SPLENDEURS ET MISERES DES LOIS DE VALEURS EXTREMES

N. BOULEAU
Centre de Mathématiques Appliquées
de l'Ecole Nationale des Ponts et Chaussées, Paris.

Ayant observé durant une période d'un siècle dans une certaine région des enregistrements sismiques de magnitude comprise entre 0 et 2, est-il possible d'en déduire avec quelle probabilité se produira dans la même région durant le siècle suivant un séisme de magnitude supérieure à 4 ? A un problème ainsi posé, rares seraient ceux qui répondraient par l'affirmative, néanmoins l'usage de plus en plus répandu dans le milieu des ingénieurs de procédures rapides utilisant les lois de valeurs extrêmes conduit à des affirmations de ce type, dont l'enjeu socio-politique est important notamment par l'habit de scientificité qui leur est donné. Après avoir rappelé les fondements de la théorie des lois de valeurs extrêmes et relevé quelques unes des hypothèses cruciales, difficiles à vérifier en pratique qui la sous-tendent, nous montrons que la méthode qui consiste à caler les paramètres d'une des trois lois de valeurs extrêmes à partir des extrêmes d'un échantillon fini dont la loi est mal connue, est fortement encouragée par la pression sociale de quantifier les risques graves d'autant plus que de tels errements, par la rareté même des événements considérés, sont peu réfutables.

1 Deux belles theories.

A - **Les lois des grands nombres pour les extrêmes.** Etant données n variables aléatoires réelles $X_1, \cdots, X_n$ on appelle "extrêmes" les variables aléatoires

$$M_n = \max_{i=1,\cdots,n} (X_i)$$

et

$$m_n = \min_{i=1,\cdots,n} (X_i)$$

L'étude, sous certaines hypothèses sur la suite (X_n), du comportement asymptotique de quantités liées aux extrêmes est plus récente dans le développement de la théorie des probabilités que l'étude des moyennes qui apparaissent dans les théorèmes des grands nombres ou de limite centrale. Après les travaux de Maurice Fréchet [1] (1927) et de Fischer et Tippett (1928), le théorème clef de la théorie mis en forme par Gnedenko (1943) peut être énoncé ainsi :

Théoreme 1 *Théorème des types extrêmes.*

Soient $X_1, \cdots, X_n, \cdots$ des variables aléatoires réelles indépendantes et de même loi. On pose $M_n = \max_{i=1,\cdots,n} (X_i)$. Il existe des nombres $a_n > 0$ et b_n tels que :

$$a_n(M_n - b_n)$$

converge en loi lorsque $n \uparrow \infty$.

Alors, soit la loi limite est dégénérée, soit la loi limite est la loi de $aX + b$ avec $a > 0$ où X a l'une des fonctions de répartition suivantes

$$\begin{array}{llcll} I) & G(x) & = & \exp(-e^{-x}), & x \in \mathbb{R}, \quad \textit{(Gumbel)} \\ II) & G(x) & = & \exp(-x^{-\alpha}) 1_{x>0}, & \alpha > 0, \quad \textit{(Fréchet)} \\ III) & G(x) & = & 1 - [1 - \exp(-(-x)^{\alpha})] 1_{x \leq 0}, & \alpha > 0, \quad \textit{(Weibull)} \end{array}$$

Soit F la fonction de répartition commune des X_i.

a) Si il existe $g(t) > 0$ telle que

$$lim_{t \uparrow \sup\{s: F(s) < 1\}} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x}$$

on est dans le cas I.

b) Si $\sup\{s : F(s) < 1\} = +\infty$ *et si*

$$\lim_{t\uparrow\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\alpha}, \quad \alpha > 0, \quad \forall x > 0$$

on est dans le cas II

c) Si $a = \sup\{1 : F(s) < 1\} < +\infty$ *et si*

$$\lim_{n\downarrow 0} \frac{1-F(a-xh)}{1-F(a-h)} = x^{\alpha}, \quad \alpha > 0, \quad \forall x > 0$$

on est dans le cas III.

Un résultat analogue peut s'énoncer pour les minima évidemment. Autrement dit les extrêmes peuvent être renormalisés de sorte qu'ils suivent asymptotiquement une loi de l'un ou l'autre des trois types, dont deux dépendent d'un paramètre $\alpha > 0$, et qui sont obtenus selon certaines propriétés de la queue de la distribution commune des (X_i). Par exemple si les (X_i) sont des variables normales réduites on peut prendre $a = \sqrt{2 \log n}$ et

$$b_n = \sqrt{2\log n} - \frac{\log\log n + \log 4\pi}{2\sqrt{2\log n}}$$

et la loi de $a_n M_n + b_n$ converge vers une loi de Gumbel.

Ou bien si les X_i suivent une loi de Cauchy de paramètre c, $\mathbb{P}(\frac{\pi}{cn} M_n \leq x) \to e^{-1/x}$ (loi de Fréchet de paramètre 1).

Ce théorème a reçu, depuis, de nombreuses extensions plus ou moins compliquées, notamment au cas où les variables X_i ne sont plus supposées indépendantes (valeurs extrêmes de processus stationnaires) ou même à des situations plus générales. On a également étudié les statistiques d'ordre multivariées, notamment les lois limites concernant les k plus grandes valeurs atteintes, etc.

La loi de type III ou de Weibull [3] qui est obtenue lorsque la loi des X_i a un support borné à droite (ou borné à gauche si l'on étudie le minimum) se rattache également à une autre théorie, qui lui a valu et lui vaut encore un succès assez étonnant, qui concerne la mécanique de la rupture des barres.

B - **La théorie de Weibull de la résistance des barres à la traction**.

L'idée est de s'appuyer sur le fait que tant que la barre soumise à une force de traction résiste sans se rompre, c'est que tous les tronçons successifs de la barre résistent, et qu'inversement la résistance de la barre (c'est à dire la valeur de la force qui produit la rupture) est la plus petite des résistances de la barre en sous-barres, ceci quelque soit le partage de la barre en sous-barres.

Soit ξ_L la résistance (aléatoire) d'une barre de longueur L partagée en n sous-barres de longueur $\ell_1, \cdots, \ell_n$ ($\ell_1 + \cdots + \ell_n = L$) de résistances respectives $\xi_{\ell_1}, \cdots, \xi_{\ell_n}$.

Si on fait les hypothèses suivantes :

i) $\xi_L = \min_{i=1,\cdots,n} \xi_{\ell_i}$. Ce qu'on exprime parfois en disant que le matériau est <u>stochastiquement fragile</u>.

ii) Les lois des ξ_ℓ ne dépendent que de la longueur ℓ. On dit alors que la matériau est <u>stochastiquement homogène</u>.

iii) Les variables ξ_{ℓ_i} relatives à des sous-barres disjointes sont <u>indépendantes</u>.

iv) La dépendance de la loi de ξ_ℓ vis à vis des changements d'échelle est affine, c'est à dire que ξ_ℓ a une fonction de répartition de la forme

$$\mathbb{P}(\xi_\ell \leq x) = F(a_\ell(x - b_\ell)).$$

Alors par un raisonnement analogue à ceux qui permettent d'établir le théorème des types extrêmes on obtient que nécessairement la fonction de répartition de la résistance de la barre $F(x) = \mathbb{P}(\xi_L \leq x)$ est de la forme

$$F(x) = 1 - \exp\{-a[(x - x_0)^+]^\alpha\} \qquad \alpha, a > 0,\ x_0 \geq 0$$

c'est à dire une loi de Weibull (de minimum).

Ce raisonnement ne se limite pas à la résistance des matériaux, il se transpose avec des hypothèses analogues à l'étude des temps de première panne des systèmes en série et la loi de Weibull apparaît dans les travaux d'ingénieurs de nombreux domaines.

2 De l'usage de méthodes non justifiées à la fabrication de conclusions ex nihilo.

A - Nous commencerons par des commentaires sur la théorie de la résistance des barres car le problème est relativement plus simple et mieux circonscrit.

Le point fondamental est que la conclusion dépend de façon cruciale de l'hypothèse (iv) qui revient à dire

$$\xi_\ell \overset{Loi}{=} a(\ell)(\xi_1 - b(\ell))$$

et qui n'est fondée sur aucune explication véritable.

A propos de cette hypothèse Leadbetter et als [4] écrivent "This is an ad hoc notion [$\cdots$] Unlike i), ii), iii) it has no strict physical meaning, but is sometimes defended from the intuitively appealing principle that the simplest model is also the physically more realistic one". Ceci est d'autant plus important que sans l'hypothèse iv) on peut obtenir n'importe qu'elle loi pour la résistance de la barre :

Proposition 1 *Sans l'hypothèse iv) mais en conservant i) ii) et iii) on peut prendre n'importe quelle loi donnée à l'avance sur $\mathbb{R}_+$ pour la résistance de la barre.*

∎ Soit $h : \mathbb{R}_+ \to \bar{\mathbb{R}}_+$, croissante, continue à droite, telle que $\lim_{t\uparrow\infty} h(t) = +\infty$.

Prenons pour les fonctions de répartitions des sous-barres de longueur $\frac{L}{n}$ $G_n(t) = \left(\frac{1}{n}h(t)\right) \wedge 1$. La fonction de répartition de la barre constituée de n sous-barres est alors F_n donnée par

$$1 - F_n(t) = (1 - G_n(t))^n \longrightarrow_{n\uparrow\infty} e^{-h(t)}$$

Donc la résistance de la barre obtenue lorsque $n \uparrow \infty$ a pour fonction de répartition

$$F(t) = 1 - e^{-h(t)}$$

ce qui est, compte tenu des hypothèses sur h, toutes les fonctions de répartition sur $\mathbb{R}_+$. ∎

La statistique fournit des instruments beaucoup plus fins si l'on est dans une situation où le phénomène est régit par une loi de probabilité qui appartient à une famille connue. Ceci pousse les praticiens à rechercher des raisons a priori pour se limiter à telle ou telle famille de lois. Une fois cette famille choisie les arguments de l'inférence statistique, quoique de nature particulière s'établissent et peuvent être validés rigoureusement. Le point faible de ces méthodologies est souvent les justifications du choix de la famille. Par exemple les hypothèses gaussiennes fondées sur une application sous une forme ou une autre du théorème de limite centrale, cachent nécessairement une hypothèse quant aux moments d'ordre deux. De la même façon, la théorie de Weibull est invoquée souvent pour limiter les techniques statistiques à une famille à 3 paramètres particulière, mais on a vu que ceci ne fait qu'adopter implicitement une hypothèse douteuse sans laquelle tout est possible. La situation est tout à fait semblable à celle que dénonce William Feller [5] à propos de la loi logistique :

"The *logistic distribution* function

$$F(t) = \frac{1}{1 + e^{-\alpha t - \beta'}}$$

may serve as a warning. An unbelievably huge literature tried to establish a transcendental "law of logistic growth" : measured in appropiate units, practically all growth processes were supposed to be represented by a function of the form (4.10) with t representing time. Lengthy tables, complete with chi-square tests, supported this thesis for human populations, for bacterial colonies, development of railroads, etc. Both height *and* weight of plants and animals were found to follow the logistic law even though it is theoretically clear that these two variables cannot be subject to the same distribution. Laboratory experiments on bacteria showed that not even systematic disturbances can produce other results. Population theory relied on logistic extrapolations (even though they were demonstrably unreliable). The only trouble with the theory is that not only the logistic distribution but also the normal, the Cauchy, and other distributions can fitted to the *same material with the same or better goodness of fit.* In this competition the logistic distribution plays no distinguished role whatever ; most contradictory theoretical models can be supported by the same observational material."

B - A propos des valeurs extrêmes, notons d'abord quelques remarques techniques.

1. Dans le théorème des types extrêmes, si la loi des (X_i) est discrète on a aucun des trois types mais une loi dégénérée.

2. Les vitesses des convergences dont fait état le théorème sont très variées suivant la loi des X_i.

3. Un point assez important est que le théorème des types fournit des descriptions asymptotiques incompatibles avec les changements d'échelle fortement non linéaires : prenons par exemple des (X_i) positives et posons $Y_i = \log(X_i)$.

Si on pose

$$M_n = \max_{i=1,\cdots,n} X_i \text{ et } M'_n = \max_{i=1,\cdots,n} Y_i$$

on a évidemment

$$M'_n = \log M_n \text{ et } M_n = \exp M'_n.$$

Cependant si

$$a_n M_n + b_n \overset{Loi}{\to} aM + b$$

et

$$a'_n M'_n + b'_n \overset{Loi}{\to} a'M' + b'$$

les lois de $aM + b$ et $a'M' + b'$ ne sont reliées par rien de simple en général car ce qu'on peut déduire de la seconde convergence c'est :

$$(M_n)^{a'_n} e^{b'_n} \overset{Loi}{\to} e^{a'M'+b'}$$

convergence qui ne relève pas du théorème des 3 types.

Il en résulte que l'utilisation des lois de valeurs extrêmes pour des phénomènes qui ne sont pas des grandeurs additives mais uniquement des grandeurs <u>repérables</u> pose des difficultés de principe quant au choix de l'échelle de mesure.

C - L'usage courant des lois de valeurs extrêmes.

Il est intéressant de dénoncer un usage courant des lois de valeurs extrêmes trop largement répandu dans des domaines aussi variés que :

- la prévision des crues et des inondations,

- la pollution en liaison avec les phénomènes météorologiques,

- la résistance des matériaux,

- la détection des valeurs aberrantes dans les données,

- les pannes des systèmes complexes,

- la durabilité des ouvrages,

- etc.

Cet usage peut être schématisé de la façon suivante :

1. Evaluation du type : D'après les informations dont on dispose sur la loi commune des X_i on cherche laquelle des trois lois de valeurs extrêmes est atteinte dans le théorème des types.

2. Calage des paramètres de la loi asymptotique ainsi trouvée (image affine d'une des trois lois) grâce aux extrêmes de l'échantillon lui-même.

3. Déduction par la loi ainsi calée de la probabilité d'événements rares (queue de loi).

Cette méthode est merveilleuse car elle fournit des informations là où il n'y en avait pas : la queue de la loi des (X_i) et les phénomènes rares. Evidemment, en fait, comme on l'a montré,le résultat de la méthode est très sensible à la queue de la loi des (X_i) précisément. Il est curieux que des ingénieurs (cf [6] par exemple) ne soient pas davantage génés par le côté magique de cette méthode, il est vrai que la science est à certains égards merveilleuse et que l'on s'y habitue.

3 Remarques sur l'épistémologie des phénomènes rares.

Les considérations précédentes permettent de dégager quelques réflexions sur le contexte scientifique et culturel des phénomènes rares notamment lorsqu'ils représentent des risques majeurs.

L'éventualité de diverses catastrophes, qui hante l'esprit humain, de façon récurrente tout au long de son histoire et notamment actuellement, crée une pression sociale très grande qui pénétre les institutions et les organisations pour que la science, qu'on appelle toujours à la rescousse lorsqu'il faut apprivoiser l'inconnu car elle est capable parfois de tisser des liens solides avec des chocs connus, fournisse le maximum de ce qu'elle peut donner.

Cette demande est si forte qu'elle conduit certains techniciens à utiliser la science comme tranquillisant social ou comme placebo, ce qui dans certains des domaines cités est une forme de technocratie qui n'est pas anodine. Ceci est rendu possible parce que les affirmations ainsi produites portent sur des événements, certes graves, mais aussi rares et sont par ce fait même peu réfutables de sorte que les fabricants de conclusions douteuses ne prennent aucun risque au sens épistémologique que Popper a donné à ce terme. Il est d'ailleurs légitime de s'interroger, comme l'ont fait beaucoup d'auteurs, sur le sens même du concept de probabilité lorsqu'on l'applique à des événements exceptionnels. D'un point de vue strictement mathématique l'attribution d'une probabilité à un tel événement suppose que la loi de probabilité soit très bien connue. Ce peut être le cas dans des problèmes de combinatoire, ou lorsque le hasard provient de phénomènes physiques précis (agitation thermique, mécanique quantique, etc).Mais les phénomènes rares de la nature (tempêtes, séismes, etc.) ou de la vie économique ont des lois de probabilité nécessairement mal connues.A la limite, les événements intrinsèquement uniques, peuvent être dans le domaine de l'incertain, sans être probabilisables : quelle était la probabilité d'apparition des dinosaures ? Quelle était la probabilité d'extinction de la branche de Néanderthal quelle est la probabilité que l'arithmétique de Peano soit contradictoire etc. Autant de questions qui se rattachent au mythe d'un univers régi par une probabilité et qu'on peut ranger sans perte dans le domaine de l'insignifiant. Historiquement la science a toujours profité des éclaircissements quant à ses propres limites, et les travaux mathématiques, physiques et économiques récents concernant les systèmes dynamiques, les jeux indéterminés et les évolutions chaotiques à quoi il conviendrait d'ajouter la logique mathématique (cf. par exemple (7), fournissent de nombreux exemples de modèles, parfois très simples, dont l'évolution reste, au moins partiellement inconnue mais l'étude pertinente et féconde ne relève pas de la théorie des probabilités ni de la statistique du moins pas pour l'essentiel.

Que conclure de constructif de ces réflexions sur les lois des valeurs extrêmes ? Un principe épistémologique se dégage : toute démarche attribuant une valeur numérique précise pour la probabilité d'un phénomène rare est suspecte, sauf si les lois physiques régissant le phénomène sont explicitement et exhaustivement connues. A quoi nous ajouterons une remarque afin de nous prémunir contre toute assimilation abusive : ceci ne signifie pas que l'usage de probabilités ou de concepts probabilistes soit pour autant à rejeter des règlements de construction, dans la mesure ou un tel langage peut permettre une expression plus fine de certaines règles de conception. Dire que l'ouvrage devra résister pendant un certain temps à une houle gaussienne de spectre donné avec une probabilité fixée revient à demander au concepteur de faire certains calculs qui peuvent l'astreindre de façon différente que les règles déterministes.

References

[1] M. FRECHET, Les mathématiques et le concept, *P.U.F.* (1955).

[2] J. GALAMBOS, The asymptotic theory of extreme order statistics. *J. Wiley and Sons* (1978).

[3] W. WEIBULL, A statistical theory of the strength of materials, *Ing. Vetenskaps Akad. Handl*, **151**, (1939). The phenomenou of rupture in solids, *Ing. Vetenskaps Akad. Handl*, **153**, (1939).

[4] M.R. LEADBETTER, G. LINDGREN, H. ROOTZEN, Extremes and related properties of random sequences and processes, *Springer*, (1983).

[5] W FELLER, An application to probability theory and its application, Vol II, *J. Wiley and Sons*, (1971),

[6] R.R. KUININSON, Applied extreme value statistics, *Battelle Press* (1985).

[7] N. BOULEAU, J.Y. GIRARD, A. LOUVEAU, Cinq conférences sur l'indécidabilité, *Presses ENPC*, (1983).