The Annals of Applied Probability
2007, Vol. 17, No. 2, 537–571
DOI: 10.1214/105051606000000835

LIMIT LAWS FOR RANDOM VECTORS WITH AN EXTREME COMPONENT

BY JANET E. HEFFERNAN AND SIDNEY I. RESNICK[1]

Lancaster University and Cornell University

Models based on assumptions of multivariate regular variation and hidden regular variation provide ways to describe a broad range of extremal dependence structures when marginal distributions are heavy tailed. Multivariate regular variation provides a rich description of extremal dependence in the case of asymptotic dependence, but fails to distinguish between exact independence and asymptotic independence. Hidden regular variation addresses this problem by requiring components of the random vector to be simultaneously large but on a smaller scale than the scale for the marginal distributions. In doing so, hidden regular variation typically restricts attention to that part of the probability space where all variables are simultaneously large. However, since under asymptotic independence the largest values do not occur in the same observation, the region where variables are simultaneously large may not be of primary interest. A different philosophy was offered in the paper of Heffernan and Tawn [*J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** (2004) 497–546] which allows examination of distributional tails other than the joint tail. This approach used an asymptotic argument which conditions on one component of the random vector and finds the limiting conditional distribution of the remaining components as the conditioning variable becomes large. In this paper, we provide a thorough mathematical examination of the limiting arguments building on the orientation of Heffernan and Tawn [*J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** (2004) 497–546]. We examine the conditions required for the assumptions made by the conditioning approach to hold, and highlight simililarities and differences between the new and established methods.

1. Introduction. Extreme value theory motivates statistical models for the tails of multivariate probability distributions. All such theory relies on some form of asymptotic argument; it is this limiting argument which forces us into the distributional tails and allows the examination of the extremal behavior of random vectors.

The first such arguments relied upon limiting behavior imposed by considering componentwise maxima of random vectors [19, 15, 30, 34]. This approach was extended by Coles and Tawn [5, 6], de Haan and de Ronde [16] in a multivariate

Received February 2005; revised October 2006.
[1]Supported in part by NSF Grant DMS-03-03493.
AMS 2000 subject classifications. Primary 60G70; secondary 62G32.
Key words and phrases. Conditional models, heavy tails, regular variation, coefficient of tail dependence, hidden regular variation, asymptotic independence.

analogue of the one-dimensional threshold methods of Davison and Smith [39], Smith [8]. The methods provide a rich class of models to describe asymptotic dependence but cannot distinguish between asymptotic independence and exact independence. In response to this weakness, theory and models offering a richer description of asymptotic independence behavior have been developed by Heffernan and Resnick [20], Ledford and Tawn [24–26], Maulik and Resnick [27] and Resnick [36]. The assumptions underlying this broader class of models have been termed *hidden regular variation* which elaborates the concept of the *coefficient of tail dependence*.

Models based on assumptions of multivariate regular variation and hidden regular variation have a common reliance on limiting procedures in which all vector components are scaled by functions increasing to infinity. In the case of asymptotic dependence, reliance only on multivariate regular variation is sufficient since in this case the largest values of the components of the random vector tend to occur together. However, models based on multivariate regular variation fail to distinguish between asymptotic independence and exact independence and as such provide an inadequate description of dependence within the asymptotic independence class. Hidden regular variation attempts to repair this defect by allowing a different scale function which gives nontrivial limit behavior when vector components are simultaneously large. Although the hidden regular variation as typically formulated provides a more satisfactory description of the joint tail of the distribution for asymptotically independent variables, this approach still has practical limitations in applications where interest is in tail regions other than the joint tail. These other tail regions are of practical significance since under asymptotic independence, the largest values of the components of the random vector tend not to occur in the same observation.

The philosophy of examining distributional tails in which one or more but not necessarily all of the vector components are simultaneously large was explained in [21]. They focused on a single variable being large by conditioning on one component of the random vector and finding the limiting conditional distribution of the remaining components as the conditioning variable becomes large. Simulation studies in [21] suggested that this alternative approach is useful in accurately describing a range of qualitatively different dependence structures including asymptotic dependence, asymptotic independence and negative dependence. The approach is flexible and readily applicable to general d-dimensional distributions. However, this new basis for modeling multivariate extremes was criticized in the discussion to the paper as lacking a rigorous theoretical underpinning. The discussion highlighted the need for further work to clarify how the approach extends and/or differs from established methodologies which rely on multivariate regular variation and hidden regular variation.

In this paper, we use the philosophy of Heffernan and Tawn [21] and offer a mathematical framework for a theory of conditional distributions given a component is large. We have changed the formulation of Heffernan and Tawn [21] for two reasons. First, it is difficult to construct an asymptotic theory based on regular conditional distributions which are readily manageable only for the case in which smooth densities are assumed and secondly our formulation readily allows for connections to classical multivariate extreme value theory and regular variation.

1.1. *Content of the paper.* Here are more details about the content of the paper. We consider the distribution of a bivariate random vector (X, Y) on $\mathbb{R}^2$ under the condition that Y is large. Generalizations could be made to the case of a $(d + 1)$-dimensional vector

$$(\mathbf{X}, Y) := \big(X^{(1)}, \dots, X^{(d)}, Y\big)$$

where we seek conditional limits of $\mathbf{X}$ given Y is large. However, we leave such generalizations to subsequent investigations. We assume the distribution function F of Y is in a *domain of attraction* of an extreme value distribution $G_\gamma(x)$, written $F \in D(G_\gamma)$. This means there exist functions $a(t) > 0$, $b(t) \in \mathbb{R}$, such that,

$$F^t\big(a(t)y + b(t)\big) \to G_\gamma(y) \qquad (t \to \infty), \tag{1}$$

weakly, where

$$G_\gamma(y) = \exp\{-(1 + \gamma y)^{-1/\gamma}\}, \qquad 1 + \gamma y > 0, \gamma \in \mathbb{R}, \tag{2}$$

and the expression on the right is interpreted as $e^{-e^{-y}}$ if $\gamma = 0$. See, for example, [7, 9, 12, 31, 34]. We can and do assume

$$b(t) = \left(\frac{1}{1 - F(\cdot)}\right)^{\leftarrow}(t),$$

where for a nondecreasing function U we define the left continuous inverse

$$U^{\leftarrow}(t) = \inf\{y : U(y) \ge t\}.$$

Setting $\overline{F} = 1 - F$, we have relation (1) is equivalent to

$$t\overline{F}\big(a(t)y + b(t)\big) \to (1 + \gamma y)^{-1/\gamma}, \qquad 1 + \gamma y > 0, \tag{3}$$

or taking inverses

$$\frac{b(tx) - b(t)}{a(t)} \to \frac{x^\gamma - 1}{\gamma}, \qquad x > 0. \tag{4}$$

For convenience we write $\mathbb{E}_\gamma := \{y \in \mathbb{R} : 1 + \gamma y > 0\}$. When considering vague convergence, it is convenient to close the interval $\{y \in \mathbb{R} : 1 + \gamma y > 0\}$ on the right and denote by $\overline{\mathbb{E}}_\gamma$ this closure. So, for instance, $\overline{\mathbb{E}}_0 = (-\infty, \infty]$.

In Section 2, we explore the implications of assuming the existence of:

1. Scaling function $a(\cdot) > 0$, and centering function $b(\cdot) \in \mathbb{R}$ so that (1) holds for $F(x) = P[Y \le x]$;

2. Scaling function $\alpha(\cdot) > 0$, and centering function $\beta(\cdot) \in \mathbb{R}$ and a nonnull Radon measure μ on Borel subsets of $[-\infty, \infty] \times (-\infty, \infty]$, such that for each fixed $y \in \mathbb{E}_\gamma$,

   (a) $\mu([-\infty, x] \times (y, \infty])$ is not a degenerate distribution function in x,
   (b) $\mu([-\infty, x] \times (y, \infty]) < \infty$,
   (c) and

$$tP\left[\frac{X-\beta(t)}{\alpha(t)} \le x, \frac{Y-b(t)}{a(t)} > y\right] \to \mu\big([-\infty, x] \times (y, \infty]\big), \tag{5}$$

   at continuity points (x, y) of the limit.

If we interpret (5) as vague convergence (cf. Section A.3) in $M_+([-\infty, \infty] \times \overline{\mathbb{E}}_\gamma)$, the Radon measures on $[-\infty, \infty] \times \overline{\mathbb{E}}_\gamma$, then in fact (5) implies $F \in D(G_\gamma)$ for some $\gamma \in \mathbb{R}$. Also, we will see that (5) is equivalent to assuming the existence of the conditional limiting distribution of the scaled and centered X variable given Y is extreme:

$$P\left[\frac{X-\beta \circ b^{\leftarrow}(t)}{\alpha \circ b^{\leftarrow}(t)} \le x \,\Big|\, Y > t\right] \to \mu\big([-\infty, x] \times (0, \infty]\big), \tag{6}$$

as t converges to the right end point of F. This observation motivates our focusing on the convergence (5).

Thus we make a different assumption from that of Heffernan and Tawn [21], in that in (6) we condition on the event $Y > t$ rather than $Y = t$ as in [21] which requires regular conditional distributions which are only defined up to almost everywhere equivalence. Our formulation also has a natural connection with extreme value theory as it implies Y is in a domain of attraction. In cases where densities exist, the two formulations are similar. See Section 2.5.

Having established conditions for the existence of a limit in (5), in Section 3 we characterize the class of attainable limiting measures. These measures are found to be either product measures or to have a spectral form after a standardization procedure and then transformation to polar coordinates. The standardization renders (5) into a standard multivariate regular variation condition on the cone $[0, \infty] \times (0, \infty]$ and puts us in familiar territory. Relating (5) to standard multivariate regular variation allows us to identify the class of possible limit measures [32, 34, 37].

Section 4 is motivated by the Heffernan and Tawn [21] approach. Instead of normalizing X by deterministic functions of the threshold t, we normalize by functions of the precise value of Y occurring with X. This leads to a product limit form in all cases.

In Section 5, we highlight connections between assumption (5) and standard assumptions of multivariate regular variation and hidden regular variation, and in particular show that under multivariate regular variation, (5) assumes something additional beyond multivariate regular variation only in the presence of asymptotic independence.

Section 6 illustrates our results with a range of examples. Of particular interest is the bivariate Normal example which shows a transformation of X for which the limit (5) does not exist. This leads to Section 7, in which we explore how flexible one can be in the choice of measurement units in which to record X such that the limit measure in (5) does exist. Our results suggest how to construct change of variable functions which will give such a limit.

Section 8 returns in more detail to the modeling assumptions made by Heffernan and Tawn [21] which motivated the work of this paper, and discusses the implications of the new results for their conditional approach to modeling multivariate extreme values.

1.2. *Symbol and concept glossary.* The Appendix contains several appendices reviewing and referencing needed background. We merely list here some concepts and symbols; explanations and references in the appendices can be consulted as needed.

vectors	Bold lower case is reserved for deterministic vectors and bold upper case is reserved for random vectors. Relations are interpreted componentwise. See Section A.1.
$\mathbb{E}$	A nice subset of compactified finite dimensional Euclidean space.
$M_+(\mathbb{E})$	The class of Radon measures on Borel subsets of $\mathbb{E}$.
$U^{\leftarrow}$	The left continuous inverse of the nondecreasing function U.
RV_ρ	The class of regularly varying functions with index ρ defined in (64).
Π	The function class Π reviewed in Section A.2 along with subclasses $\Pi_+(a(\cdot))$ and $\Pi_-(a(\cdot))$ and auxiliary function $a(\cdot)$.
Γ	The function class Γ reviewed in Section A.2 along with $\Gamma(f)$ and auxiliary function f.
$\xrightarrow{v}$	Vague convergence of measures; see Section A.3.
G_γ	An extreme value distribution given by (2) in the Von Mises parameterization.
$\mathbb{E}_\gamma$	$\{x : 1 + \gamma x > 0\}$.
$\overline{\mathbb{E}}_\gamma$	The closure on the right of the interval $\mathbb{E}_\gamma$.
$D(G_\gamma)$	The domain of attraction of the extreme value distribution G_γ. This is the set of F's satisfying (1). Note for $\gamma > 0$, $F \in D(G_\gamma)$ is equivalent to $1 - F \in RV_{1/\gamma}$.

2. Basic results.

In this section we give some implications of (5) and the assumptions (1), (2) given in Section 1.

2.1. *Standardization of Y.* Without loss of generality, we may assume Y is heavy tailed and $F \in D(G_1)$. The usual standardization procedure in extreme value theory (e.g., [34], Chapter 5, [17], Chapter 6.1.2, [32], Section 6.5.6) means that (1) implies for $x > 0$, as $t \to \infty$,

$$\begin{aligned} tP\left[\frac{b^{\leftarrow}(Y)}{t} > x\right] &= tP\left[\frac{Y - b(t)}{a(t)} > \frac{b(tx) - b(t)}{a(t)}\right] \\ &\to \left(1 + \gamma\frac{(x^{\gamma} - 1)}{\gamma}\right)^{-1/\gamma} = x^{-1}. \end{aligned}$$

Note if the distribution F of Y is continuous, $b^{\leftarrow}(Y)$ has a Pareto distribution and, in any case, $b^{\leftarrow}(Y)$ will always have a distribution tail which is asymptotically Pareto. For $y > 0$, (5) and (4) imply

$$\begin{aligned} & tP\left[\frac{X - \beta(t)}{\alpha(t)} \le x, \frac{b^{\leftarrow}(Y)}{t} > y\right] \\ & \quad = tP\left[\frac{X - \beta(t)}{\alpha(t)} \le x, \frac{Y - b(t)}{a(t)} > \frac{b(ty) - b(t)}{a(t)}\right] \\ & \quad \to \begin{cases} \mu\left([-\infty, x] \times \left(\dfrac{y^{\gamma} - 1}{\gamma}, \infty\right]\right), & \text{if } \gamma \ne 0, \\ \mu\big([-\infty, x] \times (\log y, \infty]\big), & \text{if } \gamma = 0. \end{cases} \end{aligned} \tag{7}$$

So at the expense of replacing Y by $b^{\leftarrow}(Y)$, theoretical development proceeds without loss of generality by replacing the conditions around (5) with

$$\begin{cases} \mu\big([-\infty, x] \times (y, \infty]\big) \text{ is not a degenerate distribution function in } x, \\ \qquad \text{for each } y > 0, \\ P[Y \le t] \in D(G_1), \qquad \lim_{t\to\infty} tP[Y > t] = 1, \\ tP\left[\dfrac{X - \beta(t)}{\alpha(t)} \le x, \dfrac{Y}{t} > y\right] \to \mu\big([-\infty, x] \times (y, \infty]\big), \\ \qquad x \in \mathbb{R}, y > 0, \text{ at continuity points } (x, y) \text{ of the limit.} \end{cases} \tag{8}$$

We refer to (8) as the basic convergence with the Y-variable *standardized*.

REMARK 1. The argument leading to (8) shows that we are free to change the marginal distribution of the Y-variable without disturbing the conditional convergence (6). We will see in Section 6, that this is not always possible for the X-variable.

We reiterate the connection with conditional modeling when (8) is assumed. For x which are continuity points of $H(x) := \mu([-\infty, x] \times (1, \infty])$,

$$\begin{aligned} H_t\big(\alpha(t)x + \beta(t)\big) &:= P\left[\frac{X - \beta(t)}{\alpha(t)} \le x \,\middle|\, Y > t\right] \\ &= \frac{P[(X - \beta(t))/\alpha(t) \le x, Y > t]}{P[Y > t]} \end{aligned} \tag{9}$$

$$\begin{aligned}&\sim tP\left[\frac{X-\beta(t)}{\alpha(t)}\le x, \frac{Y}{t}>1\right]\\&\to \mu\big([-\infty,x]\times(1,\infty]\big) =: H(x).\end{aligned}$$

Interpreting (8) as vague convergence on $M_+([-\infty,\infty]\times(0,\infty])$, we obtain from marginal convergence that

$$H(\infty)=\mu\big([-\infty,\infty]\times(1,\infty]\big)=1.$$

2.2. *Properties of the functions* $\alpha(\cdot)$ *and* $\beta(\cdot)$. The following is an initial attempt to understand the properties of the functions $\alpha(\cdot)$ and $\beta(\cdot)$.

PROPOSITION 1. *Suppose* (X,Y) *satisfy the standard form condition* (8). *Then there exist two functions* $\psi_1(\cdot)$, $\psi_2(\cdot)$, *such that for all* $c>0$,

$$\lim_{t\to\infty}\frac{\alpha(tc)}{\alpha(t)}=\psi_1(c) \tag{10}$$

and

$$\lim_{t\to\infty}\frac{\beta(tc)-\beta(t)}{\alpha(t)}=\psi_2(c). \tag{11}$$

The convergence in (10) *and* (11) *is uniform on compact subsets of* $(0,\infty)$.

PROOF. Pick $c>0$. For all but an at most countable set Λ of x-values, $(x,1)$ and (x,c^{-1}) are continuity points of μ. For $x\in\Lambda^c$, on the one hand we have (9) and on the other we have

$$\begin{aligned}&\lim_{t\to\infty}P\left[\frac{X-\beta(tc)}{\alpha(tc)}\le x\,\middle|\,\frac{Y}{t}>1\right]\\&\quad=\lim_{t\to\infty}tP\left[\frac{X-\beta(tc)}{\alpha(tc)}\le x,\frac{Y}{t}>1\right]\\&\quad=\lim_{t\to\infty}\frac{tc}{c}P\left[\frac{X-\beta(tc)}{\alpha(tc)}\le x,\frac{Y}{tc}>c^{-1}\right]\\&\quad=\frac{\mu([-\infty,x]\times(c^{-1},\infty])}{c}=:H^{(c)}(x).\end{aligned} \tag{12}$$

Thus the family $\{H_t\}$ converges with two different normalizations:

$$H_t\big(\alpha(t)x+\beta(t)\big)\to H(x),\qquad H_t\big(\alpha(tc)x+\beta(tc)\big)\to H^{(c)}(x).$$

The convergence to types theorem (see, e.g., [10] or [35], page 275) implies that (10) and (11) hold and also

$$H^{(c)}(x)=H\big(\psi_1(c)x+\psi_2(c)\big). \tag{13}$$

To prove local uniform convergence in (10) and (11), replace $c > 0$ in the argument with $c(t)$ where $c(t) \to c \in (0, \infty)$. Then (10) and (11) still hold and since ψ_1, ψ_2 are continuous (see next paragraph), the result follows from *continuous convergence*. See [34], page 2, or [23]. $\square$

From (10), we have that $\alpha(\cdot)$ is regularly varying with some index $\rho \in \mathbb{R}$, written $\alpha \in RV_\rho$, so that $\psi_1(x) = x^\rho$. (See [34], page 14, [4, 10–12, 38].) The function $\psi_2(x)$ may be identically zero. However, if it is not, then from [11], page 16, we have

$$\psi_2(x) = \begin{cases} k(x^\rho - 1)/\rho, & \text{if } \rho \neq 0, x > 0, \\ k \log x, & \text{if } \rho = 0, x > 0, \end{cases} \tag{14}$$

for $k \neq 0$. Also, there is more detailed information:

(i) If $\rho > 0$, then $\beta(\cdot) \in RV_\rho$ and $\beta(t) \sim \frac{1}{\rho}\alpha(t)$. So it is enough to scale X in (8) with a consequent location change in the x-variable for μ.

(ii) If $\rho = 0$, then $\beta(\cdot) \in \Pi(\alpha)$ and $\alpha \in RV_0$. So α is the auxiliary function of the Π-function β.

(iii) If $\rho < 0$, then $\beta(\infty) = \lim_{t\to\infty} \beta(t)$ exists finite and

$$\beta(\infty) - \beta(t) \in RV_\rho; \qquad \bigl(\beta(\infty) - \beta(t)\bigr) \sim \frac{1}{|\rho|}\alpha(t).$$

Case (iii) can be reduced to case (i) by a change of variable. From case (iii) of (8) we get

$$tP\left[\frac{X - \beta(\infty) + [\beta(\infty) - \beta(t)]}{|\rho|(\beta(\infty) - \beta(t))} \le x, \frac{Y}{t} > y\right] \to \mu\bigl([-\infty, x] \times (y, \infty]\bigr).$$

Write

$$\widetilde{X} := \frac{1}{X - \beta(\infty)}, \qquad \tilde{\beta}(t) := \frac{1}{|\rho|(\beta(\infty) - \beta(t))}, \tag{15}$$

so that

$$\begin{aligned} &tP\left[\frac{\widetilde{X}}{\tilde{\beta}(t)} \le x, \frac{Y}{t} > y\right] \\ &\quad = tP\left[\frac{X - \beta(\infty)}{|\rho|(\beta(\infty) - \beta(t))} \ge \frac{1}{x}, \frac{Y}{t} > y\right] \\ &\quad = tP\left[\frac{X - \beta(\infty)}{|\rho|(\beta(\infty) - \beta(t))} + \frac{1}{|\rho|} \ge \frac{1}{x} + \frac{1}{|\rho|}, \frac{Y}{t} > y\right] \\ &\quad \to \mu\left(\left[\frac{1}{x} + \frac{1}{|\rho|}, \infty\right] \times (y, \infty]\right) =: \tilde{\mu}\bigl([-\infty, x] \times (y, \infty]\bigr). \end{aligned} \tag{16}$$

Since case (iii) can be reduced to case (i), it does not need separate theoretical attention.

2.3. *Conditions for the limit μ to be a product measure.* It turns out that μ being a product measure is equivalent to $\psi_1 \equiv 1$ and $\psi_2 \equiv 0$.

PROPOSITION 2. *We have $\mu = H \times \nu_1$, where $\nu_1((y,\infty]) = y^{-1}$, $y > 0$ (i.e., $\mu([-\infty, x] \times (y,\infty]) = H(x)y^{-1}$), iff for all $c > 0$,*

$$\psi_1(c) = \lim_{t\to\infty} \frac{\alpha(tc)}{\alpha(t)} = 1, \qquad \psi_2(c) = \lim_{t\to\infty} \frac{\beta(tc)-\beta(t)}{\alpha(t)} = 0. \tag{17}$$

PROOF. Given that μ is a product, we have from (9) and (12), that $H^{(c)}(x) = H(x)$. Hence (17) follows from the convergence to types theorem. Conversely, if (17) holds, $H^{(c)}(x) = H(x)$ and from (12) we have, for all $c > 0$, $\mu([-\infty, x] \times (c^{-1}, \infty]) = cH(x)$. So for all $y > 0$, $\mu([-\infty, x] \times (y,\infty]) = H(x)y^{-1}$. □

REMARK 2. What if $\psi_2 \equiv 0$ but $\psi_1 \not\equiv 1$? Then $\alpha \in RV_\rho$ for some $\rho \in \mathbb{R}$, $\rho \neq 0$ and $\psi_1(c) = c^\rho$, for $c > 0$. The reasoning in the previous proof shows that μ has the form

$$\mu\big([-\infty, x] \times (y,\infty]\big) = y^{-1} H(x/y^\rho), \tag{18}$$

for $x \in \mathbb{R}$, and $y > 0$ and where H is a proper nondegenerate probability distribution.

2.4. *When the X-variable can be standardized.* Standardization is the process of transforming variables so that their distributions have regularly varying tails *in standard form*. See [34], Chapter 5, [17], Chapter 6.1.2, [32], Section 6.5.6. Once standard form regular variation is achieved, limit measures have a scaling property and characterization of these limits becomes possible. We know we can standardize the Y variable. What about the X variable?

It is possible to standardize the X-variable if $\beta(t) \geq 0$ and $\psi_2(\cdot)$ in (11) is not constant and $\beta^{\leftarrow}$ is nondecreasing on the range of X since in this case we have for $x > 0$,

$$\begin{aligned} tP\left[\frac{\beta^{\leftarrow}(X)}{t} \leq x, \frac{Y}{t} > y\right] &= tP\left[\frac{X-\beta(t)}{\alpha(t)} \leq \frac{\beta(tx)-\beta(t)}{\alpha(t)}, \frac{Y}{t} > y\right] \\ &\to \mu\big([-\infty, \psi_2(x)] \times (y,\infty]\big), \end{aligned} \tag{19}$$

at continuity points of the limit. We emphasize there are important cases where $\psi_2(x)$ is identically zero and thefore where X cannot be standardized by the procedure in (19); see Section 6.1.

Standardization is also possible if $\psi_2 \equiv 0$, provided $X > 0$ and $\psi_1 \not\equiv 1$; that is if $\alpha(\cdot) \in RV_\rho$ with $\rho \neq 0$. If $\rho > 0$, then [4], Theorem 3.1.12a, c, page 136, gives $\beta(t)/\alpha(t) \to 0$ and by the convergence to types theorem (8) can be rewritten as

$$tP\left[\frac{X}{\alpha(t)} \leq x, \frac{Y}{t} > y\right] \to \mu\big([0, x] \times (y,\infty]\big), \qquad x > 0, y > 0.$$

Therefore, supposing without loss of generality that $\alpha(\cdot)$ is strictly increasing and continuous (e.g., [38]), we have

$$
\begin{aligned}
tP\left[\frac{\alpha^{\leftarrow}(X)}{t} \le x, \frac{Y}{t} > y\right] &= tP\left[\frac{X}{\alpha(t)} \le \frac{\alpha(tx)}{\alpha(t)}, \frac{Y}{t} > y\right] \\
&\to \mu\big((0, x^{\rho}] \times (y, \infty]\big)
\end{aligned}
$$

and $(\alpha^{\leftarrow}(X), Y)$ are the standardized variables. If $\rho < 0$, [4], Theorem 3.1.10a, c, page 134, implies $\beta(\infty) := \lim_{t\to\infty} \beta(t)$ exists finite and $(\beta(\infty) - \beta(t))/\alpha(t) \to 0$. Therefore, if we suppose $P[X \le \beta(\infty)] = 1$, we have for $x > 0$,

$$
\begin{aligned}
&\lim_{t\to\infty} tP\left[\frac{1/(\beta(\infty) - X)}{1/\alpha(t)} \le x, \frac{Y}{t} > y\right] \\
&\quad = \lim_{t\to\infty} tP\left[\frac{\beta(\infty) - X}{\alpha(t)} \ge x^{-1}, \frac{Y}{t} > y\right] \\
&\quad = \lim_{t\to\infty} tP\left[\frac{\beta(\infty) - X - (\beta(\infty) - \beta(t))}{\alpha(t)} \ge x^{-1}, \frac{Y}{t} > y\right] \\
&\quad = \lim_{t\to\infty} tP\left[\frac{X - \beta(t)}{\alpha(t)} \le -x^{-1}, \frac{Y}{t} > y\right] \\
&\quad = \mu\big([-\infty, -x^{-1}] \times (y, \infty]\big),
\end{aligned}
$$

and the variables $((\beta(\infty) - X)^{-1}, Y)$ can be standardized according to the recipe for the $\rho > 0$ case.

2.4.1. *When $\beta(t)$ is monotone.* The standardization of the X variable in (19) begs the question of when β is monotone. Consider the case where $\psi_2 \not\equiv 0$ and ψ_2 is given by (14) and indexed by $\rho \in \mathbb{R}$. For discussing when $\beta(t)$ is monotone, it is important to remember that $\beta(\cdot)$ is only determined up to the asymptotic equivalence given by the convergence to types theorem.

Consider the following cases.

1. $\rho > 0$: For this case, we have $\beta \in RV_\rho$ and there exists $\tilde{\beta}(t) \in RV_\rho$ such that $\tilde{\beta}(\cdot)$ is continuous, strictly increasing to ∞ with $\beta \sim \tilde{\beta}$. (See, e.g., [38].) So without loss of generality, for the case $\rho > 0$, we may assume $\beta(\cdot)$ is continuous and strictly increasing.
2. $\rho < 0$: The transformation described in (15) and (16), show that the pair (X, Y) can be transformed to $(\widetilde{X}, Y)$ satisfying $\rho > 0$.
3. $\rho = 0$: Suppose $\beta(\cdot) \in \Pi_+(a)$ after which we consider $\beta \in \Pi_-(a)$. From [18] as reviewed in Section A.2, there exists $\tilde{\beta}(t)$ which is continuous, strictly increasing and such that $\beta - \tilde{\beta} = o(\alpha)$ so that the convergence of types theorem allows us to replace β by $\tilde{\beta}$. Assume this is done which is tantamount to dropping the tilde. Then there are two cases to consider.

(a) $\beta(\infty) = \infty$.
(b) $\beta(\infty) < \infty$.

For 3(a) it is clear that $\beta(t)$ has the desired properties of being continuous and strictly increasing to ∞. For 3(b), proceed as follows to transform (X, Y): Define

$$\begin{aligned} \widetilde{X} &= \frac{1}{\beta(\infty) - X}, \qquad \tilde{\beta}(t) = \frac{1}{\beta(\infty) - \beta(t)}, \\ \tilde{\alpha}(t) &= \frac{\alpha(t)}{(\beta(\infty) - \beta(t))^2}. \end{aligned} \tag{20}$$

Then $\tilde{\beta}(t) \uparrow \infty$ is continuous and strictly monotone and $\tilde{\beta} \in \Pi_+(\tilde{\alpha})$ and after some calculation we get

$$\begin{aligned} &tP\left[\frac{\widetilde{X} - \tilde{\beta}(t)}{\tilde{\alpha}(t)} \le x, \frac{Y}{t} > y\right] \\ &\quad = tP\left[\frac{X - \beta(t)}{\alpha(t)} \le \frac{x}{1 + \alpha(t)x/(\beta(\infty) - \beta(t))}, \frac{Y}{t} > y\right] \\ &\quad \to \mu\big([-\infty, x] \times (y, \infty]\big) \end{aligned}$$

since $\tilde{\beta} \in \Pi_+(\tilde{\alpha})$ implies $\tilde{\beta}(t)/\tilde{\alpha}(t) \to \infty$ which is identical to $(\beta(\infty) - \beta(t))/\alpha(t) \to \infty$. Thus after the transformation of (X, Y) to $(\widetilde{X}, Y)$, case 3(b) is reduced to case 3(a).

What if $\beta \in \Pi_-(a)$? Then define

$$\widetilde{X} = -X, \qquad \tilde{\beta}(t) = -\beta(t), \qquad \tilde{\alpha}(t) = \alpha(t),$$

and $\tilde{\beta} \in \Pi_+(a)$ and this case reduces to the case when $\beta \in \Pi_+(a)$ since

$$\begin{aligned} tP\left[\frac{\widetilde{X} - \tilde{\beta}(t)}{\tilde{\alpha}(t)} \le x, \frac{Y}{t} > y\right] &= tP\left[\frac{X - \beta(t)}{\alpha(t)} \ge -x, \frac{Y}{t} > y\right] \\ &\to \mu\big([-x, \infty] \times (y, \infty]\big). \end{aligned}$$

2.4.2. *Summary.* When $\psi_2 \not\equiv 0$, if we make the transformation $X \mapsto \widetilde{X}$ and consider the analogue of (8) for $(\widetilde{X}, Y)$, we can standardize the $\widetilde{X}$-variable. If $\psi_2 \equiv 0$, but $\psi_1(c) = c^\rho$, for $c > 0$, $\rho \neq 0$, then for $\rho > 0$, $(\alpha^{\leftarrow}(X), Y)$ are a standardized pair and for $\rho < 0$, $((1/\alpha)^{\leftarrow}(\widetilde{X}), Y)$ is a standardized pair.

When the limit μ is a product measure, $(\psi_1, \psi_2) \equiv (1, 0)$ and standardization is not possible; an example is given in Section 6.1.3 and a proof of the assertion is easy using the change of coordinate system techniques of Section 7.

2.5. *Densities.* In this section we see what form the basic convergence takes when (X, Y) has a density. Since it is sufficient to suppose that the Y-variable has been transformed to the standard case, for this section, we assume the following:

1. The pair (X, Y) has density $f(x, y)$.
2. The marginal density $f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,dx$ of the Y-variable satisfies

$$f_Y(y) = y^{-2}, \qquad y > 1.$$

Since we have densities, we assume the transformation to Y being standard renders Y a Pareto random variable with unit shape parameter.

3. The joint density $f(x, y)$ satisfies

$$t^2\alpha(t) f\big(\alpha(t)x + \beta(t), ty\big) \to g(x, y) \in L_1\big([-\infty, \infty] \times (0, \infty]\big), \tag{21}$$

where the limit $g(x, y) \geq 0$ is integrable, not identically zero and satisfies for each fixed $v > 0$,

$$v^2 g(u, v) \qquad \text{is a probability density in } u. \tag{22}$$

PROPOSITION 3. *With the assumptions just listed*, (8) *holds with*

$$\mu\big([-\infty, x] \times (y, \infty]\big) = \int_{u \leq x} \int_{v > y} g(u, v)\,dv\,du,$$

and $H(\infty) = \mu([-\infty, \infty] \times (1, \infty]) = 1$.

PROOF. We use standard notation for conditional densities. So for instance, $f_{X|Y=v}(u|v)$ is the conditional density of X given $Y = v$.

We need two facts:

1. First we evaluate the integrand. For $v > 0$, (21) implies

$$f_{(X-\beta(t))/\alpha(t)|Y/t=v}(u|v) \to v^2 g(u, v) \qquad (t \to \infty). \tag{23}$$

To see this, observe

$$\begin{aligned} f_{(X-\beta(t))/\alpha(t)|Y/t=v}(u|v) &= \frac{f_{(X-\beta(t))/\alpha(t),Y/t}(u, v)}{f_{Y/t}(v)} = \frac{t\alpha(t) f(\alpha(t)u + \beta(t), tv)}{t f_Y(tv)} \\ &= t^2\alpha(t) v^2 f\big(\alpha(t)u + \beta(t), tv\big) \to v^2 g(u, v). \end{aligned}$$

2. We now show convergence of the integral. The function of u

$$f_{(X-\beta(t))/\alpha(t)|Y/t=v}(u|v)$$

is a probability density for fixed v.

Now write

$$\begin{aligned} &tP\left[\frac{X - \beta(t)}{\alpha(t)} \leq x, \frac{Y}{t} > y\right] \\ &\quad = t\int_{[v>y]} \left[\int_{[u \leq x]} f_{(X-\beta(t))/\alpha(t)|Y/t=v}(u|v)\,du\right] f_{Y/t}(v)\,dv \\ &\quad = \int_{[v>y]} \left[\int_{[u \leq x]} f_{(X-\beta(t))/\alpha(t)|Y/t=v}(u|v)\,du\right] v^{-2}\,dv. \end{aligned}$$

The integral inside the square bracket has an integrand which is a family of probability densities in the variable u (with v fixed) indexed by t which converges to a limiting probability density $v^2 g(u,v)$. Hence by Scheffé's lemma (e.g., [35], page 253)

$$\left[\int_{[u\le x]} f_{(X-\beta(t))/\alpha(t)|Y/t=v}(u|v)\,du\right] \to \int_{[u\le x]} v^2 g(u,v)\,du.$$

Now the square bracket term is a conditional probability and hence is a function of v bounded almost surely by 1. So by dominated convergence, we have proven (8) as required.

To check the last assertion that $H(\infty)=1$, note

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{v>1} g(u,v)\,du\,dv &= \int_{v>1} v^{-2}\left(\int_{-\infty}^{\infty} v^2 g(u,v)\,du\right)dv \\ &= \int_{v>1} v^{-2} dv = 1.\end{aligned}$$

□

Heffernan and Tawn [21] assume that (X,Y) have been transformed to have Gumbel marginal distributions, that is, $P(X\le t)=P(Y\le t)=\exp(-\exp(-t))$ for $t\in\mathbb{R}$ and that for such (X,Y)

$$tP\left[\frac{X-\tilde{\beta}(t)}{\tilde{\alpha}(t)}\le x \,\middle|\, Y=t\right] \tag{24}$$

converges to a nondegenerate limit distribution as $t\to\infty$, for some scaling function $\tilde{\alpha}(\cdot)>0$ and centering function $\tilde{\beta}(\cdot)\in\mathbb{R}$.

Thus we see that since (23) implies [21] condition (24), (21) implies (24). This makes explicit the link between our assumptions (5) and those of Heffernan and Tawn [21] under the above conditions for densities. We have

$$\begin{aligned}P\left[\frac{X-\tilde{\beta}(t)}{\tilde{\alpha}(t)}\le x \,\middle|\, Y=ty\right] &= \int_{u\le x} f_{(X-\tilde{\beta}(t))/\tilde{\alpha}(t)|Y/t=y}(u|y)\,du \\ &\to \int_{u\le x} y^2 g(u,y)\,du,\end{aligned}$$

and letting $y=1$ gives

$$P\left[\frac{X-\tilde{\beta}(t)}{\tilde{\alpha}(t)}\le x \,\middle|\, Y=t\right] \to \int_{u\le x} g(u,1)\,du.$$

3. Characterizing the class of limit measures. Assuming the Y-variable is standardized, what is the class of limits in (8)? We divide this issue in two parts, depending on whether the limit measure μ is a product or not.

3.1. *The limit measure is a product.* For this case, there is not much discussion required since for any distribution function $H(x)$ on $\mathbb{R}$, the limit

$$\mu = H \times \nu_1 \quad \text{or} \quad \mu([-\infty, x] \times (y, \infty]) = H(x)y^{-1}$$

is possible. To achieve this limit, suppose X, Y are independent random variables with X having distribution H and Y being standard Pareto. Then with $\beta(t) = 0$ and $\alpha(t) = 1$, (8) is satisfied.

3.2. *The limit measure is not a product.* When μ is not a product, we change coordinate systems and transform X to some X^* and assume (X^*, Y) is a standard pair and

$$tP\left[\left(\frac{X^*}{t}, \frac{Y}{t}\right) \in \cdot\right] \xrightarrow{v} \mu_*(\cdot) \qquad \text{in } M_+([0, \infty] \times (0, \infty]), \tag{25}$$

where μ_* is a transformation of μ as described in Section 2.4.

From (25), we see that the distribution of (X^*, Y) is standard regularly varying with limit measure μ_* (see [3, 32, 37]) on the cone $[0, \infty] \times (0, \infty]$ and, therefore μ_* is homogeneous of order -1:

$$\mu_*(c\Lambda) = c^{-1}\mu_*(\Lambda) \qquad \forall c > 0,$$

where Λ is a Borel subset of $[0, \infty] \times (0, \infty]$. This means μ_* has a spectral form. We pick a norm. Any norm would do but for convenience define

$$\|(x, y)\| = |x| + |y|, \qquad (x, y) \in \mathbb{R}^2.$$

Of course, when restricting attention to $[0, \infty] \times (0, \infty]$, the absolute value bars can be dropped. Then the standard argument using homogeneity ([34], Chapter 5), yields for $r > 0$ and Λ a Borel subset of $[0, 1)$,

$$\begin{aligned}
&\mu_*\left\{(x, y) \in [0, \infty] \times (0, \infty] : x + y > r, \frac{x}{x + y} \in \Lambda\right\} \\
&\quad = \mu_*\left\{r(x, y) \in [0, \infty] \times (0, \infty] : x + y > 1, \frac{x}{x + y} \in \Lambda\right\} \\
&\quad = r^{-1}\mu_*\left\{(x, y) \in [0, \infty] \times (0, \infty] : x + y > 1, \frac{x}{x + y} \in \Lambda\right\} \\
&\quad =: r^{-1}S(\Lambda).
\end{aligned} \tag{26}$$

The Radon measure S need not be a finite measure on $[0, 1)$ but to guarantee that

$$H_*(x) = \mu_*([0, x] \times (1, \infty]) \tag{27}$$

is a probability measure, we need

$$\int_0^1 (1 - w)S(dw) = 1. \tag{28}$$

This will be clear from the following calculation to get the canonical form of $H_*(x)$ for $x > 0$:

Using (26), write for $x > 0$,

$$
\begin{aligned}
&\mu_*([0,x]\times(y,\infty])\\
&\quad=\iint_{\substack{0\le rw\le x\\ r(1-w)>y\\ 0\le w<1}} r^{-2}\,dr\,S(dw)\\
&\quad=\int_{r=0}^{\infty}\biggl(\int_{\substack{0\le w\le x/r\\ 1-y/r>w\\ 0\le w<1}} S(dw)\biggr)r^{-2}\,dr \qquad (29)\\
&\quad=\int_0^{\infty} S\biggl(\biggl[0,\frac{x}{r}\wedge\Bigl(1-\frac{y}{r}\Bigr)\wedge 1\biggr)\biggr)r^{-2}\,dr\\
&\quad=\int_0^{\infty} S\bigl([0,xv\wedge(1-yv)\wedge 1)\bigr)\,dv.
\end{aligned}
$$

Integrating the double integral in reverse order yields the alternate expression

$$
\begin{aligned}
&\mu_*([0,x]\times(y,\infty])\\
&\quad=\int_{w\in[0,1)}\biggl(\int_{y/(1-w)<r\le x/w} r^{-2}\,dr\biggr)S(dw)\\
&\quad=\int_{w\in[0,1)}\bigl((1-w)y^{-1}-wx^{-1}\bigr)_+S(dw) \qquad (30)\\
&\quad=y^{-1}\int_0^{x/(x+y)}(1-w)S(dw)-x^{-1}\int_0^{x/(x+y)} wS(dw).
\end{aligned}
$$

Conclusion: The class of limits μ_* or conditional limits

$$H_*(x)=\lim_{t\to\infty}P\biggl[\frac{X^*}{t}\le x\Big|Y>t\biggr]$$

is indexed by Radon measures S on $[0,1)$ satisfying the integrability condition (28).

EXAMPLE. As an example, suppose S is uniform on $[0,1)$: $S(dw)=\frac{dw}{c}$, where c is chosen so that (28) is satisfied: $\int_0^1\frac{w}{c}\,dw=1$ which implies $c=1/2$. This yields

$$\mu_*([0,x]\times(y,\infty])=\frac{x}{x+y}\biggl[\frac{2}{y}-\biggl(\frac{1+x/y}{x+y}\biggr)\biggr]$$

and setting $y=1$ we get a Pareto distribution

$$H_*(x)=\frac{x}{1+x}=1-\frac{1}{1+x},\qquad x>0.$$

4. Random norming. In [21], it was necessary to normalize X by a function of the precise value of Y occurring with X to achieve nondegeneracy of the limiting conditional distribution. Motivated by this, we consider how to normalize the X-variable with a function of Y rather than a deterministic affine transformation, using functions of the threshold t in (6). This leads to a product form limit in all cases.

It is significant that normalizing by using functions of the threshold t in (6) does not result in a product limit in all cases, but that the inclusion of the precise value of Y occurring with X adds enough detail to the normalization to allow the limit always to factorize. In statistical applications the factorization of the limit distribution will constitute a welcome simplification of models based on this limiting form. Indeed, the statistical model of Heffernan and Tawn [21]relies on such factorization to ensure that the residuals formed by normalizing observed values of X by functions of the observed values of Y are independent of the Y values.

We discuss this random normalization in two stages:

- The X-variable can be standardized and the limit in (8) is not a product.
- The limit measure μ in (8) is a product measure.

4.1. *The X-variable can be standardized and the limit measure μ is not a product.* We suppose X can be transformed to X^* so that (X^*, Y) is a standardized pair and (25) holds with limit measure μ_*. As in Section 3.2, let S be the spectral measure of μ_*. Then we have the following result which forms the basis of the estimation procedure proposed in [21].

PROPOSITION 4. *If (25) holds, then*

$$tP\left[\left(\frac{X^*}{Y}, \frac{Y}{t}\right) \in \cdot\right] \xrightarrow{v} G \times \nu_1 \qquad \text{in } M_+\big([0,\infty] \times (0,\infty]\big), \tag{31}$$

where for $x > 0$

$$\nu_1((x,\infty]) = x^{-1} \quad \text{and} \quad G(x) = \int_0^{x/(1+x)} (1-w)S(dw). \tag{32}$$

This means

$$P\left[\frac{X^*}{Y} \le x \middle| Y > t\right] \to G(x), \qquad x > 0.$$

Conversely, if (31) holds, then so does (25).

PROOF. This proof is discussed in Theorem 2.1 of [28]. The outline of the argument is as follows. Applying the map $T_1(x,y) = (\frac{x}{y}, y)$ to (25) yields after a compactification argument that

$$tP\left[\left(\frac{X^*}{Y}, \frac{Y}{t}\right) \in \cdot\right] \xrightarrow{v} \mu_* \circ T_1^{-1}.$$

So the limit evaluated on $[0, x] \times (y, \infty]$ is

$$\begin{aligned}\mu_*&\left\{(u, v): \frac{u}{v} \le x, v > y\right\}\\&= y^{-1}\mu_*\left\{(u, v): \frac{u}{v} \le x, v > 1\right\}\\&= y^{-1}\iint_{\substack{rw/(r(1-w))\le x\\ r(1-w)>1}} r^{-2}\,dr\,S(dw)\\&= y^{-1}\int_{w\le x/(1+x)}\left(\int_{r>1/(1-w)} r^{-2}\,dr\right)S(dw)\\&= y^{-1}\int_0^{x/(1+x)}(1-w)S(dw).\end{aligned}$$

The converse proceeds similarly using the map $T_2(x, y) = (xy, y) = T_1^{-1}(x, y)$. □

4.2. *The limit measure μ is a product measure.* Now we suppose (8) holds with $\mu = H \times \nu_1$. In this case, from Proposition 2, (10) and (11) hold with $\psi_1(x) \equiv 1$, $\psi_2(x) \equiv 0$.

PROPOSITION 5. *If,*

$$tP\left[\frac{X-\beta(t)}{\alpha(t)} \le x, \frac{Y}{t} > y\right] \to H(x)y^{-1} \qquad (x \in \mathbb{R}, y > 0) \tag{33}$$

for a nondegenerate probability distribution function $H(x)$, then also

$$tP\left[\frac{X-\beta(Y)}{\alpha(Y)} \le x, \frac{Y}{t} > y\right] \to H(x)y^{-1} \qquad (x \in \mathbb{R}, y > 0) \tag{34}$$

and

$$P\left[\frac{X-\beta(Y)}{\alpha(Y)} \le x \,\middle|\, Y > t\right] \to H(x).$$

Conversely, if (34) holds and $\alpha(\cdot)$ and $\beta(\cdot)$ satisfy (10), (11) locally uniformly with $\psi_1(x) \equiv 1$, and $\psi_2(x) \equiv 0$, then (33) also holds.

PROOF. For any $K > y > 0$ we have

$$\begin{aligned}tP&\left[\frac{X-\beta(Y)}{\alpha(Y)} \le x, \frac{Y}{t} \in (y, K]\right]\\&= tP\left[\frac{X-\beta(t)}{\alpha(t)} \le \frac{\alpha(tY/t)}{\alpha(t)}x + \frac{\beta(tY/t)-\beta(t)}{\alpha(t)}, \frac{Y}{t} \in (y, K]\right]\end{aligned}$$

and because of local uniform convergence in (10) and (11), this converges to

$$\mu\big([-\infty, x] \times (y, K]\big) = H(x)(y^{-1} - K^{-1}).$$

Therefore

$$\begin{aligned}\liminf_{t\to\infty} tP\left[\frac{X-\beta(Y)}{\alpha(Y)} \le x, \frac{Y}{t} > y\right] &\ge \liminf_{t\to\infty} tP\left[\frac{X-\beta(Y)}{\alpha(Y)} \le x, \frac{Y}{t} \in (y, K]\right] \\ &= H(x)(y^{-1} - K^{-1}).\end{aligned}$$

Since this is true for all $K > y$, we have

$$\liminf_{t\to\infty} tP\left[\frac{X-\beta(Y)}{\alpha(Y)} \le x, \frac{Y}{t} > y\right] \ge H(x)y^{-1}.$$

Also,

$$\begin{aligned}\limsup_{t\to\infty} tP\left[\frac{X-\beta(Y)}{\alpha(Y)} \le x, \frac{Y}{t} > y\right] &\le \lim_{t\to\infty} tP\left[\frac{X-\beta(Y)}{\alpha(Y)} \le x, \frac{Y}{t} \in (y, K]\right] \\ &\quad + \limsup_{t\to\infty} tP\left[\frac{Y}{t} > K\right] \\ &= H(x)(y^{-1} - K^{-1}) + K^{-1}.\end{aligned}$$

Letting $K \to \infty$ provides the other half of the sandwich and (34) is proven.

For the converse, write

$$\begin{aligned}tP&\left[\frac{X-\beta(t)}{\alpha(t)} \le x, \frac{Y}{t} \in (y, K]\right] \\ &= tP\left[\frac{X-\beta(Y)}{\alpha(Y)} \le \frac{\alpha(t)}{\alpha(Y)}x + \frac{\beta(t)-\beta(Y)}{\alpha(Y)}, \frac{Y}{t} \in (y, K]\right].\end{aligned}$$

Proceed as before using uniform convergence. □

5. Connection to multivariate extreme value theory and asymptotic independence. We now make some comments on the relationship between our conditioned limit condition (8) and multivariate extreme value theory.

Suppose the distribution of (X, Y) is in the domain of attraction of a multivariate extreme value distribution. This means that for i.i.d. replicates $\{(X_i, Y_i), i \ge 1\}$ of (X, Y) there exist centering $b_j(t) \in \mathbb{R}$ and scaling $a_j(t) > 0$ functions, $j = 1, 2$, and

$$P\left[\frac{\bigvee_{i=1}^n X_i - b_1(n)}{a_1(n)} \le x, \frac{\bigvee_{i=1}^n Y_i - b_2(n)}{a_2(n)} \le y\right] \to G(x, y), \tag{35}$$

where G is a multivariate extreme value distribution. Let the marginal distributions of G be G_j, $j = 1, 2$. *Asymptotic independence* means $G(x, y) = G_1(x)G_2(y)$.

Define

$$U_1(x) = \frac{1}{P[X > x]}, \qquad U_2(y) = \frac{1}{P[Y > y]},$$
$$\chi_j(x) = \left(\frac{1}{-\log G_j}\right)^{\leftarrow}(x), \qquad x > 0, j = 1, 2,$$
$$G^*(x, y) = G(\chi_1(x), \chi_2(y)), \qquad x > 0, y > 0.$$

According to Resnick [34], Proposition 5.10, page 265, we can standardize the condition (35) by transforming $(X, Y) \mapsto (X^*, Y^*) = (U_1(X), U_2(Y))$ and then

$$P\left[\frac{\bigvee_{i=1}^n X_i^*}{n} \le x, \frac{\bigvee_{i=1}^n Y_i^*}{n} \le y\right] \to G^*(x, y), \tag{36}$$

and G^* is max-stable. From [34], Proposition 5.15, page 277 and [32], Section 6.1, this is equivalent to marginal convergence and multivariate regular variation of the distribution of (X^*, Y^*):

$$tP\left[\left(\frac{X^*}{t}, \frac{Y^*}{t}\right) \in \cdot\right] \xrightarrow{v} \nu^*(\cdot), \tag{37}$$

in $M_+([0, \infty]^2 \setminus \{\mathbf{0}\})$. Here ν^* is a Radon measure on $[0, \infty]^2 \setminus \{\mathbf{0}\}$ satisfying

$$\nu^*(t\cdot) = t^{-1}\nu^*(\cdot). \tag{38}$$

Asymptotic independence means

$$\begin{aligned}\nu^*([0, x] \times [0, y])^c &= -\log G^*(x, y) = -\log G^*(x, \infty) - \log G^*(\infty, y) \\ &= \nu^*\big((x, \infty] \times [0, \infty]\big) + \nu^*\big([0, \infty] \times (y, \infty]\big),\end{aligned}$$

and ν^* concentrates on the lines $\{(x, 0) : x > 0\} \cup \{(0, y) : y > 0\}$.

Suppose the domain of attraction condition (37) holds but asymptotic independence does *not* hold. Condition (37) implies for $x > 0$, $y > 0$,

$$tP\left[\frac{X^*}{t} \le x, \frac{Y^*}{t} > y\right] \to \nu^*\big([0, x] \times (y, \infty]\big)$$

and we claim for fixed $y > 0$, $\nu^*([0, x] \times (y, \infty])$ is not degenerate in x. This follows, for instance, from (38). Conclusion: the domain of attraction condition (37) in standard form without asymptotic independence implies that (X^*, Y^*) satisfy (8). Condition (8) is equivalent to vague convergence on the cone $[0, \infty] \times (0, \infty]$ while the regular variation condition (37) gives vague convergence on the bigger cone $[0, \infty]^2 \setminus \{\mathbf{0}\}$.

Suppose (37) holds *with* asymptotic independence. Consider (8) with X^*/t in place of $(X - \beta(t))/\alpha(t)$. The nondegeneracy condition in (8) fails because for fixed $y > 0$, $\mu([-\infty, x] \times (y, \infty]) = \nu^*([-\infty, x] \times (y, \infty])$ concentrates all mass at $x = 0$. If one wants (8) to hold, one must make an additional assumption beyond the domain of attraction condition (37) and the X^* variable in (37) must be normalized differently. For a simple particular case which is somewhat familiar, consider the following: Suppose we assume the condition (37) with asymptotic independence and in addition we assume that X^* can be normalized by $\alpha(t)$ instead of by t, so that (8) holds in the form

$$tP\left[\frac{X^*}{\alpha(t)} \le x, \frac{Y^*}{t} > y\right] \to \mu([0,x] \times (y,\infty]), \qquad x > 0, y > 0. \tag{39}$$

From (39) and (37), we have for $0 < a < b \le \infty$ and $y > 0$

$$tP\left[\frac{X^*}{\alpha(t)} \in (a,b], \frac{Y^*}{t} > y\right] \to \mu((a,b] \times (y,\infty]),$$
$$tP\left[\frac{X^*}{t} \in (a,b], \frac{Y^*}{t} > y\right] \to 0.$$

We claim that $t/\alpha(t) \to \infty$ so that $\alpha(\cdot)$ is of smaller order than t. If not, there exist $t_n \to \infty$ and $0 \le c < \infty$ and $t_n/\alpha(t_n) \to c$. From the nondegeneracy condition in (8), we may pick $0 < a < b$ such that $\mu((a,b] \times (1,\infty]) > 0$. Then

$$0 < \mu((a,b] \times (1,\infty]) = \lim_{n\to\infty} t_n P\left[\frac{X^*}{t_n} \in \left(\frac{\alpha(t_n)}{t_n}a, \frac{\alpha(t_n)}{t_n}b\right], \frac{Y^*}{t_n} > 1\right] = 0$$

giving a contradiction. So $\alpha(\cdot)$ is of smaller order than t and we have the situation of *hidden regular variation* [20, 27, 36]; that is, the regular variation condition (37) holds on the big cone $[0,\infty]^2 \setminus \{\mathbf{0}\}$ but a different regular variation condition holds on the smaller cone $[0,\infty] \times (0,\infty]$.

To summarize: The multivariate extreme value paradigm without asymptotic independence subsumes our conditioned limit condition (5). However, in the presence of asymptotic independence, the multivariate extreme value condition is refined by (5) which uses a more delicate normalization to track mass into the part of the distributional tail where the conditioning variable Y is large.

6. Examples.

We give examples to illustrate some intricacies.

6.1. *Bivariate normal.* Suppose N_1, N_2 are i.i.d. $N(0,1)$ random variables and $|\rho| \le 1$. Define $(X,Y) = (\sqrt{1-\rho^2}N_1 + \rho N_2, N_2)$ which is a bivariate normal vector with means 0, variances 1 and correlation ρ. Denote the standard normal distribution function by $N(x)$. Recall (e.g., from [34], page 71) that we may set

$$\begin{aligned} a(t) &= \frac{1}{\sqrt{2\log t}}, \\ b(t) &= \left(\frac{1}{1-N}\right)^{\leftarrow}(t) = \sqrt{2\log t} - \frac{(1/2)(\log\log t + \log 4\pi)}{\sqrt{2\log t}} + o(a(t)), \end{aligned} \tag{40}$$

and then for $x \in \mathbb{R}$,

$$\lim_{t\to\infty} tP\left[\frac{N_1 - b(t)}{a(t)} > x\right] = e^{-x}.$$

6.1.1. *Conditional limits for* (X, Y). We begin by discussing the following result learned from [1]. Suppose $N(x)$ is the standard normal distribution function and $n(y)$ is its density. Then

$$tP\left[X - \rho b(t) \le x, \frac{Y - b(t)}{a(t)} > y\right] \to N\big(x/\sqrt{1-\rho^2}\big)e^{-y}, \tag{41}$$

or standardizing the Y-variable,

$$tP\left[X - \rho b(t) \le x, \frac{b^{\leftarrow}(Y)}{t} > y\right] \to N\big(x/\sqrt{1-\rho^2}\big)y^{-1}. \tag{42}$$

Here we claimed $\beta(t) = \rho b(t)$ and $\alpha(t) = 1$. It is well known (e.g., [34], page 71) that $b(\cdot) \in \Pi(a(\cdot))$ and therefore

$$\begin{aligned} \frac{\beta(tc) - \beta(t)}{\alpha(t)} &= \rho\big(b(tc) - b(t)\big) \\ &= \rho \frac{(b(tc) - b(t))}{a(t)} a(t) \sim \rho \log c \cdot a(t) \to 0. \end{aligned} \tag{43}$$

Thus $\psi_2(x)$ in (11) is identically 0 and $\psi_1(x) \equiv 1$.

We now see why (41) and (42) are true. We write,

$$\begin{aligned} &tP\left[X - \rho b(t) \le x, \frac{Y - b(t)}{a(t)} > y\right] \\ &\quad = tP\left[\sqrt{1-\rho^2}N_1 + \rho N_2 - \rho b(t) \le x, \frac{N_2 - b(t)}{a(t)} > y\right] \\ &\quad = \int_{a(t)y+b(t)}^{\infty} P\big[\sqrt{1-\rho^2}N_1 + \rho s - \rho b(t) \le x\big] tn(s)\,ds \\ &\quad = \int_y^{\infty} P\big[\sqrt{1-\rho^2}N_1 + \rho\big(a(t)u + b(t)\big) - \rho b(t) \le x\big] \\ &\qquad\qquad \times ta(t)n\big(a(t)u + b(t)\big)\,du \\ &\quad \sim \int_y^{\infty} P\big[\sqrt{1-\rho^2}N_1 \le x - \rho a(t)u\big]e^{-u}\,du \end{aligned}$$

since $ta(t)n(a(t)u + b(t)) \to e^{-u}$. Using the fact that $a(t) \to 0$, we get convergence to

$$\to \int_y^{\infty} P\big[\sqrt{1-\rho^2}N_1 \le x\big]e^{-u}\,du = N\big(x/\sqrt{1-\rho^2}\big)e^{-y},$$

as claimed.

Conclusion: The limit measure is a product measure, $(\psi_1, \psi_2) \equiv (1, 0)$ and $\alpha(t) = 1$. We have an illustration of Proposition 2.

6.1.2. *Exponential marginals for X.* In light of the standard form result (42) it is tempting to look at limits for $(b^{\leftarrow}(X), b^{\leftarrow}(Y))$ but this turns out not to work. The reason for this is explored in Section 6.1.3. Instead, following [21], we consider $(\log b^{\leftarrow}(X), \log b^{\leftarrow}(Y))$. Thus we can transform X to have exponential marginals but not Pareto marginals.

We show the standard form

$$\begin{aligned} tP\left[\frac{\log b^{\leftarrow}(X) - \log b^{\leftarrow}(\rho b(t))}{\rho b(t)} \le x, \frac{b^{\leftarrow}(Y)}{t} > y\right] \\ \to N\left(\frac{x}{\sqrt{1-\rho^2}}\right) y^{-1}. \end{aligned} \tag{44}$$

The verification of (44) needs the following lemma.

LEMMA 1. *The function*

$$V(t) := -\log \overline{N}(\log t) = \log b^{\leftarrow}(\log t) \in \Pi(\log t)$$

is Π-varying with auxiliary function $g(t) = \log t$.

PROOF. To prove membership in the Π-class, it suffices according to de Haan [14] (see alternatively [34], page 30), to show $V'(t) \in RV_{-1}$ and then the auxiliary function can be taken to be $tV'(t)$. So it suffices to show

$$(-\log \overline{N}(\log t))' \sim \frac{\log t}{t} \in RV_{-1}.$$

The derivative is

$$\frac{n(\log t)t^{-1}}{\overline{N}(\log t)} \sim \frac{n(\log t)t^{-1}}{n(\log t)/\log t} = t^{-1}\log t \in RV_{-1}. \qquad \square$$

To show (44), we use (42) and the Delta method. The left-hand side of (44) is

$$\begin{aligned} tP\left[\frac{V(e^{X-\rho b(t)}e^{\rho b(t)}) - V(e^{\rho b(t)})}{g(e^{\rho b(t)})} \le x, \frac{b^{\leftarrow}(Y)}{t} > y\right] \\ \to P\left[\log e^{N_1\sqrt{1-\rho^2}} \le x\right] y^{-1} = N\left(\frac{x}{\sqrt{1-\rho^2}}\right) y^{-1}. \end{aligned}$$

Here is the conditional form of (44), where X is transformed to have exponential marginals:

$$\begin{aligned} \lim_{t\to\infty} P\left[\frac{\log b^{\leftarrow}(X) - \log b^{\leftarrow}(\rho b(t))}{\rho b(t)} \le x \middle| Y > b(t)\right] \\ = \lim_{t\to\infty} P\left[\frac{\log b^{\leftarrow}(X) - \log b^{\leftarrow}(\rho t)}{\rho t} \le x \middle| Y > t\right] = N\left(\frac{x}{\sqrt{1-\rho^2}}\right). \end{aligned}$$

The conditional form of (42), where the marginal distribution is normal, has the same limit:

$$\lim_{t\to\infty} P[X - \rho b(t) \le x | Y > b(t)] = \lim_{t\to\infty} P[X - \rho t \le x | Y > t] = N\left(\frac{x}{\sqrt{1-\rho^2}}\right).$$

This result seems natural when one observes that the normal distribution is in the domain of attraction of the Gumbel distribution.

After transformation of X to exponential marginals, we have for (44)

$$\beta(t) = -\log \overline{N}(\rho b(t)), \qquad \alpha(t) = \rho b(t),$$

and again $\psi_2(t) = 0$, since

$$\begin{aligned}\frac{\beta(tc) - \beta(t)}{\rho b(t)} &= \frac{\log(\overline{N}(\rho b(tc))/\overline{N}(\rho b(t)))}{\rho b(t)} \sim \frac{\log(n(\rho b(tc))/n(\rho b(t)))}{\rho b(t)} \\ &\sim \frac{\log e^{(\rho^2/2)(b^2(tc)-b^2(t))}}{\rho b(t)} = \frac{\rho^2}{2}\big(b(tc) - b(t)\big)\frac{(b(tc)+b(t))}{\rho b(t)} \\ &\sim \rho\big(b(tc) - b(t)\big) \to 0,\end{aligned}$$

using the same argument as in (43). (This provides another illustration of Proposition 2.)

6.1.3. *Why X cannot be transformed to Pareto.* It is noteworthy that one cannot transform X to have Pareto marginals and expect the analogue of (41) to hold. Here is the explanation which also relates to the discussion in Section 7.

Suppose for some choice of centering and scaling $\alpha_2(t) > 0$, $\beta_2(t) \in \mathbb{R}$ we have

$$\lim_{t\to\infty} tP\left[\frac{b^{\leftarrow}(X) - \beta_2(t)}{\alpha_2(t)} \le x, \frac{b^{\leftarrow}(Y)}{t} > y\right] \tag{45}$$

exists and is nondegenerate in the sense of condition (iii) stated at the beginning of Section 2. This expression (45) equals

$$\lim_{t\to\infty} P\left[X - \rho b(t) \le b\big(\alpha_2(t)x + \beta_2(t)\big) - \rho b(t), \frac{b^{\leftarrow}(Y)}{t} > y\right] \tag{46}$$

and from (41) we would have for some nondecreasing limit $\psi(x)$, that as $t \to \infty$,

$$b\big(\alpha_2(t)x + \beta_2(t)\big) - \rho b(t) \to \psi(x). \tag{47}$$

Furthermore, the limit in (45) would have to be

$$N\left(\frac{\psi(x)}{\sqrt{1-\rho^2}}\right)y^{-1}. \tag{48}$$

Inverting (47), we would need

$$\frac{b^{\leftarrow}(y + \rho b(t)) - \beta_2(t)}{\alpha_2(t)} \to \psi^{\leftarrow}(y).$$

Changing variables leads to

$$\frac{b^{\leftarrow}(\log tx)) - \beta_2(b^{\leftarrow}(\log t/\rho))}{\alpha_2(b^{\leftarrow}(\log t/\rho))} \to \psi^{\leftarrow}(\log x).$$

If $\psi^{\leftarrow}$ is not constant, then ([11], page 16)

$$b^{\leftarrow} \circ \log = \left(\frac{1}{1-N}\right) \circ \log$$

is either regularly varying with positive index or it is Π-varying. Neither of these possibilities is true. If $\psi^{\leftarrow}$ is constant, then the limit (48) fails the nondegeneracy assumptions.

So assuming the nondegenerate limit exists in (45) leads to a contradiction. This illustrates the restrictions in our ability to standardize the X variable discussed in Section 2.4.

6.2. *Heavy tailed examples.* In this section, we present examples of heavy tailed random variables possessing asymptotic independence.

6.2.1. *Mixture of independent standard regularly varying random variables* I: *positive* ρ. Suppose nonnegative random variables (U, V) have a joint distribution which is standard regularly varying; that is, there is a limit measure ν on $[0,\infty]^2 \setminus \{\mathbf{0}\}$ such that

$$tP\left[\left(\frac{U}{t}, \frac{V}{t}\right) \in \cdot\right] \xrightarrow{v} \nu$$

in $M_+([0,\infty]^2 \setminus \{\mathbf{0}\})$. For example, (U, V) could be max-stable ([34], Chapter 5), [17] with exponent ν. Suppose (U_i, V_i), $i = 1, 2$, are i.i.d. copies of (U, V). For $0 < p < 1$, define

$$(X, Y) = B(U_1, V_1^p) + (1-B)(U_2^p, V_2), \tag{49}$$

where $P[B=0] = P[B=1] = \frac{1}{2}$, and B is independent of (U_i, V_i), $i = 1, 2$.

Observe that for any $x > 0$, $y > 0$

$$\begin{aligned} &tP\left\{\left[\frac{X}{t} \le x, \frac{Y}{t} \le y\right]^c\right\} \\ &\quad = \frac{t}{2}P\left[\frac{U_1}{t} > x \text{ or } \frac{V_1^p}{t} > y\right] + \frac{t}{2}P\left[\frac{U_2^p}{t} > x \text{ or } \frac{V_2}{t} > y\right] \qquad (50) \\ &\quad = \frac{t}{2}P[U_1 > tx] + o(1) + \frac{t}{2}P[V_2 > ty] + o(1) \to \frac{1}{2}(x^{-1} + y^{-1}). \end{aligned}$$

So (X, Y) is standard regularly varying, in a domain of attraction of a multivariate extreme value distribution, and possesses asymptotic independence. The asymptotic independence holds even if (U, V) has no asymptotic independence.

Now observe that

$$
\begin{aligned}
&tP\left[\frac{X}{t^p} \le x, \frac{Y}{t} > y\right] \\
&\quad = \frac{t}{2} P[U_1 \le t^p x, V_1^p > ty] + \frac{t}{2} P[U_2^p \le t^p x, V_2 > ty] \\
&\quad = \frac{t}{2} P[U_1 \le t^p x, V_1 > t^{1/p} y^{1/p}] + \frac{t}{2} P[U_2 \le t x^{1/p}, V_2 > ty] \\
&\quad \to 0 + \frac{1}{2}\nu\big([0, x^{1/p}] \times (y, \infty]\big) =: \mu\big([0, x] \times (y, \infty]\big).
\end{aligned} \tag{51}
$$

If (U, V) possess asymptotic independence, then $\nu((0, \infty]^2) = 0$ and the non-degeneracy assumption for μ stated in (8) fails since for fixed $y > 0$, the function of x given by $\nu([0, x^{1/p}] \times (y, \infty])$ concentrates at $x = 0$. So for this example, (X, Y) is standard regularly varying, asymptotically independent and provided (U, V) does not possess asymptotic independence, we can refine the asymptotic independence to get the limit in (8). This gives an example of case (i) of (14) with $\rho = p$, $\beta(t) = (1/\rho)\alpha(t) = t^p$. The conditional limit distribution can most simply be written as

$$
\lim_{t\to\infty} P\left[\frac{X}{t^p} \le x \,\middle|\, Y > t\right] = \frac{1}{2}\nu\big([0, x^{1/p}] \times (1, \infty]\big).
$$

(Note that the normalization of the X variable may have to be properly scaled by ct^p for some $c > 0$ to ensure the limit is a probability distribution.)

The details of this construction can be repeated in modestly greater generality with (49) modified as

$$
(X, Y) = B(U_1, h(V_1)) + (1 - B)(h(U_2), V_2), \tag{52}
$$

with $h \in RV_p$ and $h(t)/t \to 0$. As before, (X, Y) is standard regularly varying and asymptotically independent and

$$
tP\left[\left(\frac{X}{h(t)}, \frac{Y}{t}\right) \in \cdot\right] \xrightarrow{v} \mu(\cdot), \tag{53}
$$

where μ is given as in (51). The condition $h(t)/t \to 0$ is necessary and sufficient for (X, Y) to be asymptotically independent as can be seen by examining the calculations leading to (50).

6.2.2. *Mixture of independent standard regularly varying random variables* II; *negative* ρ. To exemplify case (iii) of (14) where $\rho < 0$, suppose (52), (53) still hold, $h(t)/t \to 0$ and (U, V) are not asymptotically independent. Define $\widetilde{X} = 1/X$, $\tilde{h} = 1/h \in RV_{-p}$, and a measure $\tilde{\mu}$ on $[0, \infty] \times (0, \infty]$ by

$$
\tilde{\mu}\big([0, x] \times (y, \infty]\big) = \mu\left(\left[\frac{1}{x}, \infty\right] \times (y, \infty]\right).
$$

Then

$$tP\left[\left(\frac{\widetilde{X}}{\tilde{h}(t)}, \frac{Y}{t}\right) \in \cdot\right] \xrightarrow{v} \tilde{\mu}(\cdot),$$

in $M_+([0, \infty] \times (0, \infty])$. The reason this works is that the first space in the product $[0, \infty] \times (0, \infty]$ is compact:

$$tP\left[\frac{\widetilde{X}}{\tilde{h}(t)} \le x, \frac{Y}{t} > y\right] = tP\left[\frac{X}{h(t)} \ge \frac{1}{x}, \frac{Y}{t} > y\right] \to \mu\left(\left[\frac{1}{x}, \infty\right] \times (y, \infty]\right).$$

So using $(\widetilde{X}, Y)$, we have an example of case (iii) of (14) where $\rho = -p < 0$, $\alpha(t) = \beta(t) = \tilde{h}(t)$. The conditioned limit distribution is

$$H(x) = \lim_{t\to\infty} P[\widetilde{X}/\tilde{h}(t) \le x | Y > t] = \mu\left(\left[\frac{1}{x}, \infty\right] \times (1, \infty]\right).$$

6.2.3. *Mixture of independent standard regularly varying random variables* III; $\rho = 0$. Finally, suppose (52) still holds but this time suppose $h \in \Pi(g)$ is nondecreasing and Π-varying with auxiliary function $g(t)$. [E.g., we could take $h(t) = \log t$, $g(t) = 1$.] Then $h(t)/t \to 0$ as $t \to \infty$ so (X, Y) is standard regularly varying as well as asymptotically independent. To verify this we need the fact that if ξ is either U or V, then

$$tP\left[\frac{h(\xi)}{t} > x\right] \to 0 \qquad (x > 0, t \to \infty). \tag{54}$$

To see this, let K be a large number and

$$\begin{aligned} tP\left[\frac{h(\xi)}{t} > x\right] &= tP\left[\frac{h(\xi)}{t} > x, \xi \le tK\right] + tP\left[\frac{h(\xi)}{t} > x, \xi > tK\right] \\ &\le o(1) + tP[\xi > tK] \to K^{-1}. \end{aligned}$$

The upper bound is arbitrarily small and thus we verified (54).

Now we check that (X, Y) is standard regularly varying and asymptotically independent:

$$\begin{aligned} & tP\left[\frac{X}{t} > x \text{ or } \frac{Y}{t} > y\right] \\ &\quad = \frac{t}{2}P\left[\frac{U_1}{t} > x \text{ or } \frac{h(V_1)}{t} > y\right] + \frac{t}{2}P\left[\frac{h(U_2)}{t} > x \text{ or } \frac{V_2}{t} > y\right] \\ &\quad = o(1) + \frac{t}{2}P\left[\frac{U_1}{t} > x\right] + \frac{t}{2}P\left[\frac{V_2}{t} > y\right] \to \frac{1}{2}(x^{-1} + y^{-1}). \end{aligned}$$

Note we applied (54).

Next consider

$$
\begin{aligned}
&tP\left[\frac{X-h(t)}{g(t)} \le x, \frac{Y}{t} > y\right]\\
&\quad = o(1) + \frac{t}{2}P\left[\frac{h(U_2)-h(t)}{g(t)} \le x, \frac{V_2}{t} > y\right]\\
&\quad \sim \frac{t}{2}P\left[\frac{U_2}{t} \le \frac{h^{\leftarrow}(g(t)x+h(t))}{t}, \frac{V_2}{t} > y\right] \sim \frac{t}{2}P\left[\frac{U_2}{t} \le e^x, \frac{V_2}{t} > y\right]\\
&\quad \to \frac{1}{2}\nu\big([0,e^x]\times(y,\infty]\big).
\end{aligned}
$$

This exemplifies case (ii) of (14) with $\rho = 0$, $\beta(t) = h(t)$ and $\alpha(t) = g(t)$. The form of the conditioned limit is

$$
P\left[\frac{X-h(t)}{g(t)} \le x \Big| Y > t\right] \to \frac{1}{2}\nu\big([0,e^x]\times(1,\infty]\big) =: H(x), \qquad x \in \mathbb{R}.
$$

7. Change of coordinate system. How much freedom do we have to measure the X-variable in different units? This issue was raised in the discussion to Heffernan and Tawn [21] and we try to offer further insight on the matter here. For the example in Section 6.1.3 we saw that for (X, Y) bivariate normal, it was possible to transform $X \mapsto \log b^{\leftarrow}(X)$ and get a conditional limit but the transformation $X \mapsto b^{\leftarrow}(X)$ did not preserve existence of conditional limits. Can something more general be said about this issue?

Starting with (8) where the Y-variable is standardized, for what monotone increasing functions $h(\cdot)$ do there exist centering and scaling functions $\alpha_2(t) > 0$, $\beta_2(t) \in \mathbb{R}$, such that for some limit measure μ_2 satisfying the nondegeneracy assumptions at the beginning of Section 2 we have

$$
tP\left[\left(\frac{h(X)-\beta_2(t)}{\alpha_2(t)}, \frac{Y}{t}\right) \in \cdot\right] \xrightarrow{v} \mu_2 \tag{55}
$$

in $M_+([-\infty,\infty]\times(0,\infty])$? This problem has many similarities to ones considered in [2, 33] and the experience gained in Section 6.1.3 is helpful.

In (8), assume centering by $\beta(t)$ is really necessary; that is, suppose it is not the case that $\beta(t) = o(\alpha(t))$. [If $\beta(t) = o(\alpha(t))$, the following arguments are easier and lead to regular variation of h.] Assume (55) and rewrite the left side of (55) evaluated on $[-\infty, x]\times(y,\infty]$ as

$$
tP\left[\frac{X-\beta(t)}{\alpha(t)} \le \frac{h^{\leftarrow}(\alpha_2(t)x+\beta_2(t))-\beta(t)}{\alpha(t)}, \frac{Y}{t} > y\right].
$$

Since this converges, there must exist a limit $\psi(x)$ such that

$$
\frac{h^{\leftarrow}(\alpha_2(t)x+\beta_2(t))-\beta(t)}{\alpha(t)} \to \psi(x) \tag{56}
$$

and then we see that

$$\mu\big([-\infty, \psi(x)] \times (y, \infty]\big) = \mu_2\big([-\infty, x] \times (y, \infty]\big). \tag{57}$$

The limit ψ cannot be constant without violating the nondegeneracy assumption for μ_2. Inverting (56) we get

$$\frac{h(y\alpha(t) + \beta(t)) - \beta_2(t)}{\alpha_2(t)} \to \psi^{\leftarrow}(y).$$

This suggests we set

$$\beta_2(t) = h(\beta(t)), \tag{58}$$

since

$$\frac{h(y\alpha(t) + \beta(t)) - h(\beta(t))}{\alpha_2(t)} \to \psi^{\leftarrow}(y) - \psi^{\leftarrow}(0) =: \chi(y) \tag{59}$$

and presuming $\chi(1) > 0$, we could set

$$\alpha_2(t) = h\big(\alpha(t) + \beta(t)\big) - h(\beta(t)).$$

We now look at some possible forms of h which allow change of coordinate system (55). We do not achieve necessary and sufficient conditions but come to an understanding of how to generate broad classes of functions h permitting nonlinear transformation of X.

7.1. *Case* A: $\alpha(t)$ *is asymptotically a constant.* Assume $\beta(t) \uparrow \infty$ as $t \to \infty$. If $\alpha \sim 1$, then

$$\frac{h(y + \beta(t)) - h(\beta(t))}{\alpha_2(t)} \to \chi(y),$$

and changing variables yields

$$\frac{h(y + t) - h(t)}{\alpha_2(\beta^{\leftarrow}(t))} \to \chi(y),$$

or

$$\frac{h(\log tx) - h(\log t)}{\alpha_2(\beta^{\leftarrow}(\log t))} \to \chi(\log x), \qquad x > 0. \tag{60}$$

Since $h \circ \log$ is nondecreasing, either [11]

(a) $h \circ \log \in RV_p$, $p > 0$, in which case $\alpha_2(\beta^{\leftarrow}(\log t)) \sim h(\log t)$
or
(b) $h \circ \log \in \Pi(\alpha_2 \circ \beta^{\leftarrow}(\log t))$.

Conclusion: If $\alpha \sim 1$, we may change coordinates $X \mapsto h(X)$, provided $h \circ \log \in RV_p \cup \Pi(\alpha_2 \circ \beta^{\leftarrow}(\log t))$.

REMARK 3. 1. In Section 6.1.3, $\alpha(t) = 1$. We tried $h(x) = b^{\leftarrow}(x)$ but did not get a conditioned limit law. In Section 6.1.3, $h \circ \log = b^{\leftarrow} \circ \log$ is neither regularly varying, nor Π-varying.

2. In Section 6.1.2, $\alpha(t) = 1$. We tried $h(x) = \log b^{\leftarrow}(x)$ which led to a conditioned limit law because Lemma 1 proved $h \circ \log = \log b^{\leftarrow} \circ \log \in \Pi(\log)$.
3. The result in (b) suggests how to construct other examples of h which lead to conditioned limits. If g is any slowly varying function, then $\int_1^x g(u)u^{-1}du$ is Π-varying with auxiliary function g ([14], [34], page 30). Define h by $h(\log x) = \int_1^x g(u)/u\,du$ or

$$h'(x) = g(e^x), \qquad h(x) = \int_0^x g(e^u)\,du.$$

Any such h will lead to a conditioned limit. Examples include:

- $g(x) = \log x$ and $h(x) = x^2/2$.
- $g(x) = \log\log x$ and $h(x) = \int_0^x \log u\,du \sim x\log x$.
- $g(x) = (\log x)^p$ and $h(x) = \frac{x^{p+1}}{p+1}$ for $p > 0$.

For an example where $h \circ \log \in RV_p$ for $p > 0$, set

$$h(\log x) = U(x) \in RV_p \quad \text{or} \quad h(x) = U(e^x).$$

Apply this to the convergence (42) for the bivariate normal pair (X, Y) where recall

$$\beta(t) = \rho b(t), \qquad \alpha(t) = 1, \qquad \mu([-\infty, x] \times (y, \infty]) = N\left(\frac{x}{\sqrt{1-\rho^2}}\right)y^{-1}.$$

Then evaluating (60) with $h(\log t) = U(t) \in RV_p$, $p > 0$, gives, with $\alpha_2 \circ \beta^{\leftarrow} \circ \log = U$ that

$$\frac{U(tx) - U(t)}{U(t)} \to x^p - 1 = \chi(\log x).$$

Therefore, $\chi(y) = e^{py} - 1$, and from (57)

$$tP\left[\frac{U(e^X) - U(e^{\rho b(t)})}{U(e^{\rho b(t)})} \le x, \frac{b^{\leftarrow}(Y)}{t} > y\right] = \mu([-\infty, \chi^{\leftarrow}(x)] \times (y, \infty])$$
$$= N\left(\frac{p^{-1}\log(1+x)}{\sqrt{1-\rho^2}}\right)y^{-1}.$$

So for this example, $\beta_2(t) = \alpha_2(t) = U(e^{\rho b(t)})$.

7.2. *Case* B: $\alpha(t)$ *is not asymptotically a constant.* Again assume $\beta(t) \uparrow \infty$ as $t \to \infty$. Transform (59) to get

$$\frac{h(y\alpha \circ \beta^{\leftarrow}(t) + t) - h(t)}{\alpha_2 \circ \beta^{\leftarrow}(t)} \to \chi(y) \tag{61}$$

which is of the form

$$\frac{h(t + f(t)y) - h(t)}{\alpha_*(t)} \to \chi(y).$$

To proceed further in a way that generates a broad class examples, suppose $f(t) = \alpha \circ \beta^{\leftarrow}(t)$ is *self-neglecting* [4]. A simple sufficient condition is $f'(t) \to 0$ and f self-neglecting means it is the auxiliary function of a Γ-varying function (see Appendix A.2) and that

$$H(x) := \exp\left\{\int_1^x \frac{1}{f(u)}\, du\right\} \in \Gamma(f).$$

Then defining the function V by

$$h = V \circ H \text{ or equivalently } V = h \circ H^{\leftarrow}$$

we have either ([14], page 249, [34], page 36)

(a) $V \in \Pi$ and $\chi(y) = \log e^y = y$;

or

(b) $V \in RV_p$, $p > 0$ and $\chi(y) = e^{py} - 1$.

Conclusion: We considered the case that $\beta \neq o(\alpha)$ and $\beta(t) \uparrow \infty$ and α not asymptotically a constant. For such a case, the change of variable $X \mapsto h(X)$ preserves conditioned limits provided h is either the composition of a Π-varying function and a Γ-varying function or the composition of a regularly varying function and a Γ-varying function. (The composition of a regularly varying function and a Γ-varying function is another Γ-varying function; see [12], [34], page 36).

8. Discussion and concluding remarks.

The statistical models proposed by Heffernan and Tawn [21] are based on the assumption that for (X, Y) having Gumbel marginal distributions, there exist normalizing functions $\alpha(\cdot)$ and $\beta(\cdot)$ such that the conditional distribution of $(X - \beta(y))/\alpha(y)$ given $Y = y$ can be approximated for large y by some nondegenerate, proper $G(x)$. We have built our theory by standardizing Y to have asymptotically Pareto distribution and looked at the conditional distribution of $(X - \beta(t))/\alpha(t)$ given $Y > t$ which also leads to conditional distributions for $(X - \beta(Y))/\alpha(Y)$ given $Y > t$. This formulation is consistent with the Heffernan and Tawn [21] approach and allows a mathematically precise theory which can be related to the extended theory of multivariate regular variation.

From the perspective of statistical modeling, important results are contained in Propositions 4 and 5. These propositions reveal the factorization of the limit distribution obtained when X is normalized by the value of Y that occurs with it. This factorization permits a significant simplification of models based on the limit form, as it enables the assumption of limiting independence between the conditioning and standardized variables. This independence assumption was employed in [21] and is key to statistical modeling and extrapolation.

One issue we have not resolved is consistency of different models. The definition (5) or its standardized version (8) is not symmetric in the X, Y variables. However, when fitting models to data one has a choice of which variable to condition being large and a logical issue is whether the various models obtained by conditioning on different variables are related to each other in any way. Conditions for consistency would strengthen the statistical model assumptions based on this representation and therefore potentially improve the ability of such approaches to describe the joint distribution in tail regions where there is naturally little data. Currently we have nothing terribly useful to say on this issue other than to point out that it seems important to understand consistency better.

APPENDICES

For convenience, this section collects some notation, needed background on regular variation and notions on vague convergence needed for some formulations and proofs.

A.1. Vector notation. Vectors are denoted by bold letters, capitals for random vectors and lower case for nonrandom vectors. For example: $\mathbf{x} = (x^{(1)}, \dots, x^{(d)}) \in \mathbb{R}^d$. Operations between vectors should be interpreted componentwize so that for two vectors $\mathbf{x}$ and $\mathbf{z}$

$$\begin{aligned}
\mathbf{x} < \mathbf{z} &\text{ means } x^{(i)} < z^{(i)}, \qquad i = 1, \dots, d,\\
\mathbf{x} \le \mathbf{z} &\text{ means } x^{(i)} \le z^{(i)}, \qquad i = 1, \dots, d,\\
\mathbf{x} = \mathbf{z} &\text{ means } x^{(i)} = z^{(i)}, \qquad i = 1, \dots, d,\\
\mathbf{z}\mathbf{x} &= \bigl(z^{(1)}x^{(1)}, \dots, z^{(d)}x^{(d)}\bigr),\\
\mathbf{x} \vee \mathbf{z} &= \bigl(x^{(1)} \vee z^{(1)}, \dots, x^{(d)} \vee z^{(d)}\bigr), \qquad \frac{\mathbf{x}}{\mathbf{z}} = \left(\frac{x^{(1)}}{z^{(1)}}, \dots, \frac{x^{(d)}}{z^{(d)}}\right),
\end{aligned}$$

and so on. Also define $\mathbf{0} = (0, \dots, 0)$. For a real number c, denote as usual $c\mathbf{x} = (cx^{(1)}, \dots, cx^{(d)})$. We denote the rectangles (or the higher dimensional intervals) by

$$[\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{a} \le \mathbf{x} \le \mathbf{b}\}.$$

Higher dimensional rectangles with one or both endpoints open are defined analogously, for example,

$$(\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \in \mathbb{R}^d : \mathbf{a} < \mathbf{x} \le \mathbf{b}\}.$$

A.2. The function classes Π and Γ. Continue the domain of attraction discussion: Writing (3) as

$$\left(\frac{1}{1 - F(a(t)x + b(t))}\right)\Big/ t \to (1 + \gamma x)^{1/\gamma}$$

and inverting yields as $t \to \infty$

$$\frac{b(ty)-b(t)}{a(t)} \to \begin{cases} \dfrac{y^\gamma - 1}{\gamma}, & \text{if } \gamma \neq 0, \\ \log y, & \text{if } \gamma = 0. \end{cases} \tag{62}$$

In case $\gamma = 0$, (62) says that $b(\cdot) \in \Pi(a(\cdot))$; that is, the function $b(\cdot)$ is Π-varying with auxiliary function $a(\cdot)$ ([34], pages 26ff, [4, 11, 12]).

More generally ([4], Chapter 3, [18]) define for an auxiliary function $a(t) > 0$, $\Pi_+(a)$ to be the set of all functions $\pi : \mathbb{R}_+ \mapsto \mathbb{R}_+$ such that

$$\lim_{t\to\infty} \frac{\pi(tx) - \pi(t)}{a(t)} = k \log x, \qquad x > 0, k > 0. \tag{63}$$

The class $\Pi_-(a)$ is defined similarly except that $k < 0$ and

$$\Pi(a) = \Pi_+(a) \cup \Pi_-(a).$$

By adjusting the auxiliary function in the denominator, it is always possible to assume $k = \pm 1$.

Two functions $\pi_i \in \Pi_\pm(a)$, $i = 1, 2$, are $\Pi(a)$-equivalent if for some $c \in \mathbb{R}$

$$\lim_{t\to\infty} \frac{\pi_1(t) - \pi_2(t)}{a(t)} = c.$$

There is usually no loss of generality in assuming $c = 0$.

The class of regularly varying functions with index $\rho \in \mathbb{R}$ is denoted by RV_ρ so that $U : \mathbb{R}_+ \mapsto \mathbb{R}_+$ satisfies $U \in RV_\rho$ if

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\rho, \qquad x > 0. \tag{64}$$

The following are known facts about Π-varying functions.

1. We have $\pi \in \Pi_+(a)$ iff $1/\pi \in \Pi_-(a/\pi^2)$.
2. If $\pi \in \Pi_+(a)$, then ([4], page 159 or [18], page 1031) there exists a continuous and strictly increasing $\Pi(a)$-equivalent function π_0 with $\pi - \pi_0 = o(a)$.
3. If $\pi \in \Pi_+(a)$, then

   $$\lim_{t\to\infty} \pi(t) =: \pi(\infty)$$

   exists. If $\pi(\infty) = \infty$, then $\pi \in RV_0$ and $\pi(t)/a(t) \to \infty$. If $\pi(\infty) < \infty$, then $\pi(\infty) - \pi(t) \in \Pi_-(a)$ and $\pi(\infty) - \pi(t) \in RV_0$ and $(\pi(\infty) - \pi(t))/a(t) \to \infty$. (Cf. [11], page 25.) Furthermore,

   $$\frac{1}{\pi(\infty) - \pi(t)} \in \Pi_+\big(a/\big(\pi(\infty) - \pi(t)\big)^2\big).$$

In addition to the function class Π we need de Haan's class Γ ([4, 11–13, 34]). A function $V:\mathbb{R}_+ \mapsto \mathbb{R}_+$ is a Γ-function with auxiliary function f [written $V \in \Gamma(f)$] if, as $t \to \infty$,

$$\frac{V(t + xf(t))}{V(t)} \to e^x, \qquad x > 0.$$

For V nondecreasing, $V \in \Gamma(f)$ iff $V^{\leftarrow} \in \Pi(f \circ V^{\leftarrow})$.

A.3. Vague convergence. For a nice space $\mathbb{E}$, that is, a space which is locally compact with countable base (e.g., a finite dimensional Euclidean space), denote $M_+(\mathbb{E})$ for the nonnegative Radon measures on Borel subsets of $\mathbb{E}$. This space is metrized by the vague metric. The notion of vague convergence in this space is as follows: If $\mu_n \in M_+(\mathbb{E})$ for $n \geq 0$, then μ_n converge vaguely to μ_0 (written $\mu_n \stackrel{v}{\to} \mu_0$) if for all bounded continuous functions f with compact support we have

$$\int_{\mathbb{E}} f\,d\mu_n \to \int_{\mathbb{E}} f\,d\mu_0 \qquad (n \to \infty).$$

This concept allows us to write (3) as

$$tP\left[\frac{Y - b(t)}{a(t)} \in \cdot\right] \stackrel{v}{\to} m_\gamma(\cdot), \tag{65}$$

vaguely in $M_+((-\infty, \infty])$ where

$$m_\gamma((x, \infty]) = (1 + \gamma x)^{-1/\gamma}.$$

Standard references include [22, 29] and [34], Chapter 3.

Acknowledgments. The authors would like to thank two anonymous referees for careful reading of the initial submission, and their comments which contributed to a substantially improved final version.

Thanks to Lancaster University and Cornell University's School of Operations Research and Industrial Engineering and Department of Statistics for funding and hospitality during a visit to Cornell in September 2004.

REFERENCES

[1] ABDOUS, B., FOUGÈRES, A.-L. and GHOUDI, K. (2005). Extreme behaviour for bivariate elliptical distributions. *Canad. J. Statist.* **33** 317–334. MR2193978

[2] BALKEMA, A. A. (1973). *Monotone Transformations and Limit Laws*. Mathematisch Centrum, Amsterdam. MR0334307

[3] BASRAK, B., DAVIS, R. and MIKOSCH, T. (2002). A characterization of multivariate regular variation. *Ann. Appl. Probab.* **12** 908–920. MR1925445

[4] BINGHAM, N., GOLDIE, C. and TEUGELS, J. (1987). *Regular Variation*. Cambridge Univ. Press. MR0898871

[5] COLES, S. G. and TAWN, J. A. (1991). Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B* **53** 377–392. MR1108334

[6] COLES, S. G. and TAWN, J. A. (1994). Statistical methods for multivariate extremes: An application to structural design (with discussion). *J. R. Stat. Soc. Ser. C* **43** 1–48.

[7] COLES, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer, London. MR1932132

[8] DAVISON, A. C. and SMITH, R. L. (1990). Models for exceedances over high thresholds (with discussion). *J. Roy. Statist. Soc. Ser. B* **52** 393–442. MR1086795

[9] EMBRECHTS, P., KLUPPELBERG, C. and MIKOSCH, T. (1997). *Modelling Extreme Events for Insurance and Finance*. Springer, Berlin. MR1458613

[10] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications*, 2nd ed. Wiley, New York.

[11] GELUK, J. L. and DE HAAN, L. (1987). *Regular Variation*, *Extensions and Tauberian Theorems*. Stichting Mathematisch Centrum, Amsterdam. MR0906871

[12] DE HAAN, L. (1970). *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes*. Math. Centrum, Amsterdam. MR0286156

[13] DE HAAN, L. (1974). Equivalence classes of regularly varying functions. *Stochastic Process. Appl.* **2** 243–259. MR0365664

[14] DE HAAN, L. (1976). An Abel–Tauber theorem for Laplace transforms. *J. London Math. Soc. (2)* **13** 537–542. MR0407542

[15] DE HAAN, L. (1985). Extremes in higher dimensions: The model and some statistics. In *Proceedings of the 45th Session of the International Statistical Institute* **4** (*Amsterdam*, *1985*) 185–192. *Bull. Inst. Internat. Statist.* **51**. ISI, Hague, Netherlands. MR0886266

[16] DE HAAN, L. and DE RONDE, J. (1998). Sea and wind: Multivariate extremes at work. *Extremes* **1** 7–46. MR1652944

[17] DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory*: *An Introduction*. Springer, New York. MR2234156

[18] DE HAAN, L. and RESNICK, S. I. (1979). Conjugate π-variation and process inversion. *Ann. Probab.* **7** 1028–1035. MR0548896

[19] DE HAAN, L. and RESNICK, S. (1977). Limit theory for multivariate sample extremes. *Z. Wahrsch. Verw. Gebiete* **40** 317–337. MR0478290

[20] HEFFERNAN, J. and RESNICK, S. (2005). Hidden regular variation and the rank transform. *Adv. in Appl. Probab.* **37** 393–414. MR2144559

[21] HEFFERNAN, J. and TAWN. J. (2004). A conditional approach for multivariate extreme values (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** 497–546. MR2088289

[22] KALLENBERG, O. (1983). *Random Measures*, 3rd ed. Akademie-Verlag, Berlin. MR0818219

[23] KURATOWSKI, K. (1966). *Topology*. I. New edition, revised and augmented. Translated from the French by J. Jaworowski. Academic Press, New York. MR0217751

[24] LEDFORD, A. and TAWN, J. (1996). Statistics for near independence in multivariate extreme values. *Biometrika* **83** 169–187. MR1399163

[25] LEDFORD, A. and TAWN, J. (1997). Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B* **59** 475–499. MR1440592

[26] LEDFORD, A. and TAWN, J. (1998). Concomitant tail behaviour for extremes. *Adv. in Appl. Probab.* **30** 197–215. MR1618837

[27] MAULIK, K. and RESNICK, S. (2005). Characterizations and examples of hidden regular variation. *Extremes* **7** 31–67. MR2201191

[28] MAULIK, K., RESNICK, S. and ROOTZÉN, H. (2002). Asymptotic independence and a network traffic model. *J. Appl. Probab.* **39** 671–699. MR1938164

[29] NEVEU, J. (1977). Processus ponctuels. *École d'Été de Probabilités de Saint-Flour VI—1976*. *Lecture Notes in Math.* **598** 249–445. Springer, Berlin. MR0474493

[30] PICKANDS, J. (1981). Multivariate extreme value distributions. In *Proceedings of the 43rd Session of the International Statististical Institute* **2** 859–878. MR0820979

[31] REISS, R.-D. and THOMAS, M. (2001). *Statistical Analysis of Extreme Values*, 2nd ed. Birkhäuser, Basel. MR1819648

[32] RESNICK, S. (2006). *Heavy Tail Phenomena*: *Probabilistic and Statistical Modeling*. Springer, New York.

[33] RESNICK, S. (1973). Limit laws for record values. *Stochastic Process. Appl.* **1** 67–82. MR0362454

[34] RESNICK, S. (1987). *Extreme Values*, *Regular Variation and Point Processes*. Springer, New York. MR0900810

[35] RESNICK, S. (1999). *A Probability Path*. Birkhäuser, Boston. MR1664717

[36] RESNICK, S. (2002). Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* **5** 303–336. MR2002121

[37] RESNICK, S. (2004). On the foundations of multivariate heavy-tail analysis. In *Stochastic Methods and Their Applications* (J. Gani and E. Seneta, eds.) 191–212. *J. Appl. Probab.* **41A**. Papers in honour of C. C. Heyde. MR2057574

[38] SENETA, E. (1976). *Regularly Varying Functions*. Springer, New York. MR0453936

[39] SMITH, R. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground level ozone (with discussion). *Statist. Sci.* **4** 367–393. MR1041763

DEPARTMENT OF MATHEMATICS AND STATISTICS
LANCASTER UNIVERSITY
LANCASTER, LA1 4YF
UNITED KINGDOM
E-MAIL: j.heffernan@lancaster.ac.uk

SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK 14853
USA
E-MAIL: sir1@cornell.edu