, Klüppelberg & Mikosch [1997], Chapters 2 and 3. In the multivariate setting, in Chapter II below, the CTT yields the well-known multivariate max-stable laws; in Chapter III the CTT yields a continuous one-parameter family of limit laws; and in Chapter IV the CTT yields two semi-parametric families of high risk limit laws, one for exceedances over (horizontal) linear thresholds, one for exceedances over elliptic thresholds.

Let us try to give the intuition behind the CTT in the case of a Gaussian limit.

A sequence of random vectors Z_n is asymptotically normal if there exist affine normalizations α_n such that

$$W_n := \alpha_n^{-1}(Z_n) \Rightarrow W,$$

where W is standard normal. The validity of the term asymptotic normality would seem to derive from geometric insight. In geometric terms one may try to associate with each Z_n an ellipsoid E_n that is transformed into the unit ball B by the normalization. These ellipsoids $E_n = \alpha_n(B)$ may be related to the expectation and covariance of Z_n (if these exist and converge), or to certain convex *level sets* of the density of Z_n (if the density exists and is *unimodal*). Perhaps the correct intuition is that large sample clouds from distributions that are asymptotically Gaussian are asymptotically elliptic, and that affine transformations that map the elliptic sample clouds into spherical sample clouds may be used to normalize the distributions. The normalizations are thus determined geometrically. The same geometric intuition forms the background to these lectures. Instead of convergence of the whole sample cloud we now assume convergence at the edge. Since we want to keep sight of individual sample points, we assume convergence to a point process.

Affine transformations are needed to pull back the distributions as the halfspaces diverge. Let us say a few words about the space $\mathcal{A} = \mathcal{A}(d)$ of affine transformations on $\mathbb{R}^d$. Recall that an *affine transformation* has the form

$$w \mapsto z = \alpha(w) = Aw + a, \tag{1}$$

where a is a vector in $\mathbb{R}^d$, and A an invertible matrix of size d. The inverse is

$$z \mapsto w = \alpha^{-1}(z) = A^{-1}(z - a).$$

The set $\mathcal{A}$ is a group since $\alpha^{-1} \in \mathcal{A}$, and the composition of two affine transformations is an affine transformation:

$$w \mapsto (\alpha\beta)(w) = \alpha(\beta(w)) = A(Bw + b) + a.$$

Convergence $\alpha_n \to \alpha$ means $a_n \to a$ and $A_n \to A$ componentwise, or, equivalently, $\alpha_n(w) \to \alpha(w)$ for all $w \in \mathbb{R}^d$, or for $w = 0, e_1, \dots, e_d$, where $e_1, \dots, e_d$ are linearly independent vectors. From linear algebra it is known that $A \mapsto A^{-1}$ is continuous on the group $\mathrm{GL}(d)$ of invertible matrices of size d. Hence $\alpha \mapsto \alpha^{-1}$ is continuous. Obviously $(\alpha, \beta) \mapsto \alpha\beta$ is continuous. So $\mathcal{A}$ is a topological group. It is even a Lie group. It is possible to see $\mathcal{A}(d)$ as a subgroup of $\mathrm{GL}(1 + d)$ by writing

$$\alpha(w) = Aw + a = z \iff \begin{pmatrix} 1 & 0 \\ a & A \end{pmatrix} \begin{pmatrix} 1 \\ w \end{pmatrix} = \begin{pmatrix} 1 \\ a + Aw \end{pmatrix} = \begin{pmatrix} 1 \\ z \end{pmatrix}. \tag{2}$$

This representation makes it possible to apply standard results from linear algebra when working with affine transformations. We usually write Z for the random vector with components Z_i and

$$W = \alpha^{-1}(Z) = A^{-1}(Z - a) \tag{3}$$

for the normalized vector. Now assume $\alpha_n^{-1}(Z_n) \Rightarrow W$ where W is non-degenerate. The normalizations α_n are not unique. They may be replaced by normalizations β_n which are *asymptotic* to α_n:

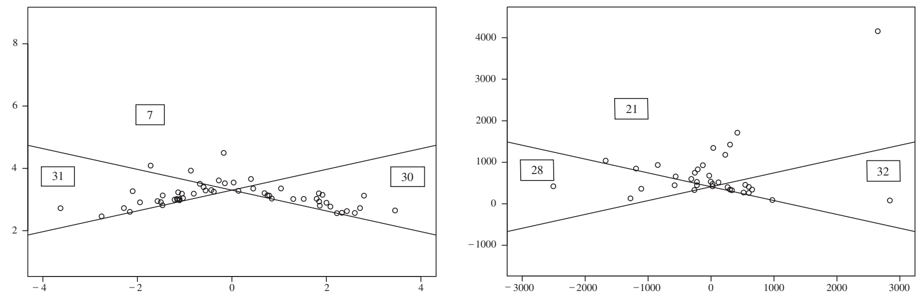
$$\beta_n \sim \alpha_n \iff \alpha_n^{-1}\beta_n \to \mathrm{id}, \tag{4}$$

where id stands for the identity transformation. Asymptotic equality is an equivalence relation for sequences in $\mathcal{A}$.

Warning. If $\alpha_n \to \mathrm{id}$ then $\alpha_n^{-1} \to \mathrm{id}$. However $\alpha_n \sim \beta_n$ does not imply $\alpha_n^{-1} \sim \beta_n^{-1}$, not even in dimension $d = 1$. Here is a simple counterexample:

Example 2. Let X_n be uniformly distributed on the interval $(1, n+1)$. Properly normalized, the X_n converge in distribution to a rv U which is uniformly distributed on the interval $(0, 1)$. Indeed $(X_n - 1)/n \Rightarrow U$; but also $X_n/n \Rightarrow U$. Set $\alpha_n(u) = nu + 1$ and $\beta_n(u) = nu$. Then $\beta_n^{-1}\alpha_n(u) \to u$ but $\alpha_n\beta_n^{-1}(x) = x + 1$. So $\alpha_n \sim \beta_n$ does not imply $\alpha_n^{-1} \sim \beta_n^{-1}$. Indeed, asymptotic equality means that the normalized variables $\alpha_n^{-1}(X_n)$ and $\beta_n^{-1}(X_n)$ are close, not the approximations $\alpha_n(U)$ and $\beta_n(U)$. ◊

After this digression on shape, geometry and affine transformations, let us return now to the basic question of determining the distribution on a halfspace containing only a few (or no) points of the sample, and to our Ansatz that high risk scenarios



Standard normal — Spherical Cauchy

Exceedances over linear thresholds with varying direction, for 40 000 points. Boxed are the number of sample points in the halfplanes, and in the intersection. In the Cauchy sample many points lie outside the figure; the highest is (27 000, 125 000).

on halfspaces with relatively large overlap have distributions with approximately the same shape. Here a *high risk scenario* of a random vector Z for a given halfspace H is just the vector Z conditioned to lie in H. For halfspaces far out this corresponds to our interpretation of a rare or extreme event. The reader may wonder whether the Ansatz implies that all high risk scenarios asymptotically have the same shape. Note that the condition of a relatively large overlap is different for light tails and for heavy tails. For a Gaussian distribution the directions of two halfspaces far out have to be close to have a relatively large overlap; for a spherical *Cauchy distribution* there is considerable overlap even if the directions of the halfspaces are orthogonal. See the figure above.

In the univariate case the condition that high risk scenarios, properly normalized, converge leads to a one-parameter family of limit shapes, the *generalized Pareto distributions*. These *GPD*s may be standardized to form a continuous one-parameter family, indexed by $\tau \in \mathbb{R}$:

$$G_\tau(v) = 1 - (1 + \tau v)_+^{-1/\tau}, \quad v \geq 0. \tag{5}$$

By continuity G_0 is the standard exponential df. The associated univariate limit theory has been applied in many fields. It is our aim to develop a corresponding theory in the multivariate setting.

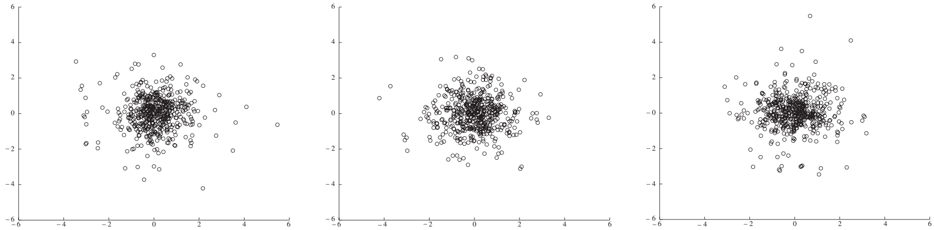
We shall denote the *high risk scenario* for Z associated with the halfspace H by Z^H. By definition, Z^H lives on H, and for any Borel set E

$$\mathbb{P}\{Z^H \in E\} = \mathbb{P}\{Z \in E \cap H\}/\mathbb{P}\{Z \in H\}.$$

Halfspaces are assumed to be closed, and $\mathbb{P}\{Z \in H\}$ is assumed positive.

One may impose the condition that the high risk scenarios Z^{H_n}, properly normalized, converge for any sequence of halfspaces H_n under the sole restriction that $\mathbb{P}\{Z \in H_n\}$ is positive and vanishes for $n \to \infty$. This assumption is quite strong. It presupposes a high degree of directional homogeneity in the halo of the sample cloud. In order to understand multivariate tail behaviour, a thorough analysis of the consequences of this strong assumption seems like a good starting point. This analysis is given in Chapter III, the heart of the book. As an illustration we exhibit below the three plane projections of a data cloud in $\mathbb{R}^3$, consisting of the log-returns of three stocks on the Dutch stock exchange AEX over the period from 2-2-04 until 31-12-05. The data were kindly made available by Newtrade Research.

One may also start with the weaker assumption that the high risk scenarios converge for halfspaces which diverge in a certain direction. This is done in Chapter IV for horizontal halfspaces. The Ansatz now holds only for horizontal halfspaces. Write $z = (x, y)$ where y is the vertical component of z and x the h-dimensional horizontal part, with $h = d - 1$. Similarly write $Z = (X, Y) \in \mathbb{R}^{h+1}$. High risk scenarios for *horizontal halfspaces* $H^y = \mathbb{R}^h \times [y, \infty)$ correspond to *exceedances* over *horizontal*



Bland sample clouds: bivariate marginals of daily log-returns ING - Allianz - ASML.

thresholds. Let $\alpha(y)$ be affine transformations mapping $H_+ = \{y \geq 0\}$, the *upper halfspace*, onto H^y. The vectors $W_y = \alpha(y)^{-1}(Z^{H^y})$ live on H_+. Now suppose that the $\alpha(y)$ yield a limit vector:

$$W_y := \alpha(y)^{-1}(Z^{H^y}) \Rightarrow W, \quad \mathbb{P}\{Y \geq y\} \to 0. \tag{6}$$

Assume the limit is non-degenerate. What can one say about its distribution? It is not difficult to see that $\alpha(y)$ maps horizontal halfspaces into horizontal halfspaces, and that this implies that the high risk scenarios $Y^{[y,\infty)}$ of the vertical coordinate, with the corresponding normalization, converge to the vertical coordinate V of the limit vector, $W = (U, V)$. By the univariate theory the vertical coordinate of the limit vector has a GPD, see (5).

One may prove more. Suppose (6) holds. Let $Z_1, Z_2, \ldots$ be independent observations from the distribution π of Z. Choose y_n so that $\mathbb{P}\{Y \geq y_n\} \sim 1/n$. Set $\alpha_n = \alpha(y_n)$ where $\alpha(y)^{-1}(Z^{H^y}) \Rightarrow W$ as above. Then the normalized sample clouds converge in distribution to a Poisson point process:

$$N_n := \{\alpha_n^{-1}(Z_1), \ldots, \alpha_n^{-1}(Z_n)\} \Rightarrow N_0. \tag{7}$$

The mean measures of the sample clouds N_n converge weakly to the *mean measure* ρ of the limiting Poisson point process N_0 on all horizontal halfspaces J on which ρ is finite:

$$\rho_n = n\alpha_n^{-1}(\pi) \to \rho \text{ weakly on } J, \quad \rho(J) < \infty. \tag{8}$$

The restriction of ρ to H_+ is a probability measure, the distribution of W.

The equivalence of the two limit relations (6) and (7) is the Extension Theorem in Section 14.6. It is a central result. The first limit relation is analytical. It raises questions such as:

1) What limit laws are possible?

2) For a given limit law, what conditions on the distribution of Z will yield convergence?

The second relation is more geometric. Here one may ask:

1) Does convergence in (8) also hold for halfspaces J which are close to horizontal?

2) Will the convex hull of the normalized sample cloud converge to the convex hull of N_0?

For the novice to the application of point process methodology to extreme value problems this all may seem to go a bit too fast. Modern *extreme value theory* with its applications to more involved problems in risk management, however, needs this level of abstraction. See McNeil, Frey & Embrechts [2005] for a good discussion of these issues. In the one-dimensional case one already needs such a theory for understanding the *Peaks Over Thresholds* method based on (5), or the limit behaviour of several order statistics, as will be seen in Section 6.4. So bear with us and try to follow the general scheme.

The vector W in (6) is the limit of high risk scenarios for horizontal halfspaces. It follows that the high risk scenarios W^{H^y} all have the same shape. This result follows from the trivial fact that a high risk scenario of a high risk scenario is again a high risk scenario, at least for horizontal halfspaces. In the univariate setting, the exponential distribution, the uniform distribution and the Pareto distributions all have the *tail property*: Any tail of the distribution is of the same type as the whole distribution. In fact this tail property characterizes the class of GPDs. It may also explain why these distributions play such an important role in applications in insurance and risk theory. In the multivariate setting the tail property is best formulated in terms of the infinite measure ρ in (8): There exists a *one-parameter group*[4] of affine transformations γ^t, $t \in \mathbb{R}$, such that

$$\gamma^t(\rho) = e^t \rho, \quad t \in \mathbb{R}. \tag{9}$$

These equations form the basis of the theory developed in the lectures below. The equations are simple. They succinctly express the stability inherent in the limit law in terms of symmetries of the associated infinite measure. The stability allows us to tackle our basic problem of describing the distribution tail on a halfspace that contains few sample points.

Definition. A measure on an open set in $\mathbb{R}^d$ is a *Radon measure* if it is finite for compact subsets. An *excess measure* is a Radon measure ρ on an open set in $\mathbb{R}^d$ that satisfies (9) and gives mass $\rho(J_0) = 1$ to some halfspace J_0.

Excess measures play a central role in this book. They are infinite, but have a simple probabilistic interpretation in terms of point processes. The significance of point processes for extreme value theory has been clear since the appearance of

[4]One-parameter groups of matrices should not frighten a reader who has had some experience with finite state Markov chains in continuous time or with linear differential equations in $\mathbb{R}^d$ of the form $\dot{x} = Ax$.

the book Resnick [1987]. In our more geometric theory the excess measure is the *mean measure* of the Poisson point process which describes the behaviour of the sample cloud, properly normalized, at its edge, as the number of data points tends to infinity. An example should make clear how excess measures may be used to tackle our problem of too few sample points.

Example 3. Suppose γ^t, $t \in \mathbb{R}$, is the group of vertical translations, $\gamma^t : (u, v) \mapsto (u, v+t)$ on $\mathbb{R}^{h+1}$. A measure ρ of the form $d\rho(u, v) = d\rho^*(u)e^{-v}dv$ will satisfy (9) for any probability measure ρ^* on $\mathbb{R}^h$. The halfspace $J_0 = H_+$ has mass one and ρ is an excess measure. Conversely one may show that any excess measure with the symmetries γ^t above has the form $d\rho(u, v) = d\rho^*(u)e^{-v}dv$ if one imposes the condition that $\rho(H_+) = 1$. ◊

The probability measure ρ^* in the example above is called the *spectral measure*. The product form of the excess measure in the example makes it possible to estimate the spectral measure even if the upper halfspace contains few points. One simply chooses a larger horizontal halfspace, containing more points. Something similar may be done for any excess measure for exceedances over horizontal thresholds. We shall not go into details here. Suffice it to say that such an excess measure is determined by its symmetry group γ^t, $t \in \mathbb{R}$, and a probability measure ρ^* on the horizontal coordinate plane $\mathbb{R}^h$. The spectral measure ρ^* may be interpreted as the conditional distribution of U given $V = 0$ for the limit vector $W = (U, V)$ on H_+ in (6). The exponential distribution on the vertical axis enters the picture via the Representation Theorem for the limit vector:

$$W = \gamma^T(U^*, 0), \tag{10}$$

where the vector U^* in $\mathbb{R}^h$ has distribution ρ^*, and T is standard exponential, independent of U^*. This decomposition of W reflects the symmetry of the excess measure expressed in (9). It enables one to build probability distributions on halfspaces H far out, and then to estimate probabilities $\mathbb{P}\{Z^H \in E\}$ for $E ß H$, and expectations $\mathbb{E}\varphi(Z^H)$ for loss functions $\varphi\colon H \to [0, \infty)$. Here is the recipe. Assume $\alpha_H^{-1}(Z^H) \Rightarrow W$, where W lives on a halfspace J_0, and has a non-degenerate distribution that extends to an excess measure ρ, and where $H = \alpha_H(J_0)$ are halfspaces such that $\mathbb{P}\{Z \in H\} \to 0$.

Recipe. *Replace Z^H by $\alpha_H(W)$ and compute $\mathbb{P}\{\alpha_H(W) \in E\} = \rho(\alpha_H^{-1}(E))/\rho(J_0)$ and the integral $\mathbb{E}\varphi(\alpha_H(W)) = \int_{J_0} \varphi \circ \alpha_H d\rho/\rho(J_0)$ in terms of the excess measure. Given the symmetry group and the normalization α_H one only needs to know the spectral measure ρ^* to compute these quantities. The spectral measure may be estimated from data points lower down in the sample cloud.* ◊

The spectral measure is dispensable. It is the symmetries in (9) that do the job.

These allow us to replace a halfspace containing few observations by a halfspace containing many observations, and on which the distribution has the same shape[5].

Given the recipe, it is clear what one should do to develop the underlying theory: determine the *one-parameter* groups γ^t in $\mathcal{A}(d)$, and for each determine the excess measures (if any) and their domains of attraction. This is done in Section 18.8 for $d = 2$. For linear transformations the program has been executed by Meerschaert and Scheffler in their book Meerschaert & Scheffler [2001] on limit laws for sums of independent random vectors. Let us give a summary of the theory in MS.

We may restrict attention to one-parameter matrix groups by (2). Such one-parameter groups are simple to handle. The group γ^t, $t \in \mathbb{R}$, is determined by its generator C. One may write $\gamma^t = e^{tC}$, where the right hand side is defined by its power series. There is a one-to-one correspondence between matrices C of size d and one-parameter groups of linear transformations on $\mathbb{R}^d$. If one chooses coordinates such that C has *Jordan form*, one may write down the matrices for γ^t by hand for any dimension d. See Section 18.12 for details. Now we have to choose ρ. Let ρ be a Radon measure on an open set OßR^d that satisfies $\gamma^t(\rho) = e^t\rho$, $t \in \mathbb{R}$. For an excess measure there still has to be a halfspace J_0 of mass one.

Example 4. *Lebesgue measure* on $\mathbb{R}^d$ satisfies (9) with $\gamma^t = \mathrm{diag}(a_1^t, \dots, a_d^t)$ for any diagonal matrix with $a_i > 0$, and $a_1 \dots a_d = 1/e$. However there are no halfspaces of measure one. There are, if one restricts the measure to an orthant, or a *paraboloid.* ◊

If γ^t for $t > 0$ maps the horizontal halfspace J_0 onto a proper subset of itself, then any probability measure ρ^* on $\mathbb{R}^h$ may act as spectral measure for a measure ρ that satisfies (9) and gives mass one to J_0. Similarly, if the γ^t for $t > 0$ are linear expansions, and the image of the open unit ball $B = \{\|w\| < 1\}$ contains the closed unit ball, a probability measure ρ^* on the sphere $\partial B = \{\|w\| = 1\}$ will generate a measure ρ on $\mathbb{R}^d \setminus \{0\}$ that satisfies (9), and gives mass one to the complement of the ball B. In the second case there are many halfspaces of finite mass: $\rho(J)$ is finite for any halfspace J that does not contain the origin. Constructing excess measures is not difficult!

Given the excess measure ρ, the halfspace J_0 of mass one, and the one-parameter group γ^t in (9), we still have to determine the domain of attraction. Recall that Z lies in the *domain of attraction* of W if (6) holds. We write $Z \in \mathcal{D}^h(W)$, and call $\mathcal{D}^h(W)$ or $\mathcal{D}^h(\rho)$ the *domain* of W or ρ *for exceedances over horizontal thresholds.* Let $Z \in \mathcal{D}^h(\rho)$ have distribution π. The basic limit relation (8) for horizontal

[5] Coles and Tawn, in their response to the discussion of their paper Coles & Tawn [1994] write: “Anderson points out that our point process model is simply a mechanism for relating the probabilistic structure within the range of the observed data to regions of greater extremity. This, of course, is true, and is a principle which, in one guise or another, forms the foundation of all extreme value theory.”

halfspaces J may be reformulated as

$$e^t \beta(t)^{-1}(\pi) \to \rho \quad \text{weakly on } J,\ \rho(J) < \infty. \tag{11}$$

The $\beta(t)$ belong to the group $\mathcal{A}^h$ of affine transformations mapping horizontal halfspaces into horizontal halfspaces. One may choose $\beta\colon [0, \infty) \to \mathcal{A}^h$ to be continuous, and to *vary like* γ^t:

$$\beta(t_n)^{-1}\beta(t_n + s_n) \to \gamma^s, \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}. \tag{12}$$

In a slightly different terminology this states that β varies regularly with index C, where C is the generator of the symmetry group γ^t. Section 18.1 contains a brief introduction to multivariate regular variation. Regular variation of linear transformations is treated in more detail in Chapter 4 of Meerschaert & Scheffler [2001]. The central result, the Meerschaert Spectral Decomposition Theorem, states that one may restrict attention to one-parameter groups γ^t for which all complex eigenvalues of γ have the same absolute value. See Section 18.4 for details.

Does one really need the theory of multivariate regular variation to handle high risk scenarios?

There are good reasons for using *regular variation* to study multivariate extremes. We list four:

1) The theory is basic. Nothing essential is lost if one assumes $t_n = n$ and $s_n = 1$ in (12). In the final resort, regular variation is about sequences of the form:

$$\beta(n) = \beta(0)\gamma_1 \dots \gamma_n, \quad \gamma_n \to \gamma. \tag{13}$$

One gets back the original curve β, or a curve asymptotic to the original curve, by interpolation. Details are given in Section 18.2.

2) The theory contains a number of deep results that clarify important issues in applications. We give two examples of questions that may be resolved by the Meerschaert *Spectral Decomposition Theorem*, the fundamental result in the multivariate theory of regular variation.

i) Suppose γ^t, $t \in \mathbb{R}$, is a group of *linear* transformations. Is it possible to choose the origin in z-space and normalizations $\bar{\beta}(t) \sim \beta(t)$, mapping w into z, that are linear in the new coordinates?

ii) Suppose the symmetries γ^t, $t \in \mathbb{R}$, in appropriate coordinates in w-space, are diagonal. Is it possible to choose coordinates in z-space and normalizations $\bar{\beta}(t) \in \mathrm{GL}$, asymptotic to $\beta(t)$ for $t \to \infty$, that are diagonal in the new coordinates?

The answer to i) is "Yes" if γ is an expansion, or a contraction ; see Lemmas 15.15 and 16.13 below. This result explains why univariate extreme value theory is so much simpler for heavy tails than for distributions in the domain of the Gumbel law. Univariate linear normalizations are non-zero scalars! The answer to ii) is "Yes" if the diagonal entries in γ are distinct.

3) Regular variation enables us to construct simple continuous densities in the domain of attraction of excess measures with continuous densities. By the transformation theorem the density g of the excess measure in (9) satisfies

$$g(\gamma^t(w)) = g(w)/q^t, \quad q = e|\det A|, \quad \gamma^t(w) = b(t) + A^t w. \tag{14}$$

For $\gamma^t \in \mathcal{A}^h$ one also has the decomposition

$$g(u,v) = g_v(u)\tilde{g}(v), \tag{15}$$

where $\tilde{g}$ is the density of a univariate GPD, see (5), and the conditional densities g_v all have the shape of the density g^* of the spectral measure ρ^*.

Example 5. The *Gauss-exponential* density $e^{-u^T u/2}e^{-v}/(2\pi)^{h/2}$ determines an excess measure ρ on $\mathbb{R}^{h+1}$. Vertical translations $\gamma^t : (u,v) \mapsto (u, v+t)$ are symmetries. The spectral density is standard Gaussian. For any curve $\beta : [0,\infty) \to \mathcal{A}^h$ that varies like γ^t there exists a vector $Z = (X,Y)$ with distribution π, and continuous density

$$f(x,y) = f_y(x)\tilde{f}(y) \tag{16}$$

such that $e^t \beta(t)^{-1}(\pi) \to \rho$ weakly on all horizontal halfspaces. The density f satisfies

$$\frac{f(\beta(t_n)(w_n))}{f(\beta(t_n)(0))} \to e^{-u^T u/2}e^{-v}, \quad t_n \to \infty, \ w_n \to (u,v) \in \mathbb{R}^{h+1}. \tag{17}$$

The density $\tilde{f}$ of Y satisfies the von Mises condition for the univariate Gumbel domain, see Section 6.6; the conditional densities f_y of X given $Y = y$ in (16) are Gaussian. ◊

A continuous density f as above will be called a *typical density* for g^* and β.

4) Multivariate regular variation has a strong geometric component. This is particularly clear if the excess measure is symmetric in a geometric sense. Let $\rho = \rho_\tau$ be a *Euclidean Pareto* measure on $\mathbb{R}^d \setminus \{0\}$. These measures are spherically symmetric with densities $c/\|w\|^{d+\lambda}$, where $\lambda = 1/\tau > 0$ describes the decay rate of the tails, and $c = c_\tau > 0$ may be chosen so that $\rho(J_0) = 1$ for the halfspace $J_0 = \mathbb{R}^h \times [1,\infty)$.

Definition. Bounded convex sets F_n and E_n of positive volume are *asymptotic* if

$$F_n \sim E_n \iff |F_n \cap E_n| \sim |F_n \cup E_n|. \tag{18}$$

Here $|A|$ denotes the volume of the set A.

Exercise 6. The reader is invited to investigate what a sequence of centered ellipses E_n in the plane looks like if $E_{n+1} \sim E_n$, and if the area $|E_n|$ is constant. ◊

Example 7. Start with a sequence of open centered ellipsoids $E_0, E_1, \dots$ such that $E_{n+1} \sim 2E_n$. Also assume E_n contains the closure of E_{n-1} for $n \geq 1$. For $c > 1$ one may use interpolation to construct a *unimodal* function f with elliptic *level sets* such that $\{f > 1/c^n\} = E_n$ for $n \geq 1$, and such that $\{f = 1\}$ is the closure of E_0. For $c > 2^d$ the function f is integrable, say $\int f d\lambda = C < \infty$. Suppose the ellipsoids $E(t) = \{f > 1/c^t\}$ *vary regularly*:

$$E(t_n + s_n) \sim 2^s E(t_n), \quad t_n \to \infty,\ s_n \to s,\ s \in \mathbb{R}. \tag{19}$$

The probability distribution π with density f/C lies in the domain of the *Euclidean Pareto* excess measure ρ_τ, where $c = 2^{d+1/\tau}$. One may choose linear transformations $\beta(t)$, depending continuously on $t \geq 0$, such that $\beta(t)(B) = E(t)$, and such that β varies like $\gamma^t : w \mapsto 2^t w$. Then

$$c^{-t}\beta(t)(\pi) \to \rho_\tau \text{ weakly on } \mathbb{R}^d \setminus \varepsilon B, \quad t \to \infty,\ \varepsilon > 0.$$

Details are given in Section 16. ◊

The reader will notice that densities occupy a central position in our discussions. In the multivariate situation densities are simple to handle. Densities are geometric: sample clouds tend to evoke densities rather than distribution functions. If the underlying distribution has a singular part, this will be reflected in irregularities in the sample cloud. Such irregularities, if they persist towards the boundary, call for a different statistical analysis. In the theory of coordinatewise maxima dfs play an all-important role. Densities have been considered too, see de Haan & Omey [1984] or de Haan & Resnick [1987], but on the whole they have been treated as stepchildren. In our more geometric approach densities are a basic ingredient for understanding asymptotic behaviour. From our point of view the general element in the domain of attraction of an excess measure with a continuous density is a perturbation of a probability distribution with a typical density. From a naive point of view we just zoom in on the part of the sample cloud where the vertical coordinate is maximal, adapting our focus as the number of sample points increases. Under this changing focus the density with which we drape the sample cloud should converge to the density of the limiting excess measure.

Proper normalization is essential for handling asymptotic behaviour and limit laws in probability theory. The geometric approach allows us to ignore numerical details, and concentrate on the main issues. Let us recapitulate: In order to estimate the distribution on a halfspace containing few sample points one needs some form of stability. The stability is formulated in our *Ansatz*: High risk scenarios far out in a given direction have the same shape. If one assumes a limit law, then there is an excess measure. The symmetries of the excess measure make it possible to estimate the distribution on halfspaces far out by our recipe above. The symmetries also impose conditions on the normalizations. These conditions have a simple formulation in terms of regular variation. One may choose the normalizing curve β in (11) to vary like γ^t. Roughly speaking, the group of symmetries γ^t of the excess measure enforces regular variation on the normalizations.

The four arguments above should convince the reader that regular variation is not only a powerful, but also a natural tool for investigating the asymptotic behaviour of distributions in the domain of attraction of excess measures.

In these notes we take an informal approach to regular variation, dictated by its applications to extremes. Attention is focussed on three situations:

1) for coordinatewise extremes the symmetries γ^t and the normalizations α_n are coordinatewise affine transformations (CATs);

2) for exceedances over *horizontal thresholds* the symmetries γ^t and normalizations α_n belong to the group $\mathcal{A}^h$: they map horizontal halfspaces into horizontal halfspaces;

3) for exceedances over *elliptic thresholds* the symmetries γ^t, $t > 0$, are linear expansions, and so are the renormalizations $\alpha_n^{-1}\alpha_{n+1}$.

The theory of coordinatewise extremes is well known, and there exist many good expositions. Our treatment in Chapter II is limited to essentials. Exceedances are treated in Chapter IV. Exceedances over horizontal thresholds describe high risk scenarios associated with a given direction; exceedances over elliptic thresholds may be handled by linear expansions. The theory developed in MS is particularly well suited to exceedances over elliptic thresholds. Arguments for using elliptic thresholds for heavy tailed distributions are given in Section 16.1. The basic limit relation (8) now reads

$$n\alpha_n^{-1}(\pi) \to \rho \text{ weakly on } \varepsilon B^c, \quad \varepsilon > 0, \tag{20}$$

where B is the open unit ball, and α_n are linear expansions. If (20) holds we say that π lies in the *domain* of ρ *for exceedances over elliptic thresholds*, and write $\pi \in \mathcal{D}^\infty(\rho)$. Example 7 is exemplary. It treats an excess measure on $\mathbb{R}^d \setminus \{0\}$ with a spectral measure ρ^* which is uniformly distributed over the unit sphere, and a symmetry group of scalar expansions

$$\gamma^t(w) = e^{\tau t}w, \quad t \in \mathbb{R}. \tag{21}$$

If we allow ρ^* to have any distribution on the unit sphere ∂B, but assume the normalizations $\beta(t)$ to be scalar, then the ellipsoids $\beta(t)(B)$ are balls. The limit relation for the high risk scenarios simplifies:

$$Z^r/r \Rightarrow W, \quad r \to \infty, \tag{22}$$

where Z^r is the vector Z conditioned to lie outside the open ball rB. In this situation it is natural to use polar coordinates and write $Z = R\zeta$ with $R = \|Z\|$. The distribution of $(\zeta, R/r)$, conditional on $R \geq r$, converges to a product measure $d\rho^* \times dG$ on

$\partial B \times [1, \infty)$, where ρ^* is the *spectral measure*, and G a Pareto distribution on $[1, \infty)$ with density $\lambda/r^{\lambda+1}$, $\lambda = 1/\tau$. The spectral measure gives an idea of the directions in which the data extremes cluster; the parameter τ in (21) describes the decay rate of the tails.

Here we have another example of the close relation between symmetry and independence! In this model it is again obvious how to estimate the distribution of the high risk scenarios Z^r for values of r so large that only one or two sample points fall in the complement of the ball rB.

Asymptotic *independence* is *not* the subject of these lectures. Our theory is based on concepts like scale invariance, self-similarity and symmetry. It is geometric and local. Independence is a global analytic assumption. It allows one to draw far-reaching conclusions about extremes, but the techniques are different from those developed here.

So far we have assumed convergence of a one-parameter family of high risk scenarios indexed by horizontal halfspaces H^y, $y \uparrow y_\infty$, or by an increasing family of ellipsoids $E_t = \alpha_t(B)$. These situations yield a limit measure with a one-parameter family of symmetries, the excess measure described in (9). Let us now return to high risk scenarios Z^H where the halfspaces H are allowed to diverge in any direction. For simplicity assume Z has a density. Assume convergence of the normalized high risk scenarios to a non-degenerate vector W on H_+: For each halfspace H of positive mass there exists an affine transformation α_H mapping H_+ onto H such that

$$\alpha_H^{-1}(Z^H) \Rightarrow W, \quad \mathbb{P}\{Z \in H\} \to 0. \tag{23}$$

The limit describes the tail asymptotics in every direction. In Section 13 we shall exhibit a continuous one-dimensional family of excess measures ρ_τ, $\tau \geq -2/h$, $h = d - 1$, corresponding to the multivariate GPDs. The densities, standardized to have a simple form, and without norming constant, are

$$g_0(u, v) = e^{-(v+u^T u/2)}, \qquad w = (u, v) \in \mathbb{R}^{h+1}, \quad \tau = 0,\ J_0 = \{v \geq 0\};$$

$$g_\tau(w) = 1/\|w\|^{d+\lambda}, \qquad w \neq 0, \quad \tau = 1/\lambda > 0,\ J_0 = \{v \geq 1\};$$

$$g_\tau(u, v) = (-v - u^T u/2)_+^{\lambda-1}, \qquad v < -u^T u/2, \quad \tau = -1/(h/2 + \lambda) < 0,\ J_0 = \{v \geq -1\}.$$

The reader may recognize the Gauss-exponential and the Euclidean Pareto excess measure from the examples above. In all cases the vertical coordinate of the high risk limit distribution determined by the restriction of ρ_τ to J_0 has a univariate GPD, with Pareto parameter τ, see (5). For $\tau < 0, \lambda = 1$, the excess measure ρ_τ is Lebesgue measure on the paraboloid $\{v < -u^T u/2\}$. For $\tau = -2/h$ the excess measure is singular. The symmetry of the excess measures ρ_τ, $\tau \geq -2/h$, is impressive. Instead of the one-parameter group γ^t, $t \in \mathbb{R}$, in (9) there now is a symmetry group of dimension $2, 4, 7, \ldots$ for $d = 2, 3, 4, \ldots$. Many halfspaces have finite mass. The probability distributions $d\rho^J = 1_J d\rho/\rho(J)$ associated with such halfspaces all have the same shape. The measure ρ has the *tail property* to an excessive degree. The domains of attraction, $\mathcal{D}(\tau)$, of the measures ρ_τ are investigated in Chapter III.

Before giving a detailed description of the contents of the various chapters we still want to consider two topics: the relation to the multivariate theory of coordinatewise maxima, and the range where the theory will apply.

How do coordinatewise maxima fit in?

The subject of this book may be described as *geometric extreme value theory* since we are looking at the behaviour of the extreme points of sample clouds as the number of data points increases without bound. We are concerned with the convex hull, but also with the points of the cloud below the surface. Since we are zooming in at the scale of individual sample points, the limit, if we assume convergence, has to be a Poisson point process whose mean measure is finite and positive on some halfspace. Such limits were first considered by Eddy [1980].

The geometric approach and the analytic, coordinatewise approach are complementary. The geometric theory is interested in linear combinations of coordinates, the analytic theory in maxima of coordinates. There is a difference in interpretation. In the geometric theory for exceedances over horizontal or elliptic thresholds there is one variate (the vertical, or the radial) that measures risk, and an h-dimensional ancillary vector; in the analytic theory all d coordinates play an equal role. The geometric approach looks at exceedances, the analytic approach at maxima. In the univariate situation the two theories are equivalent. In higher dimensions the relation between extremes and exceedances is most clearly seen in the behaviour of sample clouds, and the limiting Poisson point process. In the geometric approach the mean measure of the limit process is called an excess measure, in the analytic approach, it is called an exponent measure; but actually these two terms denote the same[6] object.

One might say that the theory of coordinatewise maxima is concerned with high risk scenarios on sets that are not halfspaces, but complements of shifted negative orthants. Instead of divergent sequences of halfspaces H_n one considers sets of the form

$$[-\infty, \infty)^d \setminus [-\infty, a_n), \quad a_n \in \mathbb{R}^d, \tag{24}$$

where the points a_n increase towards the upper endpoint of the df. Since the complement of a shifted negative orthant contains many halfspaces, convergence of the coordinatewise maxima implies convergence of high risk scenarios on many half-

[6] Exponent measures may give mass to hyperplanes at $-\infty$; excess measures live on open subsets of $\mathbb{R}^d$. The differences will be discussed more fully at various points in these lectures.

spaces. Below we formulate a result that expresses these ideas. One has to distinguish between heavy and light tails. The upper tail of a df F on $\mathbb{R}$ is *light* if

$$t^m(1 - F(t)) \to 0, \quad t \to \infty, \ m = 1, 2, \ldots. \tag{25}$$

It is *heavy* if there exists an integer $m \geq 1$ such that $t^m(1 - F(t)) \to \infty$ for $t \to \infty$.

If all components have heavy upper tails the relation between coordinate maxima and exceedances is simple. One assumes that Z has non-negative components. The exponent measure lives on $[0, \infty)^d \setminus \{0\}$. It is an excess measure, whose symmetries γ^t are linear diagonal expansions for $t > 0$. The max-stable limit $G = \lim F^n \circ \alpha_n$ has the form $G = e^{-R}$, where R is the distribution function of the excess measure. So the df G determines the mean measure, and hence the distribution, of the Poisson point process describing the asymptotic behaviour of the sample clouds. The normalizations α_n are diagonal matrices. Weak convergence $F^n \circ \alpha_n \to G$ implies weak convergence $n\alpha_n^{-1}(dF) \to dR$ on the complement of any ε-ball εB centered in the origin.

Vectors whose components have light upper tails have exponent measures that may charge planes and lines at $-\infty$. The normalizations are *CAT*s, *coordinate affine transformations*, $z_i = a_i w_i + b_i$, $i = 1, \ldots, d$. Let us show how coordinatewise extremes for light tails fit in.

Proposition 8. *Let Z have df F with marginals F_i having light upper tails. Suppose Z lies in the domain of attraction $\mathcal{D}^\vee(\rho)$ for coordinatewise maxima: There exist CATs α_n such that*

$$\begin{aligned} &F^n(\alpha_n(w)) \to G(w) = e^{-R(w)} \text{ weakly,} \\ &R(w) = \rho([-\infty, \infty)^d \setminus [-\infty, w)), \quad w \in \mathbb{R}^d. \end{aligned} \tag{26}$$

Choose $q \in \mathbb{R}^d$ such that $H_i(q_i) < 1$ for $i = 1, \ldots, d$, and set $a_n = \alpha_n(q)$ and $Q = [-\infty, \infty)^d \setminus [-\infty, q)$. Then $\alpha_n^{-1}(Z^{(-\infty, a_n)^c}) \Rightarrow W$, where W has distribution $1_Q d\rho/\rho(Q)$. Let $J = \{\xi \geq c\}$ß$\mathbb{R}^d$ with $\xi \in (0, \infty)^d$. If $\mathbb{P}\{W \in \mathbb{R}^d\}$ is positive, then

$$\alpha_n^{-1}(Z^{H_n}) \Rightarrow W^J, \quad H_n = \alpha_n(J), \tag{27}$$

where W^J has distribution $1_J d\rho/\rho(J)$.

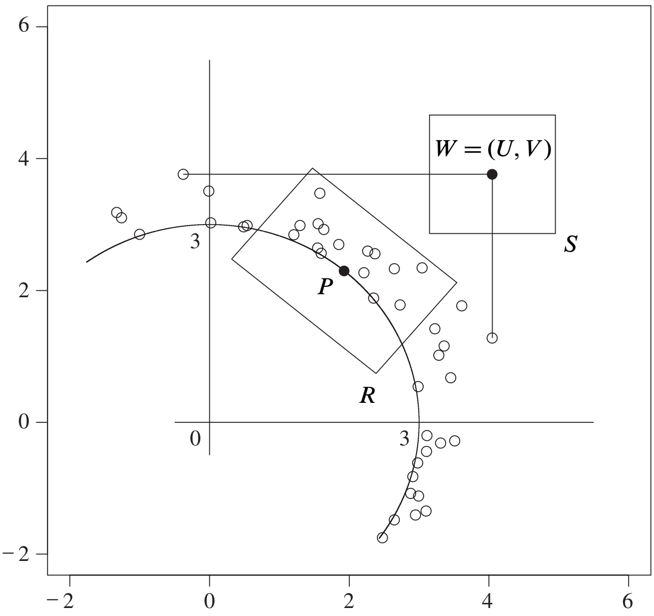
Proof. Relation (26) is standard; see Theorem 7.3. In the limit relation (27) the crucial point is that the condition $\mathbb{P}\{W \in \mathbb{R}^d\} > 0$ ensures that $\rho(J)$ is positive. This implies

$$\mathbb{P}\{Z \in H_n\}/\mathbb{P}\{Z \notin (-\infty, a_n)\} \to \rho(J)/\rho(Q).$$

For JßQ the result follows by a simple conditioning argument. The general case follows by the symmetry of ρ. $\square$

What happens if $\mathbb{P}\{W \in \mathbb{R}^d\} = 0$ in the proposition above?

The figure below suggests that a more geometric approach which zooms in on a boundary point of the sample rather than on the max-*vertex* may be useful in certain situations.



The rectangle R around the point P contains more information about the edge of this 10 000 point sample from the normal distribution than the square S around the max-vertex $W = (U, V)$.

The standard normal distribution on the plane lies in the domain $\mathcal{D}^{\vee}(\rho)$ of the max-stable df $\exp -(e^{-u} + e^{-v})$ (independent *Gumbel* marginals). In order to describe the coordinatewise maximum, the sample cloud is enclosed in a coordinate rectangle. The coordinatewise maximum is the upper right hand corner of the rectangle, the max-vertex. Now the scaling is crucial. For a heavy tailed distribution, a *spherical Student* distribution for instance, the scaling preserves the origin, which remains in the picture. For the light tailed *Gaussian distributions*, however, the normalization zooms in on a small (empty) square around the max-vertex. It fails even to see the shape of the sample cloud. As a result all bivariate Gaussian densities with standard normal marginals yield the same bivariate extreme asymptotics.

Let us now say a few words on the applicability of the theory presented in this book.

Our approach to risk is that of an observer, rather than a risk manager. Given a multivariate data set describing the past behaviour, and a loss function, our aim is to describe the tail behaviour of the distribution underlying the data set. Such a description enables one to construct large synthetic samples, and to study the behaviour of the associated random losses. This procedure is known as *stress testing* and *scenario analysis*. We are not concerned with the problem of changing the parameters of the underlying distribution, redirecting the dynamics which produced the data set, or altering the loss function by a suitable form of risk transfer. These issues are treated in McNeil, Frey & Embrechts [2005] for financial risk; and for risk in the realm of reliability engineering in Bedford & Cooke [2001].

Einstein showed that the erratic movement of pollen grains suspended in a drop of water, as observed by Brown at the beginning of the 19th century, could be described by smooth probability distributions exhibiting a large degree of symmetry. Complex dynamical systems may give rise to symmetric probability distributions. Symmetries in a data set may reflect regularities inherent in the dynamical system which produces the data. If so, the symmetries are likely to persist. The validity of our model depends on this persistence of the symmetry. For Brownian motion as a model for the movement of pollen grains, Einstein [1906] imposed a bound of 10^{-7} seconds for applicability. So too, in financial or meteorological or biological applications the symmetry will break down at a certain level[7].

To fix ideas let us posit an ultimate probability p_0 in the range 10^{-99} to 10^{-20}. Halfspaces with probabilities below this value have no reality for risk management. Replacing the conditional distribution on such a halfspace by any other distribution does not influence the policy of the risk manager. This means that the endless variety of the ever slower pirouettes performed by the sequence of ellipses E_n in Exercise 6 above, forms part of the mathematical theory, but has no bearing on risk management, since the probabilities $\mathbb{P}\{Z \in E_n^c\}$ lie below the threshold value p_0 after a few hundred terms. By the same argument the existence of moments falls outside the range of realistic risk theory. (Convergence of integrals is determined by the behaviour of the distribution on invisible halfspaces.) This collateral result is not as disturbing as it may seem on first sight. For heavy tails the value of the exponent where the moment first fails to exist, is a convenient measure of risk, but for a realist the difference between a Gaussian distribution and a *Cauchy distribution* is established by the behaviour of samples of size a hundred. She is not interested in the tail behaviour at risk levels 10^{-99}. The assumption of an ultimate probability p_0 also has advantages. It provides

[7] A finite universe does not preclude a model with infinite upper tail for spatial variables. String theory tells us that four-dimensional space-time may break down at magnitudes of 10^{-30} meters to reveal six hidden dimensions curled up into compact sets. This does not mean that models with continuous densities for spatial or temporal quantities are invalid.

us with the liberty to choose the behaviour of the distribution on invisible halfspaces to suit our fancy. We find it convenient to assume convergence of *high risk scenarios*.

Having said this much on applicability, we may now proceed with our proper task, the mathematical investigation of the consequences of the assumption that high risk scenarios converge.

Contents

The book consists of twenty lectures, grouped into five chapters. There is a basics chapter on point processes, a final chapter listing open problems, and in between there are three chapters covering three different topics: coordinatewise extremes, multivariate GPDs, and exceedances over thresholds (horizontal and elliptic).

Chapter II treats the basic univariate extreme value theory, and provides an overview of the theory of coordinatewise maxima. Our focus is on exponent measures rather than max-stable dfs. The Chapters III and IV form the body of the book. They present two different views on *high risk scenarios*. In Chapter III the high risk scenarios Z^H converge to a common limit law, in whatever direction the halfspaces H diverge. This restricts the class of limit laws. We present a one-parameter family of limit laws, the multivariate GPDs. It is not known whether other limit laws exist. Chapter III is a relatively self-contained account of what is known about the domains of attraction of the multivariate GPDs. The elegant theory of multivariate GPDs, and their domains of attraction, should be useful in situations where the sample cloud is bland or where the dimension is high, and where one is interested in the overall extremal behaviour rather than the asymptotic behaviour in a particular direction. Such an approach may be of interest to the supervisor or regulator; it allows a diversified view of the extremal behaviour of widely varying positions in the underlying market. The theory presented in Chapter IV is different. In this chapter we look at exceedances. For linear thresholds this means that we look at halfspaces moving off in a given direction. Such a model is of interest to the trader or risk manager taking directional positions in the underlying market. For simplicity we assume the thresholds horizontal. For heavy tails, elliptic thresholds are more natural since there is no difference between the local and the global theory. This is explained in Section 16.1. Heavy tailed vectors are normalized by linear contractions. The theory of exceedances presented in Chapter IV has the same structure as the theory of coordinatewise maxima. The limit laws are known. The excess measure, like the exponent measure, satisfies a one-parameter group of symmetries. The normalizations are more complex than the CATs used for coordinatewise extremes, and call for a geometric approach. The asymptotics may be handled by regular variation. A complete characterization of the domains of attraction is available for a number of limit laws. It is presented in Sections 15.2 and 16.7.

These notes offer probability theory rather than statistics. If one accepts the idea that excess measures may occur as the mean measure of a limiting Poisson point process describing the asymptotic behaviour of sample clouds at their edge, then good estimates of the excess measure and the normalizations allow one to simulate large samples that may then be used in risk analysis. The task of the probabilist is to analyse the model. What do excess measures look like? For any given excess measure, what does the domain of attraction look like? What normalizations are allowed? What moments will converge? What does the convex hull of the sample cloud look like? Does it converge? These are some of the questions that will be addressed in the present text.

Chapter I treats point processes. The first four sections are standard theory: An intuitive introduction; the Poisson point process as a limit of superpositions of sparse point processes; the distribution of point processes; and their convergence. In Section 5 extremes enter the scene. We consider the n-point sample cloud N_n from the probability distribution $\pi_n = \alpha_n^{-1}(\pi)$ on $\mathbb{R}^d$, and assume vague convergence $N_n \Rightarrow N_0$ to a Poisson point process N_0. In applications the mean measure ρ of the limit process is an infinite Radon measure on an open set O, for instance $\mathbb{R}^d$ or $\mathbb{R}^d \setminus \{0\}$. We are interested in convergence of the convex hulls of the sample clouds. That means that we have to determine the halfspaces $H ß O$ on which ρ is finite, and on which the mean measures $n\pi_n$ converge weakly to ρ. The class of such halfspaces determines two open cones in the dual space, the *intrusion cone* Δ and the *convergence cone* Γ. Modulo some minor regularity conditions convex hulls converge if Δ is non-empty, and $\Gamma = \Delta$. We also discuss loss functions, and approximate their integrals by sample sums.

Chapter II treats the theory of maxima. It consists of two sections. Section 6 treats the univariate situation. The domains of attraction for exceedances and maxima coincide. The domain of attraction $\mathcal{D}^+(0)$ of the exponential law is described in terms of densities which satisfy a von Mises condition. The section also contains an elementary proof of Bloom's basic theorem on self-neglecting functions. The second part, Section 7, assumes some acquaintance with the theory of coordinatewise extremes. We concentrate on the domain of max-stable distributions with standard exponential marginals on $(-\infty, 0)$. This allows us to treat exponent measures that charge coordinate planes in $-\infty$. The sample copula yields a simple tool for investigating the dependency structure. Non-linear normalization of the coordinates provides a direct link to copula theory.

Chapter III starts with an extensive introductory section treating applications, examples, and the general asymptotics of high risk scenarios. For coordinatewise maxima, powers of distribution functions play an important role; in the theory of high risk scenarios one encounters powers of densities. Unimodal densities (with convex level sets) seem to reflect quite well the shape of the sample clouds to which the theory applies. Pointwise convergence of densities often is a first step towards the derivation of limit theorems for distributions. We establish simple asymptotic expressions for excess probabilities, $\mathbb{P}\{Z \in H\}$, in terms of densities.

There is a continuous one-parameter family of multivariate GPDs, indexed by a shape parameter τ. As in the univariate case the family falls apart in three power families, corresponding to the sign of the parameter τ. Excess measures of the heavy tailed distributions, corresponding to $\tau > 0$, have a spherically symmetric density. Tails of distributions in $\mathcal{D}(\tau)$ for $\tau > 0$ may be approximated by tails of elliptic Student distributions. Distributions in $\mathcal{D}(\tau)$ for $\tau < 0$ have bounded support; the convex hull of the support is egg-shaped. The latter distributions receive only cursory treatment; their role in risk theory is limited.

Special attention is given to $\mathcal{D}(0)$, the domain of the Gauss-exponential law. As in the univariate setting, the parameter value $\tau = 0$ is the most interesting mathematically. Section 9 introduces the class $\mathcal{RE}$ of *rotund-exponential* densities. These have the form

$$f(z) \propto e^{-\psi \circ n_D(z)}$$

where the function $e^{-\psi}$ satisfies the von Mises condition for the univariate domain of attraction $\mathcal{D}^+(0)$. The function n_D is the gauge function of a *rotund* set D. Such a set is egg-shaped: convex, open, and bounded, it contains the origin, and the boundary is C^2 with continuously varying positive definite curvature. One may think of the gauge function as a norm, generated by the set D, when $-D = D$. The rotund-exponential densities extend the class of spherical *Weibull densities* $ce^{-\|z\|^r}$, $r > 0$. They allow us to treat sample clouds whose central part is egg-shaped rather than elliptic. Their simple structure should make them tractable for statistical analysis. The normalizations α_H may be written down explicitly in terms of ψ and D. In Section 9 we prove pointwise convergence as H diverges:

$$f(\alpha_H(w))/f(\alpha_H(0)) \to e^{-(v+u^T u/2)}, \quad w = (u, v) \in \mathbb{R}^{h+1}; \tag{28}$$

in Section 10 we prove $\mathbf{L}^1$-convergence of these quotients for unimodal densities. Section 11 introduces flat functions. Flat functions play the same role in the multivariate theory as slowly varying functions do in the univariate theory. Finally it will be shown that the normalizations induce a Riemannian metric on the convex open set $O = \{f > 0\}$. Conversely, the metric determines the normalizations, and hence the global structure of distributions in the domain $\mathcal{D}(0)$.

We mention two results from Section 13 that should give an impression of the scope and of the limitations of the theory of multivariate GPDs.

- Let A be a linear map from $\mathbb{R}^d$ onto $\mathbb{R}^m$. If the vector $Z \in \mathbb{R}^d$ lies in $\mathcal{D}(\tau)$, then so does $A(Z)$.
- A vector $Z \in \mathcal{D}(\tau)$ with independent components is Gaussian (and $\tau = 0$).

Chapter IV treats exceedances over horizontal and elliptic thresholds. The first two sections treat horizontal thresholds. The first is theoretical. We prove the Extension Theorem: If the high risk scenarios Z^H for horizontal halfspaces H, properly normalized, converge in distribution to a non-degenerate limit vector W, then there is an excess measure ρ, and the normalized sample clouds converge in distribution to a Poisson point process N_0 with mean measure ρ weakly on all horizontal halfspaces on which ρ is finite. This is the step from (6) to (11) and (12). Next we determine the limit laws and excess measures for exceedances over horizontal thresholds. Up to a non-essential multiplicative constant, the excess measure ρ is determined by its symmetry group γ^t, $t \in \mathbb{R}$, and a probability measure ρ_0^* on $\mathbb{R}^h$, the *spectral probability measure*, which is the conditional distribution of U given $V = 0$, where $W = (U, V)$ is the high risk limit vector on H_+. Section 14.9 describes the situation in $\mathbb{R}^3$.

A question that is important for applications is: To what extent may one relax the condition that the halfspaces be horizontal? The excess measure is finite for a horizontal halfspace J_0 by definition. Is it also finite for non-horizontal halfspaces close to J_0? Does weak convergence $n\alpha_n^{-1}(\pi) \to \rho$ in (8) hold on such halfspaces? A related question is whether the convex hull of the normalized sample cloud converges to the convex hull of the limiting Poisson point process. The book gives partial answers to these questions.

A considerable part of Chapter IV is taken up by the analysis of specific examples, and a discussion of the relation to the limit theory for coordinatewise extremes. Section 15 investigates the excess measures and domains of attraction for three simple symmetry groups γ^t: vertical translations, scalar contractions, and scalar expansions. For vertical translations a complete description of the domains of attraction is given.

The next two sections of Chapter IV treat heavy tailed distributions normalized by linear contractions. Here we shall work with elliptic thresholds. Section 16 presents the basic theory. The introduction to this section gives more information. The main result is a complete characterization of the domain of attraction $\mathcal{D}^\infty(\rho)$ for excess measures with a continuous positive density. Section 17 contains examples, and a more detailed analysis of the domain of attraction in the case of scalar symmetries. We also give a careful analysis of the relation between limit laws for exceedances over elliptic thresholds and multivariate regular variation. For very heavy tails sample sums are determined asymptotically by the extremes, and domains of attraction for *operator stable* distributions and for excess measures coincide: $\mathcal{D}^{OS}(\rho) = \mathcal{D}^\infty(\rho)$. In this situation excess measures may be interpreted as Lévy measures for multivariate stable processes without a Gaussian component. The theory for exceedances over elliptic thresholds coincides with the limit theory for sums of independent vectors developed in MS. The theory for exceedances may thus provide a simple introduction to the limit theory for operator stable distributions.

The theoretical results on regular variation in groups and regularly varying multivariate probability distributions developed by Meerschaert and Scheffler in their monograph MS are essential for a deeper understanding of the domain $\mathcal{D}^\infty(\rho)$. The Spectral Decomposition Theorem (*SDT*) clears up the mysterious disparity between the domains of attraction of excess measures with scalar symmetries and those with diagonal non-scalar symmetries. Section 18 contains a brief introduction to the theory of multivariate regular variation, and to the SDT. The second half of this section treats the general theory of excess measures on $\mathbb{R}^d$. Finally Section 18.14 presents an example that shows that the three approaches to the asymptotics of multivariate sample extremes developed in these notes – coordinatewise maxima, exceedances over linear thresholds, and exceedances over elliptic thresholds – may yield conflicting results.

Chapter V lists some fifty open problems. Together with the hundred examples scattered throughout the text these serve to enliven the presentation, and to mark the boundary of our present knowledge. The second part of the chapter describes some of the difficulties that a statistician may encounter if she decides to apply the theory to concrete data sets.

We have provided this lengthy introduction because the book does not have a clear linear structure. It is a collection of essays. Moreover, there is a certain ambiguity in the subject matter. Basically the book is about high risk scenarios. Chapter III may be read from this point of view without bothering about point processes. The reader will then observe that each of the limit laws extends naturally to an infinite measure, and he will observe that this excess measure has an extraordinary degree of symmetry. The excess measure is infinite, but has a simple probabilistic interpretation: The normalizations that are used to obtain a non-degenerate limit law for the *high risk scenarios* may be applied to the sample clouds to yield a limiting Poisson point process. The excess measure is the mean measure of this Poisson point process. This convergence of sample clouds is the second point of view. From this point of view it is natural to start with an overview of point processes, Sections 1–5. The point process approach unifies the univariate theory of extremes and exceedances in Section 6. Section 7 treats the limit theory for coordinatewise maxima under linear and non-linear coordinatewise normalizations from the same point of view. A natural counterpart to the limit theory for high risk scenarios developed in Chapter III is the limit theory for exceedances over thresholds developed in Chapter IV. The two sections on horizontal thresholds are only loosely connected, as are the next two sections on exceedances over elliptic thresholds. In fact it might be more instructive to start with one of the examples in Section 15 or in Section 17 in order to gain an impression of this part of the theory, rather than working through the technicalities leading up to the Extension Theorem in Section 14. Similarly the open problems in Chapter V should give a good impression of the scope of geometric extreme value theory, as treated in this monograph. Sections 5 and 18 have a special standing. They contain background material. Section 5 looks into the question: How does one describe convergence of sample clouds to a limiting Poisson point process in terms of halfspaces? Section 18 treats multivariate regular variation and the general theory of excess measures and their symmetries. It contains subsections on the Meerschaert Spectral Decomposition Theorem, on Lie groups, and on the Jordan form of a matrix.

The book treats only a part of extreme value theory. For extremes of stationary processes, of Gaussian fields, or of time series, the reader may consult Berman [1992], Davis & Resnick [1986], Dieker [2006], Finkenstädt & Rootzén [2004] or Leadbetter, Lindgren & Rootzén [1983], and the references cited therein. For extremes in Markov sequences see Perfekt [1997]; for exceedances see Smith, Tawn & Coles [1997]. Extremes in function spaces and for stochastic processes are treated in Giné, Hahn & Vatan [1990], de Haan & Lin [2003] and Hult & Lindskog [2005]. Limit behaviour of convex hulls has been investigated in Eddy & Gale [1981], Groeneboom [1988], Brozius & de Haan [1987], Baryshnikov [2000], Bräker, Hsing & Bingham [1998], and Finch & Hueter [2004]. Statistics for coordinatewise extremes are treated in de Haan & Ferreira [2006]; for heavy tails see Resnick [2006].

Interest in exceedances over linear thresholds is not new. We mention early papers by de Haan [1985], de Haan & de Ronde [1998], and Coles & Tawn [1994]. The last two contain nice applications to meteorological data. Exceedances over linear thresholds seem to fit snugly within the framework of coordinatewise extremes, as is shown by Proposition 8. It is only by taking a geometric point of view that one becomes aware of the limitations imposed by the coordinatewise approach, due to the restriction to CATs in the normalization. The strong emphasis on coordinates in multivariate extreme value theory so far may also explain why the relevance of the theory of multivariate regular variation developed in MS has not been realized before.

Notation

Halfspaces are closed, and denoted by $H = \{\zeta \geq c\}$ or J. Horizontal halfspaces have the form $\{y \geq c\}$, or $\{\eta \geq y\}$, where η is the vertical coordinate. We often use the decomposition $z = (x, y) \in \mathbb{R}^{h+1}$ into a *vertical component* $y \in \mathbb{R}$ and a horizontal part $x \in \mathbb{R}^h$. So $h = d - 1$ is the dimension of the horizontal coordinate plane $\{y = 0\}$.

The set of affine transformations $\alpha\colon w \mapsto Aw + b$ on $\mathbb{R}^d$ is denoted by $\mathcal{A} = \mathcal{A}(d)$. If the linear part is a diagonal matrix with positive entries we call α a *CAT* (coordinatewise affine transformation). CATs are simple to handle, and they are the transformations used in coordinatewise extreme value theory. The CATs form a closed subgroup of $\mathcal{A}$. So do the translations $w \mapsto w + b$ and the set $\mathcal{A}^h$ of affine transformations that map horizontal halfspaces into horizontal halfspaces. For the closed subgroups of linear transformations, and the compact subgroups of orthogonal and special orthogonal transformations (with determinant one) we use the standard notation $\mathrm{GL}(d)$, $\mathrm{O}(d)$, and $\mathrm{SO}(d)$. We write GL, O and SO if the dimension is not specified.

In general π denotes the distribution of a vector $Z = (X, Y)$ in $\mathbb{R}^{h+1}$. We assume that Z lies in the domain of attraction of a limit vector $W = (U, V)$. This means that

$W_n = \alpha_n^{-1}(Z^{H_n}) \Rightarrow W$ for certain sequences of halfspaces H_n where Z^H denotes the vector Z conditioned to lie in the halfspace H. We regard α_n as transformations from (u, v)-space to (x, y)-space, and hence use the inverse α_n^{-1} to normalize. This corresponds to the usual practice in the univariate case where one subtracts a location parameter and divides by a scale parameter.

The table below lists the domains of attraction introduced in the text:

$\mathcal{D}^+(\tau),\ \tau \in \mathbb{R}$	(6)	domain of the univariate GPD G_τ
$\mathcal{D}(\tau),\ \tau \geq -1/2h$	(8)	... of the multivariate GPD π_τ for high risk scenarios
$\mathcal{D}^\vee(\rho)\quad \mathcal{D}^\wedge(W)$	(7)	... for coordinatewise maxima and minima, normed by CATs
$\mathcal{D}^\uparrow(\rho)\quad \mathcal{D}^\uparrow(C)$	(6,7)	...... normed by monotone transformations
$\mathcal{D}^h(\rho)$	(14)	... for exceedances over *horizontal thresholds*
$\mathcal{D}^\infty(\rho)$	(16)	... for exceedances over *elliptic thresholds*
$\mathcal{D}^{OS}(\rho)$	(17)	domain of operator stable vectors with Lévy measure ρ

Domains of attraction (in brackets the section in which they are introduced).

The argument of $\mathcal{D}$ in the table above is the *Pareto parameter* τ, or the excess measure, exponent measure, or Lévy measure ρ, or the limit vector W, or the max-stable copula C. One could add a number of extra parameters, the dimension d, the generator of the symmetry group, restrictions on the normalizations in the form of a subgroup (CATs, linear maps, diagonal maps, scalar maps, translations, etc.); one could specify that convex hulls converge, that densities converge, or that densities be unimodal. Since the theory is still in flux we restrict the notation to essentials.

We mention three possible sources of confusion:

1) The high risk limit vector W lives on a halfspace J_0. In the limit relation $\alpha_H^{-1}(Z^H) \Rightarrow W$ it is assumed that $\mathbb{P}\{Z \in H\}$ is positive – in order to have well-defined high risk scenarios –, that $\mathbb{P}\{Z \in H\} \to 0$ – in order to have an interesting limit relation –, and that $\alpha_H(J_0) = H$ – in order to ensure that the normalized high risk scenarios $W_H = \alpha_H^{-1}(Z^H)$ live on J_0. Often $J_0 = H_+$, the upper halfspace, but it may sometimes be convenient to choose some other halfspace, for instance $J_0 = \{v \geq j_0\}$ with $j_0 = 1$ or $j_0 = -1$, or to leave the precise form of J_0 unspecified. This confusion already exists in the univariate case where Pareto distributions may be standardized to live on $[0, \infty)$ or on $[1, \infty)$.

2) The *spectral measure* is a finite measure, which together with the one-parameter group of symmetries γ^t, $t \in \mathbb{R}$, determines the excess measure. One may take it to be a probability measure by dividing ρ by a harmless positive constant. The spectral measure lives on $\mathbb{R}^h$ or on the unit sphere. It has the advantage over the excess measure that it is arbitrary. The *spectral measure* bears no relation with the *spectral decomposition*. The latter concerns the symmetries and the normalizations.

3) Exceedances over *elliptic thresholds* is an alternative to exceedances over linear thresholds which is particularly well suited to distributions with heavy tails. Actually, as explained in Section 16.1, we shall hardly consider *high risk scenarios* of the form Z^{E^c}. The limit relation $\alpha_E^{-1}(Z^{E^c}) \Rightarrow W$ only occurs in Section 17.7. The really interesting relation is

$$e^t \beta(t)^{-1}(\pi) \to \rho \text{ weakly on } \varepsilon B^c, \quad t \to \infty, \ \varepsilon > 0.$$

In the terminology of MS this just says that the probability measure π *varies regularly* with exponent C, where the one-parameter linear expansion group $\gamma^t = e^{tC}$, $t \in \mathbb{R}$, satisfies $\gamma^t(\rho) = e^t \rho$.

A function $f \geq 0$ on $\mathbb{R}^d$ is called *unimodal* if the level sets $\{f > c\}$ are convex for $c > 0$.

Limits are often one-sided. If y_∞ is the upper endpoint of a distribution on $\mathbb{R}$ then $y \to y_\infty$ always means convergence from below. In limits for sequences indexed by n we implicitly assume $n \to \infty$.

B is the open centered Euclidean unit ball, E_n are open ellipsoids.

N_n and N are point processes, usually on $\mathbb{R}^d$, or on an open set Oß$\mathbb{R}^d$.

$d = h + 1$ is the dimension of the vectors $Z = (X, Y)$ and $W = (U, V)$ in $\mathbb{R}^{h+1}$.

τ is the Pareto (shape) parameter, ρ the excess measure, π a probability distribution. The vertical component of the measure ρ on $\mathbb{R}^{h+1}$ is $\tilde{\rho}$, and $\tilde{\alpha}$ for $\alpha \in \mathcal{A}^h$ is the univariate affine transformation of the vertical coordinate induced by α.

The abbreviations rv, df and iid are standard in probability theory and statistics.

The relation $\stackrel{\mathrm{d}}{=}$ denotes equality in distribution; $\Rightarrow$ denotes convergence in distribution, $p := \mathbb{P}\{X_n \geq c\}$ defines p as the probability of a certain event. We use the notation $a_n \ll b_n$ or $a_n = o(b_n)$ to signify that $a_n/b_n \to 0$; $a_n \sim b_n$ means that $a_n/b_n \to 1$ and $a_n \asymp b_n$ means that the quotients a_n/b_n and b_n/a_n are bounded eventually. We use $\mathrm{int}(E)$ and $\mathrm{cl}(E)$ to denote the interior and the closure of a set E, and $c(E)$ to denote the *convex hull*.

The basic terminology and notation have been introduced in the Preview. Special notation could not be avoided completely. One may always consult the Index at the end of the book. In the text itself the index entry is in bold face or emphasized by printing it in italics to contrast with the surrounding text. In the case of multiple entries, the bold face page number in the Index will guide the reader to the formal definition. In the Bibliography the numbers in square brackets refer to the pages on which the item is cited.

The table of contents is detailed. Starred sections may be skipped. They are technical or treat subjects which are not used in the remainder of the text. There are a number of sections treating specialized subjects: Sections 2.6 and 17.6 treat Lévy processes and convergence to operator stable processes: Section 6.7 discusses self-neglecting functions; Section 5.3 treats halfspaces and convex sets; Section 18.1 gives a brief introduction to multivariate regular variation; Sections 16.9 and 18.4 discuss the Spectral Decomposition Theorem; Section 18.8 describes the excess measures on the plane; Section 18.9 treats orbits of one-parameter groups of affine transformations on $\mathbb{R}^d$; Section 18.13 treats Lie groups, and Section 18.12 treats the Jordan form, and the spectral decomposition of one-parameter groups.

EKM and MS denote two books which will be cited frequently. Embrechts, Klüppelberg & Mikosch [1997] contains the fundamental material on which this monograph is built, and is an excellent guide to applications in finance and risk theory. Meerschaert & Scheffler [2001] contains an in-depth exposition of the analytic theory of multivariate regular variation for linear transformations, functions, and measures.