Hasard, échelle et structure

par Georges MATHERON

Ingénieur en chef des Mines
Centre de Morphologie mathématique
Ecole nationale supérieure des Mines de Paris

Il n'est certainement pas question de résumer en quelques pages la théorie moderne des probabilités, ni de dégager la signification épistémologique des succès qu'elle a rencontrés dans des domaines d'applications très variés. Pourtant, quelques mots d'introduction à l'intention du lecteur cultivé mais non spécialiste ne seront peut-être pas inutiles : au risque de me perdre dans le marécage des lieux communs, j'évoquerai trois thèmes : hasard, échelle et structure.

HASARD

Dans le langage courant, le mot « hasard » évoque simplement ce qui n'est ni voulu ni prévu par l'homme (« je l'ai rencontré par hasard ») ou même simplement, par le sujet qui parle (« ce jour-là, par hasard, il y eut une éclipse »). Parfois aussi, comme dans le titre d'un ouvrage célèbre, on oppose « le Hasard et la Nécessité », mais il s'agit là d'une opposition toute relative : la nécessité est un hasard prévisible, et le hasard une nécessité aveugle. De fait, dans le livre de **J. Monod,** le hasard désigne simplement des mécanismes qui ne relèvent pas des disciplines biologiques : il peut s'agir des fameux rayons cosmiques qui provoquent des mutations « fortuites », mais, aussi bien, des lentes modifications climatiques qui, à diverses époques géologiques, ont entraîné un renouvellement profond de la faune et de la flore archaïques.

Ainsi, lorsqu'on a dit d'un phénomène qu'il est fortuit, on n'a en réalité rien énoncé du tout à son sujet, sinon tout au plus que l'on ignore les mécanismes qui l'ont engendré. Comme l'écrivait Emile Borel, « la caractéristique des phénomènes que nous appelons fortuits ou dûs au hasard est de dépendre de causes trop complexes pour que nous puissions les connaître toutes et les analyser ». Ce concept est donc dépourvu de contenu positif et est impropre à un usage scientifique. On croit parfois améliorer les choses en disant que le hasard résulte de « l'interférence de deux séries causales indépendantes ». Mais, si les mots sont pris dans leur sens usuel, cette définition est plutôt contradictoire : si deux séries causales (si tant est qu'il en existe dans un monde où l'interaction semble la règle) se trouvent interférer c'est qu'apparemment elles n'étaient pas indépendantes. En fait, dans cette définition, l'indépendance des séries causales est entendue, implicitement, en un sens déjà probabiliste : mais alors il y a, semble-t-il, pétition de principe, la définition faisant déjà appel au défini.

En réalité, c'est, semble-t-il, par abus de langage, que l'on parle de « phénomènes aléatoires ». Cet adjectif ne désigne pas une propriété de la réalité elle-même, mais seulement une caractéristique du modèle mathématique que l'on utilise pour décrire cette réalité et agir sur elle. Au fond, le mot « aléatoire » désigne une attitude, ou même une **décision** épistémologique. Dire d'un phénomène qu'il est aléatoire revient à **décider** qu'on l'étudiera au moyen de méthodes et de techniques dérivées de la théorie des probabilités. De ce point de vue, n'importe quel phénomène peut être considéré comme aléatoire : Il peut s'agir de la pièce de monnaie traditionnelle que l'on jette en l'air, des 10 puissances 27 particules que l'on lâche dans une enceinte, ou encore des éclipses de lune ou de soleil. Le seul problème qui se pose réellement est de savoir si cette décision épistémologique est la meilleure possible, et cela c'est à l'usage seulement que nous l'apprendrons. Dans le cas des deux premiers exemples (jeu de pile ou face et mécanique statistique) il se trouve que cette décision est effectivement la meilleure possible. Dans le cas des éclipses, on peut, et ce n'est pas absurde, étudier par des techniques probabilistes la chronique des observations passées, et en déduire des renseignements déjà assez précis, permettant même une certaine forme de prévision. Mais, naturellement, ces renseignements sont beaucoup moins précis que ceux que l'on peut déduire de la mécanique céleste, et il y a tout intérêt, dans le cas des éclipes, à laisser les astronomes travailler selon leurs propres méthodes.

Tout ceci est fort bien résumé par **B. Mandelbrot** (« les objets fractals, forme, hasard et dimension », Flammarion, 1975) : « Nous n'invoquerons le hasard, tel que le calcul des probabilités nous apprend à le manipuler, que parce que c'est le seul modèle dont dispose celui qui cherche à saisir l'inconnu et l'incontrôlable, modèle, fort heureusement, à la fois extraordinairement puissant et bien commode ». Il resterait, bien sûr, à s'interroger sur les « raisons (profondes, variées, et au fond encore mal connues) qui font que très souvent le résultat d'opérations déterministes mime l'aléatoire, tel que le décrit le calcul des probabilités ». Mais ceci est une autre histoire...

STRUCTURE

L'aléatoire n'est nullement synonyme de chaos absolu, et l'on peut, sans paradoxe, parler de « structures aléatoires ». Le rapprochement de ces deux termes ne surprendra ni le naturaliste, ni l'économiste, ni le spécialiste des sciences de l'homme. Un géologue, par exemple, a constamment affaire à des phénomènes dont la structure en grand se conforme à des lois ou à des modèles, sur lesquels il peut énoncer des propositions assez précises et en tirer certaines conclusions, alors que les détails, le comportement local du phénomène lui échappent et lui restent à peu près imprévisibles : le même phénomène présente ainsi deux aspects apparemment contradictoires, une structure en grand, susceptible d'être connue, coexistant avec une extrême irrégularité de détail empêchant toute prévision précise — bref, un aspect aléatoire et un aspect structuré.

Ces deux aspects sont, en réalité, en interaction étroite et profonde, et c'est d'un point de vue synthétique qu'il convient d'en aborder l'étude. S'ils nous semblent s'opposer l'un à l'autre et se contredire, c'est que nous avons en tête deux représentations extrêmes, presque caricaturales : d'une part, celle d'un « déterminisme » conçu comme une évolution obligatoirement très régulière et continue dans l'espace ou le temps, de l'autre celle d'un phénomène complètement chaotique et purement aléatoire (au sens que l'on donne trop souvent à ce mot dans la statistique élémentaire, où l'on a pris l'habitude de ne manipuler que des variables indépendantes les unes des autres).

Entre ces deux termes extrêmes, il existe en réalité toute une gamme d'états intermédiaires, qui possèdent à la fois et synthétiquement les caractères de l'aléatoire et du structural. Avec la théorie des processus stochastiques, des fonctions aléatoires, des ensembles aléatoires, etc., il se trouve, et c'est fort heureux, que le calcul des probabilités nous fournit un arsenal conceptuel et des techniques mathématiques qui permettent de représenter adéquatement ces états intermédiaires et de les traiter efficacement.

Si (pour abréger) nous définissons la structure d'un phénomène, ou d'un système, etc., comme l'ensemble des relations existant entre les éléments, ou les parties constitutives de ce système, on voit ce qu'il faut entendre par « structure aléatoire » : d'après ce qui précède, dire que l'on considère un système ou un phénomène donné comme une réalisation d'une fonction, ou d'un ensemble aléatoire signifie simplement que l'on décide d'exprimer en termes probabilistes les relations existant entre les éléments ou les parties de ce système.

ECHELLE

On a dit que « l'échelle crée le phénomène ». De ce point de vue, et bien plus qu'une science des « lois du hasard », le calcul des probabilités apparaît souvent comme un instrument conceptuel permettant de passer d'une échelle à l'échelle supérieure. La mécanique statistique est exemplaire à cet égard : des notions comme celles de pression, de température, d'entropie, etc., qui ont une signification purement probabiliste lorsqu'on les envisage à l'échelle microscopique, prennent à l'échelle macroscopique un sens parfaitement déterministe et représentent des propriétés physiques réelles de la matière. De même, en morphologie mathématique, on représentera, par exemple, un milieu poreux comme un ensemble aléatoire, et on s'intéressera à la probabilité pour qu'un point donné soit dans les pores ou pour qu'une sphère de rayon donné rencontre les grains, etc. Ces relations, envisagées d'un point de vue purement probabiliste au niveau granulométrique, acquièrent une signification parfaitement déterministe lorsque l'on passe au niveau macroscopique, et conduisent à des concepts physiques précis (porosité, surface spécifique, granulométries, etc).

Il y a ainsi une classe de phénomènes pour lesquels l'aléatoire n'émerge pas à l'échelle supérieure : le chaos imprévisible qui semble régner au niveau microscopique fait place, lorsque l'on s'élève au point de vue macroscopique, à un ordre et à une régularité « déterministes » de style plus classique. Les modèles probabilistes correspondants possèdent la propriété connue sous le nom d'ergodicité : sous certaines conditions, les valeurs moyennes prises sur des intervalles de temps de plus en plus grands, ou des zones de plus en plus grandes de l'espace, convergent vers des limites non aléatoires, égales aux espérances mathématiques des variables correspondantes. Mais cette circonstance n'est pas générale, loin de là. Il existe des phénomènes qui présentent le même degré d'irrégularité à toutes les échelles, et pour lesquels le caractère aléatoire émerge sans changement au niveau « macroscopique ». On en trouvera de nombreux et frappants exemples dans l'article de B. Mandelbrot, et dans son livre déjà cité. Les modèles probabilistes qu'il convient alors d'utiliser ne sont évidemment plus ergodiques.

Il ne faut d'ailleurs pas ériger cette opposition en absolu. Selon le point de vue auquel on se place, et le type de problèmes que l'on cherche à résoudre, un même phénomène peut fort bien relever de modèles probabilistes différents, les uns ergodiques, les autres non. Dans beaucoup d'applications (et c'est le cas dans la majorité des articles qui suivent) on se contente d'une hypothèse approchée de type « stationnaire et ergodique ». La raison principale en est sans doute que c'est dans le cadre d'une hypothèse de ce type que le problème de l'inférence statistique (l'estimation des paramètres numériques dont le modèle dépend) est le plus facile à résoudre. En fait, une analyse plus fine (dont on verra un exemple à propos des méthodes d'interpolation optimale) montre souvent que l'on peut en grande partie s'affranchir de cette hypothèse. C'est ainsi par exemple qu'en Géostatistique minière on utilise couramment des modèles (caractérisés par un « variogramme » en $|h|^{\lambda}$) qui possèdent la propriété d'auto-homothétie évoquée ci-dessus (même degré d'ordre ou de désordre à toutes les échelles). Ces modèles ne sont ni stationnaires ni ergodiques. Ils n'en permettent pas moins de résoudre très efficacement les problèmes concrets qui se posent au mineur.

CONCLUSION

S'il faut une conclusion, nous pouvons dire sous forme à peine paradoxale que le calcul des probabilités n'a pas grand chose à voir avec le hasard (à moins que l'on n'accepte de définir comme aléatoires les phénomènes auxquels, justement, le calcul des probabilités s'applique). Il se présente plutôt comme une méthode d'analyse puissante, permettant de dégager des traits structuraux masqués par un chaos apparent, et d'étudier leurs propriétés d'émergence lorsque l'on passe d'une échelle à l'échelle supérieure. Son statut — purement instrumental — est le même que celui de n'importe quelle autre théorie mathématique appliquée dans le champ des disciplines expérimentales : il ne nous révèle en aucune façon les derniers secrets de l'univers, mais nous permet d'agir avec efficacité dans tel ou tel secteur de la réalité où les autres méthodes, apparemment, échouent.

Petit lexique de probabilité

Pour faciliter la lecture de ce numéro, la rédaction des Annales a jugé utile de rappeler quelques-unes des définitions les plus usuelles du calcul des probabilités.

ESPACE PROBABILISABLE, OU MESURABLE :

C'est un couple $(\Omega, \mathcal{A})$ constitué d'un ensemble quelconque Ω et d'une σ-algèbre $\mathcal{A}$ (famille $\mathcal{A}$ de sous-ensembles $A \subset \Omega$, stable pour l'union et l'intersection dénombrables et pour la complémentation). Les $A \in \mathcal{A}$ sont souvent appelés **évènements.** Concrètement : Ω sera par exemple l'ensemble des états possibles d'un système physique, et $\mathcal{A}$ l'ensemble des propositions vérifiables que l'on peut énoncer à son sujet. Une proposition A est identifiée à l'ensemble des états ω du système pour lesquels A est vraie, de sorte que

$\mathcal{A}$ est une famille de parties de Ω. Aux opérations logiques « ou », « et », « non » effectuées sur des propositions vérifiables correspondent la réunion, l'intersection et la complémentation sur les ensembles correspondants dans $\mathcal{A}$.

ESPACE PROBABILISE : C'est un triplet (Ω, $\mathcal{A}$, P), où (Ω, $\mathcal{A}$) est un espace mesurable et P une **probabilité** sur $\mathcal{A}$ (application de $\mathcal{A}$ sur l'intervalle (0,1) telle que $P(\Omega) = 1$ et $P(UA_n) = \Sigma\ P(A_n)$ pour toute suite d'évènements disjoints dans $\mathcal{A}$). Un évènement de probabilité nulle est dit **négligeable.** Une propriété vraie en tout $\omega \in \Omega$ sauf au plus sur un ensemble négligeable est dite vraie **presque partout pour P.**

VARIABLE ALEATOIRE : C'est une fonction Y : $\Omega \rightarrow R$ définie sur Ω et **mesurable** pour la σ-algèbre $\mathcal{A}$ (l'ensemble des ω tels que $Y(\omega) \leqslant y$ est un évènement de $\mathcal{A}$). La probabilité $F(y) = P(\{Y \leqslant y\})$ est la **fonction de répartition** de la variable aléatoire Y. Concrètement : Y sera par exemple une caractéristique physique dépendant de l'état ω du système et susceptible d'être mesurée.

On peut aussi considérer simultanément deux (ou plusieurs) variable aléatoires Y, Z... La fonction de répartition de la loi à 2 variables (Y, Z) est $F(y, z) = P(Y \leqslant y \text{ et } Z \leqslant z)$. Si $F(y, z) = F_1(y)\ F_2(z)$ (produit des fonctions de répartition de Y et Z prises séparément) Y et Z sont indépendantes.

ESPERANCE d'une variable aléatoire Y. : C'est l'expression $E(Y) = \int Y(\omega)\ P(d\omega)$ (intégrale de Lebesgue : elle existe si et seulement si $E(|Y|) < \infty$). Elle peut aussi s'écrire $E(Y) = \int y\ F(dy)$ (intégrale de Stieltjes relativement à la fonction de répartition). Pour une loi discrète (attribuant la probabilité f_n à la valeur y_n) $E(Y) = \Sigma\ f_n\ y_n$. Pour une loi absolument continue (admettant une densité f(y), $E(Y) = \int y\ f(y)\ dy$.

LE MOMENT D'ORDRE p d'une variable aléatoire Y est l'espérance $E(Y^p)$ de Y^p.

LA VARIANCE est $\sigma^2 = E[(Y - E(Y))^2] = E(Y^2) - (E(Y))^2$. Si l'on considère plusieurs variables Y_1, Y_2,... la covariance de Y_i et Y_j est $\sigma_{ij} = E(Y_i\ Y_j) - E(Y_i)\ E(Y_j)$. La variance d'une combinaison linéaire $\Sigma\ \lambda_i\ Y_i$ est la somme double $\sum_{ij} \lambda_i\ \lambda_j\ \sigma_{ij}$.

LA FONCTION CARACTERISTIQUE $\Phi(u)$ d'une variable aléatoire Y est la transformée de Fourier $\Phi\ (u) = E(e^{iuY}) = \int e^{iuy}\ F(dy)$ de la loi F.

FONCTION ALEATOIRE : C'est une famille Y(x) de variables aléatoires où x décrit un espace donné T (si T = R est interprété comme un temps, on parle plutôt de **processus stochastique,** et la famille Y(t), $t \in T$ décrit l'évolution aléatoire d'un phénomène au cours du temps. Mais T peut aussi être un espace euclidien à 2 ou 3 dimensions, et Y(x) décrit alors la régionalisation aléatoire d'un phénomène dans l'espace). Pour chaque $x \in T$, $Y(x) = Y(x\,;\omega)$ est une variable aléatoire, c'est-à-dire une fonction mesurable sur Ω : $Y(x, \omega)$ est donc une fonction (ordinaire) sur l'espace produit T x Ω. A $\omega = \omega_0$ fixé, $x \rightarrow Y(x, \omega_0)$ est une fonction (ordinaire) sur T, d'ailleurs en général assez chaotique, appelée **réalisation** de la fonction aléatoire (ou **trajectoire** du processus si T = R représente le temps).

Pour deux points d'appui x et $y \in T$, les variables Y(x) et Y(y) ne sont pas, en général, indépendantes. Leur **covariance** σ (x, y) est une fonction sur T x T appelée covariance de la fonction aléatoire. En général elle dépend séparément de x et de y. Lorsqu'elle ne dépend que de la différence y−x, soit $\sigma(x, x+h) = C(h)$ pour une fonction C d'un seul argument h, on dit que la fonction aléatoire Y(x) est **stationnaire** d'ordre 2.

ESPERANCE CONDITIONNELLE : Si Y, Z_1, ...Z_n sont des variables aléatoires, l'espérance conditionnelle de Y en $(Z_1,...Z_n)$ est une fonction mesurable $E(Y|Z_1,...Z_n) = f(Z_1,...Z_n)$ telle que $E[Y\ g(Z_1,...Z_n)] = E[f(Z_1,...Z_n)\ g(Z_1,...Z_n)]$ pour toute fonction mesurable g de n variables. Il existe aussi une **loi conditionnelle** $F(z_1,...z_n\,;\ dy)$ de Y à $Z_1 = z_1,...\ Z_n = z_n$ fixés telle que

$$E[h(Y)|Z_1,...Z_n] = \int h(y)\ F(Z_1,...Z_n\,;\ dy)$$

pour toute fonction mesurable h.

Plus généralement, on définit l'espérance conditionnelle en termes de σ-algèbres : soit $\mathcal{B} \subset \mathcal{A}$ une σ-algèbre moins riche que $\mathcal{A}$ (par exemple, $\mathcal{A}$ représentera les évènements observables à un niveau microscopique, et $\mathcal{B}$ ceux d'entre eux, beaucoup moins nombreux, qui restent observables au niveau macroscopique). Si $Y(\omega)$ est une variable aléatoire (une fonction mesurable) pour $\mathcal{A}$, elle ne l'est pas, en général, pour $\mathcal{B}$ (elle représente une grandeur non directement observable au niveau macroscopique). L'espérance conditionnelle $Z = E(Y|\mathcal{B})$ relativement à la σ-algèbre $\mathcal{B}$ est la meilleure approximation de Y par une fonction mesurable pour $\mathcal{B}$, caractérisée par $E(Z\ Z') = E(Y\ Z')$ pour toute variable aléatoire Z' mesurable pour $\mathcal{B}$: cette notion est sans doute la plus importante de la théorie des probabilités. Elle se prête remarquablement bien à la formalisation des changements d'échelle et à l'étude des propriétés **d'émergence.**