PRINCIPES DE L'APPRECIATION QUANTITATIVE PROBABILISTE DES RISQUES

R. Pouillot[1], M. Sanaa[2] et Barbara Dufour[1]

RESUME : *L'appréciation probabiliste (encore appelée stochastique) des risques est une procédure permettant de prendre en compte et d'évaluer la variabilité du risque et l'incertitude liée à son estimation. Ce texte présente les bases théoriques de cette méthode et son intérêt comparé à l'appréciation ponctuelle (encore appelée déterministe) dans le domaine des risques liés à l'importation d'animaux vivants. La procédure générale de simulation utilisée, nommée procédure de « Monte-Carlo », est expliquée ainsi que l'interprétation des résultats obtenus. Enfin, un complément sur la modélisation de second ordre, permettant de séparer l'incertitude de la variabilité du risque, est exposé.*

SUMMARY : *The mathematical study of the risk probability (still called stochastic) is a procedure taking into account and estimating the variability of the risk and the uncertainty bound to its estimation. This paper presents the theoretical basis of this method and its interest compared with the punctual appreciation (still called determinist) in the field of the risks related to the import of alive animals. The general simulation, known as The Monte-Carlo evaluation, is explained as well as the interpretation of the obtained results. Finally, a further discussion on the second-class modelling, allowing to separate the uncertainty of the variation of the risk, is presented.*

Il existe plusieurs approches permettant l'analyse de scénarii de risque. On peut les classer en deux groupes : les analyses déterministes (encore appelées dans ce texte « analyses ponctuelles ») et les analyses stochastiques (encore appelée dans ce texte « analyses probabilistes »). L'objectif de ce texte est de présenter les bases théoriques de l'appréciation quantitative stochastique ou probabiliste des risques. La connaissance préalable des notions développées dans l'article de Toma (2002, ce numéro) sur l'analyse quantitative déterministe est nécessaire à sa compréhension : nous ne reviendrons notamment pas sur les notions de base telle que la définition du risque ou la description des différentes étapes de l'analyse. Dans un premier chapitre, nous comparerons l'approche de l'analyse quantitative déterministe et celle de l'analyse quantitative probabiliste. Nous développerons ensuite quelques aspects de la réalisation pratique de l'analyse quantitative probabiliste, puis la présentation et l'interprétation des résultats. Enfin, des compléments plus techniques sont proposés.

La théorie et la réalisation pratique d'analyses quantitatives probabilistes du risque n'est pas simple : elle demande une connaissance importante des théories probabilistes et statistiques. Le préalable minimal dans ce domaine est exposé dans l'article de Pouillot et Sanaa [2002, ce numéro].

Ce texte ne pourra être qu'une approche simplifiée. Le lecteur intéressé par des informations plus complètes pourra se référer par exemple à l'ouvrage de Vose [2000] ou de Murray [2002].

1 Afssa, 27-31 avenue du Général Leclerc, BP 19, F-94701 Maisons-Alfort Cedex, France, r.pouillot@afssa.fr.
2 ENVA, 7 avenue du général de Gaulle, F-94704 Maisons-Alfort cedex, France

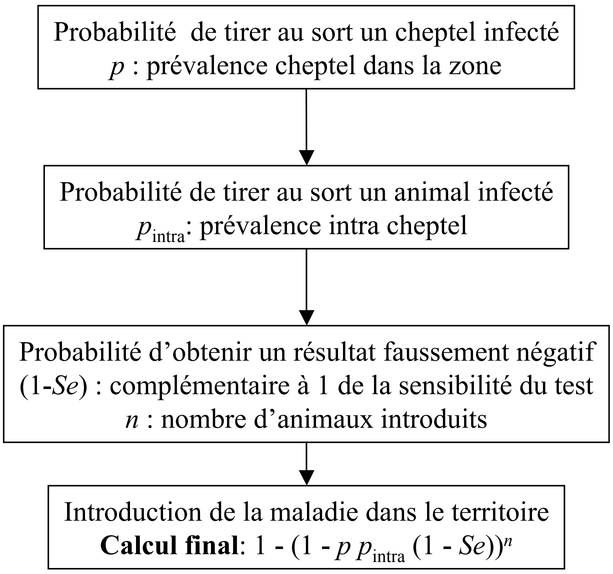
I - ESTIMATION PONCTUELLE ET ESTIMATION PROBABILISTE

Le principe de l'estimation ponctuelle du risque dans le domaine de l'importation des animaux vivants est le suivant [Toma, 2002] :

- on construit un scénario (encore appelé « arbre événementiel ») comprenant différentes phases décrivant les processus successifs nécessaires à l'introduction d'une maladie sur un territoire (*cf.* un exemple simple figure 1) ;
- à chaque paramètre utilisé dans le scénario, on associe une valeur : généralement la moyenne, mais parfois des valeurs particulières (par exemple le 95$^{\text{ème}}$ percentile ou le 99$^{\text{ème}}$ percentile) ou les valeurs extrêmes ;
- on estime finalement le risque moyen ou les risques extrêmes en utilisant les règles de calculs de probabilité.

FIGURE 1

Présentation d'un scénario simple d'introduction d'une maladie sur un territoire. Chaque cadre présente une étape et le(s) paramètre(s) ponctuel(s) permettant de caractériser cette étape. Dans le cadre final, le risque d'introduction de la maladie est estimé à l'aide d'un calcul probabiliste [*cf.* Toma, 2002].



Cependant, une estimation ponctuelle du risque peut être insuffisante pour le gestionnaire :

- ainsi, le risque peut être **variable** d'une année à l'autre, d'une région à l'autre, d'un cheptel à l'autre, et l'existence de cette variabilité doit être portée à la connaissance du gestionnaire avant qu'il prenne sa décision. Par exemple, si un pays achète entre 1 et 1 000 bovins par an en provenance d'une zone infectée, il est évident que le risque sera variable d'une année à l'autre. Le gestionnaire pourrait être intéressé par une mesure de cette variabilité ;
- d'autre part, le risque peut être estimé avec plus ou moins de précision : par exemple, si le pays exportateur affirme qu'il est indemne sur la base d'un résultat négatif obtenu sur un petit échantillon d'animaux, le caractère « indemne » de la zone est incertain ; le risque réel encouru lors de l'importation est donc incertain. Le gestionnaire peut être intéressé par une mesure de cette **incertitude**.

Ces deux notions de « variabilité » et d'« incertitude » [Anderson et Hattis, 1999] sont essentielles dans l'appréciation du risque (encadré A). L'absence de prise en compte de ces notions peut gravement compromettre la solidité de l'analyse (encadré B). L'analyse quantitative du risque probabiliste permet précisément de quantifier la variabilité du risque et l'incertitude autour de son estimation.

Encadré A : Variabilité et incertitude

Il est essentiel de comprendre et distinguer précisément les deux notions complexes de variabilité et d'incertitude :

- la **variabilité** d'un paramètre (prise dans le sens de l'analyse de risque) est la variation naturelle du paramètre dans la population. Cette variation peut refléter ce que l'on appelle la variabilité naturelle (exemple : la taille d'un troupeau à l'autre est variable, la taille d'un animal à l'autre est variable), ou les fluctuations liées à un phénomène aléatoire (exemple : le nombre exact d'animaux infectés tirés au sort d'une population est variable). La population étudiée (animaux ou élevages) peut également être segmentée en plusieurs sous-populations avec des niveaux de risque variables dans chacune des sous-populations ;
- l'**incertitude** autour d'un paramètre est une variation possible du paramètre résultant de la méconnaissance de la valeur exacte du paramètre. Par exemple, la prévalence exacte dans une population est incertaine.

Il peut être intéressant de considérer la variabilité comme une variation irréductible d'un paramètre dans une population : quel que soit le nombre de données collectées, l'effectif des cheptels sera par exemple toujours variable. Au contraire, l'incertitude est réductible par l'apport de données complémentaires : l'application d'un test parfait sur l'ensemble des animaux d'une zone pourrait permettre de connaître avec certitude la valeur exacte de la prévalence animal dans la population ; l'utilisation d'un échantillon ou l'usage d'un test imparfait entraîne une certaine incertitude sur l'estimation de la prévalence réelle, liée aux fluctuations d'échantillonnage.

La complexité s'accroît encore lorsqu'un paramètre est à la fois variable et incertain : par exemple, dans notre domaine d'intérêt, la prévalence dans une zone est variable d'une année à l'autre, et incertaine en raison de l'inévitable sondage à effectuer dans la population pour la mesurer.

Une meilleure description de la variabilité permettra d'estimer des risques spécifiques à certaines sous-populations et de proposer des options de gestion spécifiques des risques que chacune représente. Lorsque l'incertitude sur un paramètre influence le résultat final, cela permet d'orienter les activités de recherche ultérieures afin de diminuer celle-ci. On pourra également proposer des options de gestion portant sur des paramètres qui ne sont pas importants, mais au moins incertains.

Encadré B : Exemples de l'intérêt de la prise en compte de la variabilité et de l'incertitude

Variabilité :

On désire évaluer le risque lié à l'introduction d'une maladie contagieuse dans les cheptels. On évalue le risque moyen d'introduction à 0,01%, et on considère que ce risque est acceptable. Cette mesure moyenne peut, en fait, refléter plusieurs configurations :

- le risque de 0,01% est identique pour tous les cheptels de la zone ;
- le risque est variable d'un cheptel à l'autre : le risque est en fait de 0% pour 99,99% des cheptels et de 100% pour 0,01% des cheptels. Dans ce cas, il y aura, de manière sûre, introduction de la maladie dans certains cheptels. Ce risque est alors probablement inacceptable, contrairement à ce que l'estimation ponctuelle laissait supposer ;
- toutes les autres configurations intermédiaires de variabilité du risque entre les cheptels sont possibles.

Il est évident que l'attitude du gestionnaire pourra être différente selon la configuration. L'utilisation de l'estimation de la moyenne du risque réduit l'information.

Un exemple illustratif peut être proposé : un expert évalue la profondeur moyenne d'un lac à 1 m, et autorise la baignade. Ce chiffre peut ne pas représenter une certaine variabilité de cette profondeur : la profondeur d'eau pouvant être généralement de 50 cm, mais avec la présence ponctuelle de « trous » de 3 m de fond… à l'origine de nombreuses noyades.

Incertitude :

Une estimation du risque annuel d'importation de la fièvre aphteuse dans un pays est de 1%, signifiant que la fièvre aphteuse peut être introduite en moyenne une fois tous les 100 ans. Si la prise en compte de l'incertitude des paramètres d'entrée dans le modèle permet de préciser que cette valeur de 1% est la plus probable, mais que la « vraie » valeur peut se situer raisonnablement entre 50% (1 introduction en moyenne tous les 2 ans) et 1 pour mille (1 introduction en moyenne tous les 1000 ans), le gestionnaire pourra avoir une attitude tout à fait différente.

Pour reprendre l'exemple précédent, si notre expert propose une estimation la plus probable de la moyenne de la profondeur du lac à 1 m, mais que ses données ne permettent pas de préciser que la vraie moyenne est comprise entre 10 cm et 5 m, on considérera avec réserve ses conclusions.

Le principe général de l'estimation probabiliste des risques est très proche de celui de l'estimation ponctuelle des risques, voire de l'estimation qualitative des risques : il comprend notamment la première phase commune de construction du scénario d'introduction de la maladie sur le territoire.

Pour chaque paramètre du modèle ainsi créé :

- l'analyse qualitative utilisera une appréhension qualitative du paramètre ;
- l'analyse déterministe utilisera une valeur particulière (généralement la moyenne) ;
- l'analyse probabiliste utilisera **non pas une valeur, mais une distribution de probabilité** de cette valeur. Cette distribution de valeurs pourra refléter soit **l'incertitude**, soit la **variabilité** du paramètre, soit une combinaison des deux.

Alors que le résultat d'une analyse qualitative est une appréhension qualitative du risque, que le résultat d'une analyse ponctuelle est une estimation ponctuelle du risque, le résultat d'une analyse probabiliste sera une **distribution de probabilité** du risque. L'incertitude et la variabilité des paramètres seront incorporées et leur influence propagée dans le modèle : la loi de distribution du risque finalement estimée reflétera sa variabilité et l'incertitude quant à son estimation.

Outre l'absence de prise en compte de la variabilité et de l'incertitude, d'autres difficultés de l'estimation ponctuelle du risque, comparée à l'estimation probabiliste du risque, sont proposées dans l'encadré C.

Encadré C : Difficultés liées à l'estimation par la méthode ponctuelle

Estimation du risque moyen

Comme nous l'avons vu dans l'article sur les généralités [Pouillot et Sanaa, 2002], il est parfois possible de combiner les moyennes : par exemple, la moyenne de la somme de deux variables indépendantes est égale à la somme de chacune des moyennes. En revanche, il n'est pas possible dans certains cas de faire un calcul si simple : la moyenne du rapport de deux variables n'est pas égale au rapport des moyennes ; la moyenne d'une variable élevée au carré n'est pas égale au carré de la moyenne de la variable, … En fait, on rencontre les limites de l'estimation ponctuelle lorsque le modèle d'appréciation des risques n'est pas linéaire, ou encore en présence de corrélations entre les variables. La méthode probabiliste (par l'utilisation de la méthode de Monte-Carlo) permet de passer outre ces limites.

Estimation des cas extrêmes

Un type d'analyse ponctuelle permettrait d'obtenir une estimation des cas extrêmes : disposant de paramètres moyens, on réalise le calcul de l'estimation du risque moyen. Puis pour chacun des paramètres, on effectue une estimation d'une valeur « limite minimale ». A l'aide de ces valeurs, on effectue le calcul d'un « risque minimal » possible. Enfin, pour chacun des paramètres, on effectue une estimation d'une valeur « limite maximale ». A l'aide de ces valeurs, on effectue le calcul d'un « risque maximal » possible [Toma, 2002].

Ce type d'estimation, fréquemment rencontré, permet de donner une idée de l'incertitude et de la variabilité de l'estimation. Cependant, il n'est pas fondé sur une base statistique :

- les « raccourcis » rapides effectués lors de l'analyse par la multiplication ou l'addition de facteurs ne sont en théorie valides que pour certains calculs effectués sur la moyenne seulement, et non sur d'autres valeurs (*cf.* point précédent de cet encadré) ;
- les paramètres « minimaux » et « maximaux » correspondent à des valeurs pas forcément bien définies : s'agit il des $5^{ème}$ et $95^{ème}$ percentiles de la loi de distribution des paramètres? s'agit-il des véritables extrêmes ? s'il s'agit des $5^{ème}$ et $95^{ème}$ percentiles, il est important de noter que le risque estimé ne correspondra pas aux $5^{ème}$ et $95^{ème}$ percentiles du risque, mais à une « fourchette » non définie, puisque les calculs probabilistes ne sont applicables que sur les moyennes (le $95^{ème}$ percentile d'une somme de variables n'est pas égal à la somme des $95^{èmes}$ percentiles de chacune des variables) ;
- enfin, l'incertitude (par exemple : la prévalence réelle dans la zone) et la variabilité (par exemple : le nombre d'animaux achetés par an) sont considérées mais ne sont pas séparées.

II - PRINCIPE DE LA MISE EN ŒUVRE DE L'APPRECIATION QUANTITATIVE PROBABILISTE DES RISQUES

1. SPECIFICATION PRECISE DE LA QUESTION

La première étape, essentielle bien que souvent négligée, consiste à poser précisément la question. Cette étape permettra le développement plus aisé de l'étape suivante : la construction du modèle.

Ainsi, une analyse probabiliste de risque ne peut se limiter en pratique qu'à un risque restreint : dans le domaine des risques liés à l'importation, ce risque ne pourra, dans la plupart des cas, être évalué que :

- pour une maladie ou un agent pathogène donné ;
- pour une espèce donnée ;
- pour des animaux provenant d'un pays donné ;
- à un instant donné ;
- pour un type d'animaux (ou de produits) donné.

2. CONSTRUCTION DU MODELE

La construction du modèle pour une estimation probabiliste est globalement similaire à celle effectuée pour une estimation ponctuelle : elle nécessite une analyse de la situation et une disposition en ordre chronologique des différentes séquences pouvant conduire, par exemple, à l'introduction d'un agent pathogène dans un territoire indemne. En pratique, le scénario d'importation devra être précisé le mieux possible. La construction de cet arbre événementiel devra permettre [Murray, 2002] :

- d'identifier clairement les différentes séquences permettant d'obtenir l'événement non désiré ;
- d'identifier l'ensemble des paramètres nécessaires à la construction de chacune des étapes ;
- d'identifier le caractère « fixe », « variable » ou « incertain » de chacun des paramètres ;
- d'identifier les informations requises pour l'estimation de chaque paramètre.

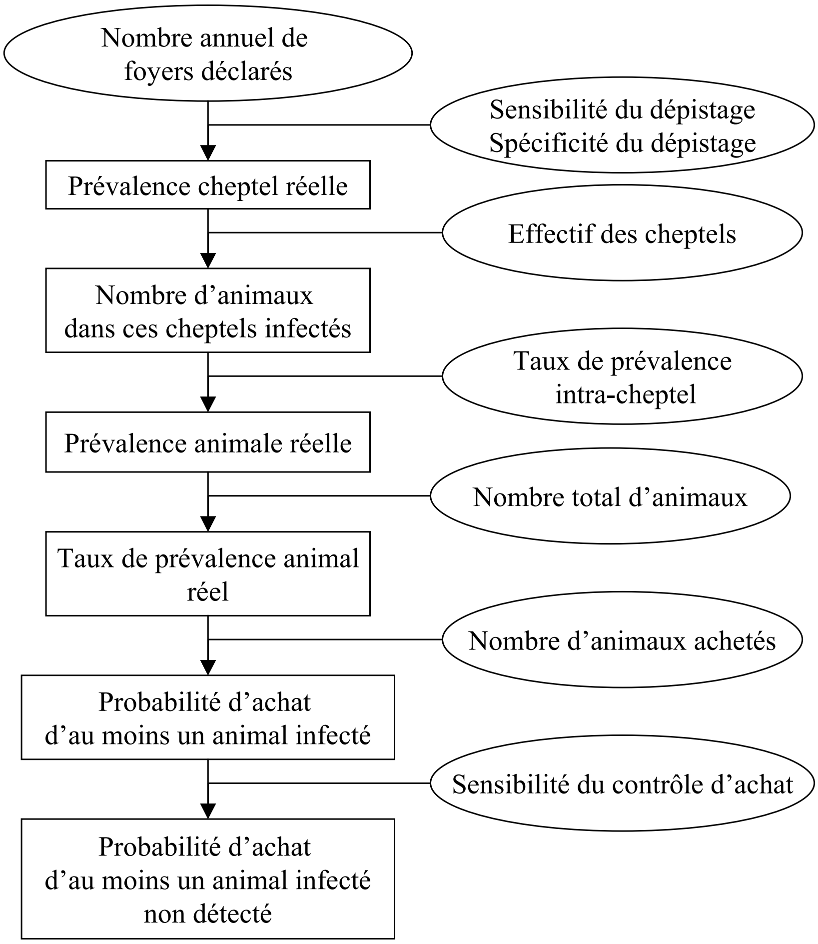
Cette analyse permettra en outre à l'évaluateur de clarifier les idées pour la spécification du risque. Fourni avec les résultats, il permettra une communication plus aisée sur le modèle.

La figure 2 représente un modèle assez simple pour la construction d'une évaluation quantitative du risque lié à l'importation d'animaux vivants. On constate que, pour un modèle aussi simple, le nombre de paramètres nécessitant des données (représentées dans la figure par un ovale) est déjà important. On pourra classer les paramètres en :

- paramètres considérés comme « fixes » : nombre total d'animaux dans le pays d'importation ;
- paramètres considérés comme « variables » : nombre annuel de foyers déclarés (variable d'une année à l'autre), effectif des cheptels (variable d'un cheptel à l'autre), taux de prévalence intra-cheptel (variable d'un cheptel à l'autre), nombre d'animaux achetés (variable d'une année à l'autre) ;
- paramètres considérés comme « incertains » : sensibilité du test de dépistage et du contrôle d'achat (dont on ne connaît qu'une estimation sur quelques animaux infectés).

FIGURE 2

**Exemple de scénario d'introduction d'animaux infectés.
Les ovales représentent des paramètres nécessitant l'obtention de données ;
les rectangles représentent des paramètres calculés.**



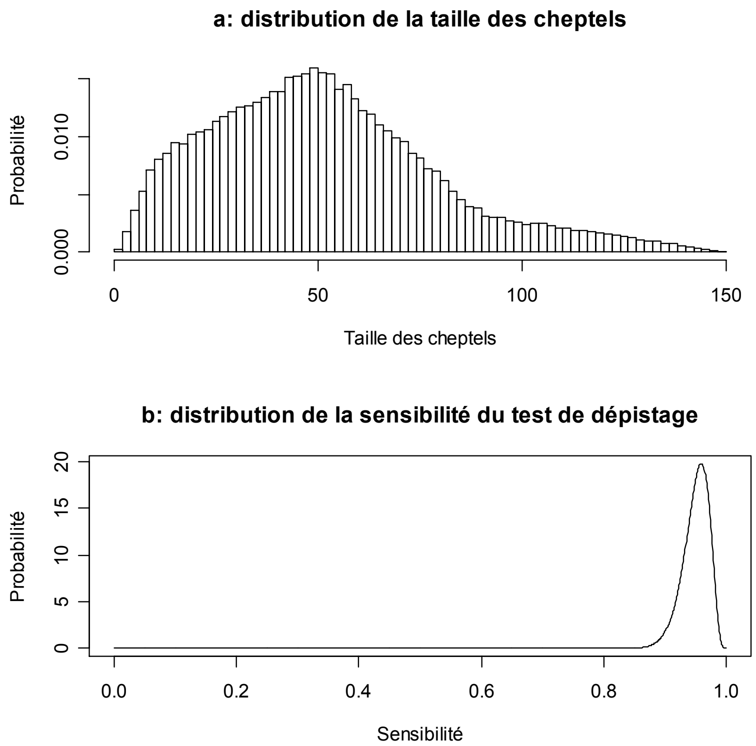
CHOIX DE LOIS DE DISTRIBUTION DES VARIABLES ENTRANT DANS LE MODELE

A chaque variable aléatoire (variable issue d'un tirage au sort reflétant soit une incertitude, soit une variabilité) utilisée dans le modèle, on associe non pas **une** valeur ponctuelle, mais **une distribution de valeurs possibles** associées à leur probabilité, appelée loi de probabilité [Pouillot et Sanaa, 2002].

Dans l'exemple précédent (figure 2), il sera donc nécessaire de choisir des lois de probabilité pour chacun des paramètres suivants : nombre annuel de foyers déclarés, effectif des cheptels, taux de prévalence intra-cheptel, nombre d'animaux achetés, sensibilité et spécificité du dépistage, sensibilité du contrôle d'achat. La figure 3a présente une loi de probabilité possible pour la taille des cheptels : la taille moyenne des cheptels est de 51 animaux. Cependant, l'effectif peut varier dans la population de 1 à 150, avec une probabilité plus élevée pour les cheptels de faible effectif (moins de 50 animaux) que pour les cheptels de plus fort effectif. La figure 3b présente une loi de probabilité possible pour la sensibilité du test de dépistage : la sensibilité de ce test est généralement estimée à 0,95. Cependant, il est possible de considérer que cette sensibilité n'est pas connue avec certitude : l'utilisation de cette loi de probabilité permettra de prendre en compte une possible sur-estimation ou sous-estimation de ce paramètre.

FIGURE 3

Exemple de distributions. (a) taille des cheptels : cette variable discrète est représentée sous la forme d'un histogramme ; (b) sensibilité du test de dépistage : cette variable continue est représentée sous la forme d'une courbe.



La principale difficulté de l'analyse quantitative des risques provient du choix de ces lois de densité pour chacune des variables utilisées dans le modèle d'estimation du risque. Toute loi de densité est caractérisée par sa forme et ses paramètres. Une bonne connaissance des théories statistiques et probabilistes est souvent nécessaire.

Derrière chaque variable aléatoire, on peut considérer qu'il existe un modèle expliquant les différentes réalisations des valeurs observées. Le choix de la loi de distribution dépendra de la nature de la variable et du processus, appelé « processus stochastique » sous-jacent. Ce processus est plus ou moins bien connu. Par exemple, le modèle décrivant le nombre d'animaux présentant une caractéristique particulière issus d'un tirage au sort au sein d'une population est bien connu [il s'agit du processus « binomial », Pouillot, 2002]. En revanche, le modèle permettant de choisir la loi de distribution de la taille des cheptels français est moins bien connu.

On peut, en fait, résumer la procédure de choix des lois de densité pour les quatre situations suivantes :

1. **le processus permettant d'observer la variable est un processus connu** : certaines lois de probabilité sont particulières, car elles correspondent à la réalisation d'un type d'expérience particulier, rencontré dans de nombreux domaines et généralisable. Ainsi, d'un point de vue probabiliste, il existe une certaine analogie que l'on tire au sort des animaux infectés dans une population comportant des animaux indemnes et infectés, des boules noires dans un sac contenant des boules noires et des boules blanches, ou des pièces défectueuses issues d'une machine produisant des pièces correctes et défectueuses. Ce type d'expériences, appelées « processus stochastiques », a des propriétés bien connues dans le domaine des probabilités. Si le processus d'obtention de la variable est un processus connu, il suffit d'appliquer la loi de distribution adéquate ;
2. **la théorie statistique permet de connaître la loi de distribution** : la théorie statistique permet de spécifier certaines lois de variables. L'exemple le plus connu est celui de la moyenne d'un

grand nombre d'observations. Le théorème central limite permet d'affirmer que, lorsque le nombre d'observations est grand, la moyenne des observations suit une loi normale. Outre la forme de cette loi, la théorie statistique permet de spécifier ses paramètres à partir d'un échantillon de valeurs observées ;

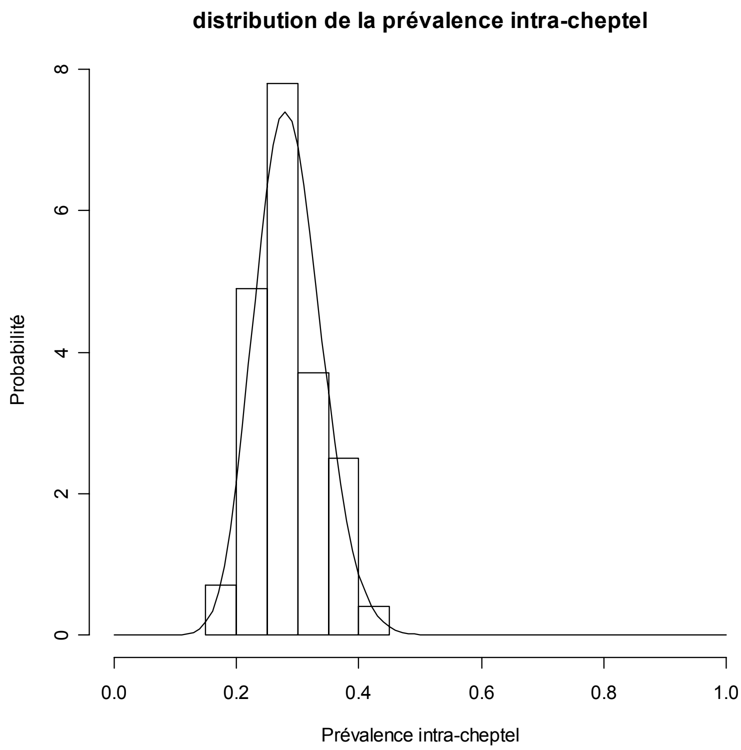
3. **on dispose d'un grand nombre de données observées** : il existe alors deux possibilités : *i*) soit l'on utilise la distribution des données observées comme loi de probabilité du paramètre ; *ii*) soit l'on « ajuste » une loi de probabilité classique à ces données. Par exemple, si l'on dispose de 200 valeurs observées de prévalence intra-cheptel, il sera possible soit d'utiliser la distribution observée (dite « empirique »), soit d'ajuster une loi de densité adéquate (figure 4) ;

4. **on ne dispose pas de données** : si l'on ne dispose pas de données, il sera nécessaire d'avoir recours à des dires d'experts. Il faudra extraire des connaissances des experts un certains nombre de données permettant de spécifier la loi de probabilité possible pour le paramètre. Nous ne développerons pas ici l'ensemble des techniques permettant d'obtenir un consensus auprès d'un groupe d'experts. On retiendra simplement que le consensus est d'autant plus simple à obtenir que le nombre de valeurs demandées est limité, et possède une signification évidente pour les experts. Ainsi, il sera généralement plus facile d'obtenir une valeur moyenne qu'un écart-type.

L'article de Pouillot [2002, ce numéro] présente un complément pour le choix des lois de distributions, selon le type de variable et les données disponibles.

FIGURE 4

Exemple d'utilisation d'un grand nombre de données observées : l'histogramme présente la distribution des données observées : cette loi de probabilité peut être utilisée directement dans l'analyse. Il est également possible d'ajuster sur ces données une loi de probabilité connue, telle que celle représentée par la courbe (ici, une loi bêta).

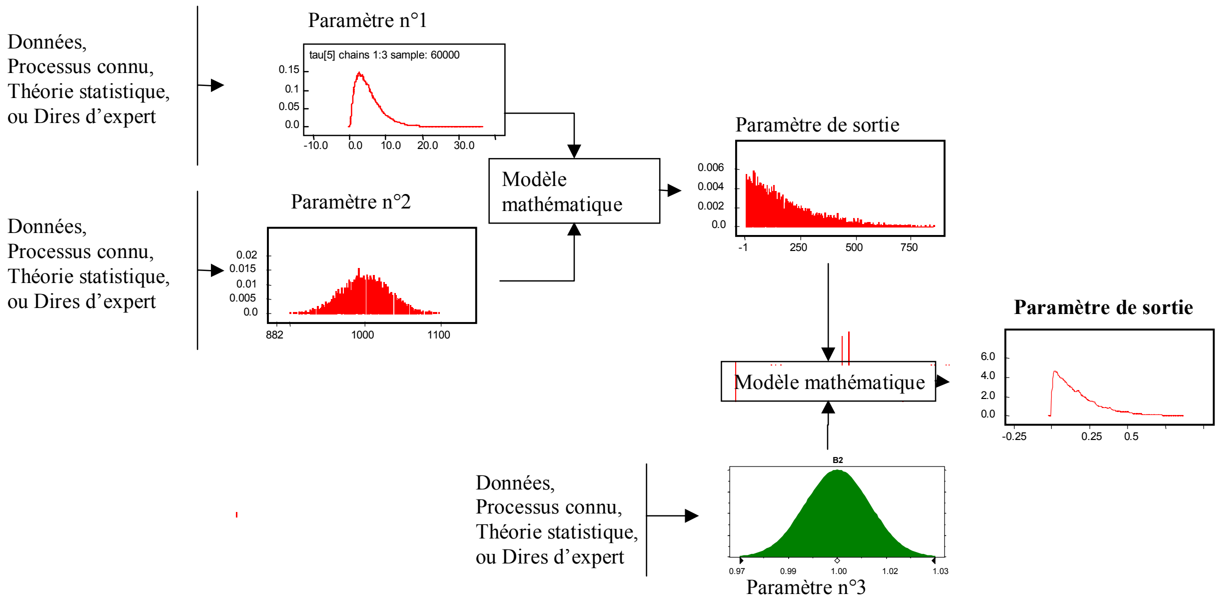


3. OBTENTION DE LA COURBE DE PROBABILITE DU RISQUE, SIMULATIONS DE « MONTE-CARLO »

L'estimation ponctuelle de risque peut généralement être obtenue directement à l'aide de calculs probabilistes. Si les différents paramètres ne sont plus des valeurs ponctuelles, mais des distributions de probabilité, les spécifications des distributions de paramètres issues d'une combinaison de distributions peuvent être plus compliquées. La figure 5 présente le processus global : à chaque paramètre correspond une loi de probabilité. La combinaison mathématique ou probabiliste entre les différents paramètres va aboutir à une nouvelle loi de probabilité, qui soit sera le paramètre final de l'analyse, soit sera elle-même combinée à une autre loi de distribution … jusqu'à obtention du paramètre d'intérêt. Généralement, les calculs nécessaires à l'obtention de la combinaison de plusieurs lois de probabilités sont impossibles à réaliser manuellement ou à l'aide d'un tableur classique type Excel (© Microsoft, Corp.).

FIGURE 5

Schéma général de la démarche de l'appréciation quantitative probabiliste des risques



La simulation dite de « Monte-Carlo » permet de résoudre ce problème. Son principe est simple. La loi de distribution d'un paramètre (que l'on nommera Y) issu de la combinaison de deux paramètres (que l'on nommera X_1 et X_2) peut être approchée à l'aide de l'algorithme suivant :

1. on tire au sort une valeur (que l'on nommera x_1) dans la loi de distribution de X_1 ;
2. on tire ensuite au sort une valeur (que l'on nommera x_2) dans la loi de distribution de X_2 ;
3. on combine les valeurs x_1 et x_2 pour obtenir une valeur y.

On répète les procédures 1 à 3 un grand nombre de fois. La distribution observée des différentes valeurs de y obtenues tend vers la loi de distribution exacte de Y quand le nombre de répétitions tend vers l'infini.

Exemple : Dans le cadre de contrôle à l'importation on utilise un test dont la valeur de la sensibilité est incertaine. La valeur exacte de la sensibilité d'une méthode diagnostique a été estimée par des experts comme suivant une loi telle que présentée à la figure 6a. On teste 100 animaux infectés avec ce test, et l'on désire connaître la distribution du nombre d'animaux infectés détectés par la méthode. La procédure de Monte-Carlo va opérer comme suit :

1ère répétition : tirage au sort d'une valeur issue de la distribution de la sensibilité : par exemple 0,61. Puis tirage au sort du nombre d'animaux infectés que l'on détecterait si la vraie valeur de sensibilité était 0,61 : par exemple 70 ;

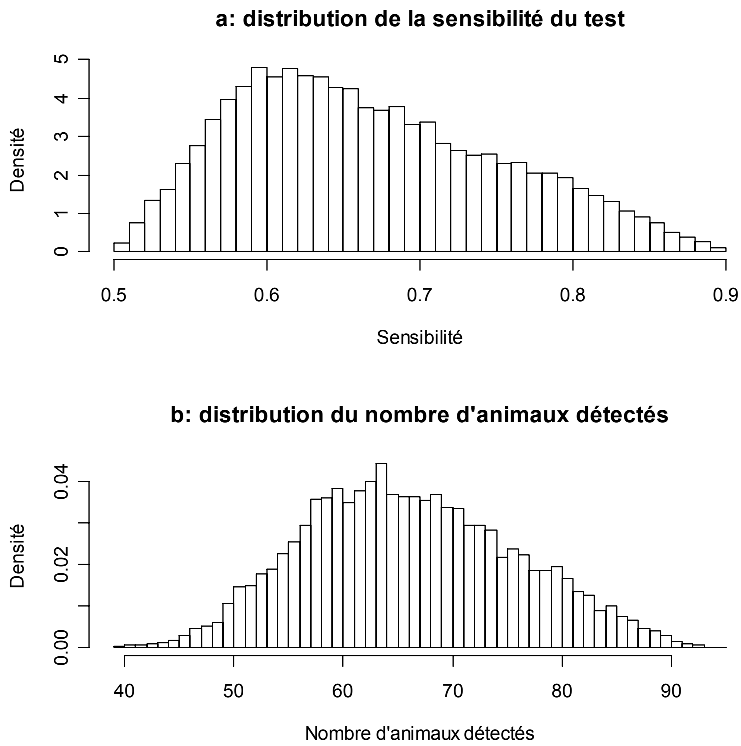
2ème répétition : tirage au sort d'une valeur issue de la distribution de la sensibilité : par exemple 0,82. Puis tirage au sort du nombre d'animaux infectés que l'on détecterait si la vraie valeur de sensibilité était 0,82 : par exemple 75 ;

3ème répétition ….

La figure 6 présente la distribution de la sensibilité de la méthode et du nombre d'animaux infectés détectés parmi les 100 animaux infectés. Cette dernière distribution n'est pas « exacte » car il n'est pas possible de lui donner une forme mathématique simple. Cependant, elle reflète cette distribution réelle.

FIGURE 6

Exemple de simulation de Monte-Carlo : a) histogramme des valeurs de *sens*, sensibilité du test ; b) histogramme de *n*, nombre d'animaux infectés détectés par le test si la sensibilité de ce test suit la loi de distribution présentée dans le graphique a (*cf*. texte)



III - PRESENTATION ET INTERPRETATION DES RESULTATS

Le résultat de l'analyse de risque est une courbe de probabilité. La figure 7 présente deux types de présentations possibles, résultats de l'analyse du modèle présenté sur la figure 2 après spécification de loi de la distribution adéquate pour chacun des paramètres (démarche non montrée).

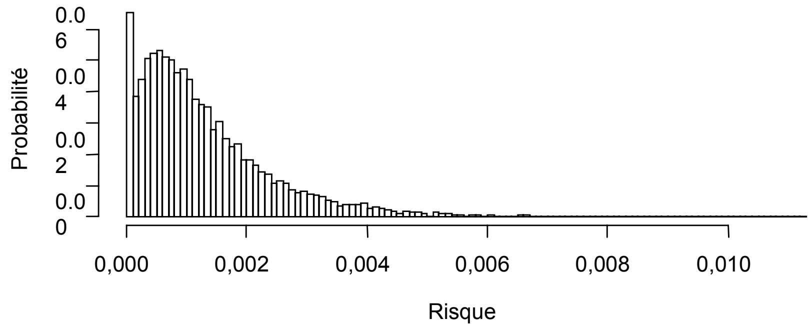
La distribution du risque associe à chaque valeur possible simulée du risque sa probabilité. La figure 7a présente simplement l'histogramme des valeurs simulées du risque. Cette présentation permet de donner quelques indications sur la forme de la densité du risque : ici, le risque 0 est fréquent, la valeur maximale du risque simulée est environ 0,12.

La distribution cumulative du risque (figure 7b) est de lecture plus complète, mais moins aisée : on peut connaître immédiatement la probabilité correspondant à toute valeur de risque d'introduction de l'axe des abscisses. En effet, la verticale passant par n'importe quelle valeur de l'axe des abscisses coupe la courbe en un point dont la projection sur l'axe des ordonnées (ligne horizontale) indique le pourcentage de probabilité. Ainsi, la verticale passant par 10^{-3} coupe la courbe en un point dont la projection sur l'axe des ordonnées correspond à environ 50%. Ceci signifie qu'il y a 50% de chances pour que la probabilité d'introduction du danger soit inférieure à 10^{-3}, 50% de chances pour quelle soit supérieure. De même, il y a 80% de chances pour que la probabilité d'introduction du danger soit inférieure à 2×10^{-3}, 20% de chances pour quelle soit supérieure.

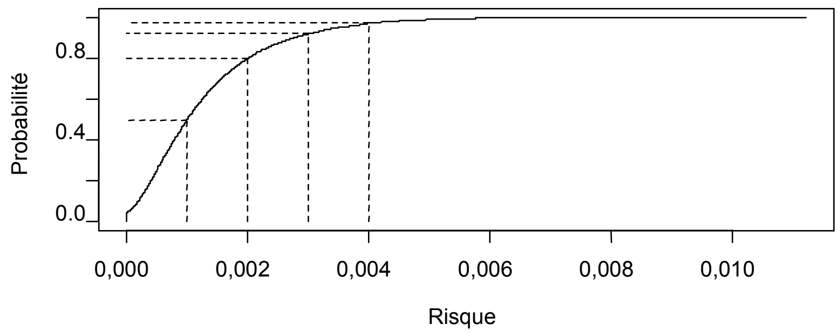
FIGURE 7

Exemple de présentation des résultats de l'appréciation quantitative probabiliste des risques

a: Distribution du risque



b: Distribution cumulative du risque



La diffusion et la compréhension des résultats d'une analyse de risque peuvent être complexes pour les non-initiés : il faut concevoir qu'il s'agit d'une loi de probabilité d'une probabilité. En pratique, la compréhension sera plus simple en extrayant quelques statistiques de la courbe. Par exemple, le risque moyen, le risque médian, la probabilité que le risque soit supérieur à quelques valeurs particulières. Il peut être intéressant également de fournir le 2,5^{ème} et le 97,5^{ème} percentile de la distribution : l'intervalle correspondant est un intervalle de confiance du risque. Le tableau I propose quelques statistiques issues de notre modèle.

L'intérêt pour le gestionnaire de posséder une estimation de la probabilité que le risque soit supérieur à certaines valeurs est évident. Dans notre exemple, si le risque acceptable est de 0,2%, le risque moyen est acceptable (car égal à 0,13%). En revanche, le gestionnaire est informé que ce risque peut être dépassé dans 20% des cas.

TABLEAU I

Exemple de statistiques descriptives de la loi de probabilité du risque

Statistique	**Valeur du risque**
moyenne	0,13%
médiane	0,10%
2,5^{ème} percentile	0,00%
97,5^{ème} percentile	0,42%
Probabilité que le risque soit supérieur à :	
0 %	95 %
0,01%	93 %
0,1 %	50 %
0,2 %	20 %
…	
1 %	0 %

IV - REALISATION PRATIQUE DE L'APPRECIATION QUANTITATIVE PROBABILISTE DES RISQUES

En pratique, l'appréciation quantitative des risques n'est généralement pas réalisable sans l'aide de logiciels. Les deux plus couramment utilisés dans le domaine biologique sont des compléments du logiciel Excel© : @Risk (© Palisade, Corp) et Crystal Ball(© Decisioneering, Inc). Les relations entre les différents paramètres sont tout d'abord construites dans une feuille de calcul Excel© classique, exactement comme s'il s'agissait d'une appréciation quantitative ponctuelle des risques, puis des lois de distribution sont choisies pour chacun des paramètres. Ces deux logiciels comportent de nombreuses fonctions permettant de spécifier les lois de distribution à partir de données observées ou à partir de dires d'experts. Une application graphique permet de visualiser directement la forme des lois de distribution. Ils permettent ensuite de procéder aux simulations de Monte-Carlo. Les résultats de l'analyse peuvent être extraits sous forme graphique. Enfin, des outils complémentaires d'étude du modèle sont proposés. Les lecteurs intéressés sont invités à se reporter au manuel de ces logiciels pour des compléments d'information.

V - COMPLEMENTS

1. SPECIFICATION DES CORRE-LATIONS ENTRE LOIS DE DISTRIBUTIONS DES VARIABLES ENTRANT DANS LE MODELE

Un des grands intérêts de l'appréciation quantitative probabiliste des risques est la possibilité de prendre en compte, de manière assez simple avec les logiciels dédiés actuellement sur le marché, une éventuelle corrélations (c'est-à-dire un éventuel lien) entre les différents paramètres.

Rappelons que la présence de corrélations invalide les évaluations quantitatives ponctuelles si elles ne sont pas prises en compte. Il n'est ainsi pas possible, en cas de présence d'une corrélation, de multiplier simplement des probabilités[3] ou des moyennes[4].

Prenons l'exemple d'un taux de prévalence intra-cheptel corrélé positivement à l'effectif du cheptel, ce qui signifie que les taux de prévalence intra-cheptel les plus faibles sont rencontrés dans des troupeaux de plus faible effectif. La moyenne du nombre d'animaux atteints par cheptel ne sera pas égale à la moyenne du taux de prévalence intra-cheptel multiplié par l'effectif moyen. L'estimation ponctuelle du risque est alors fausse, si l'on ne prend pas en compte cette corrélation. Si, dans la procédure de Monte-Carlo, le lien entre ces deux variables n'est pas simulé, le logiciel tirera au sort indépendamment les deux variables dans leur loi de distribution respective, sans considérer la présence plus probable d'une faible prévalence intra-cheptel dans les petits troupeaux : l'estimation probabiliste sera également erronée.

L'absence de prise en compte de liens entre les variables peut invalider l'estimation de la moyenne du risque, mais son influence est encore plus importante sur la mesure de la variabilité (ou de l'incertitude) du risque.

Les logiciels dédiés d'analyse de risque permettent, lors de la procédure de Monte-Carlo, de simuler des jeux de paramètres issus de lois de distribution en assurant un certain lien entre les paramètres, spécifié par l'utilisateur.

La spécification de l'intensité du lien entre deux variables se mesure par le coefficient de corrélation (« classique » ou « des rangs »). Cette valeur est comprise entre -1 (corrélation totale négative : toute valeur plus élevée de la première variable entraînera une valeur moins élevée de la seconde variable) et 1 (corrélation totale positive : toute valeur plus élevée de la première variable entraînera une valeur plus élevée de la seconde variable) en passant par 0 (absence de corrélation : la valeur de la première variable n'a aucune influence sur la valeur de la seconde variable). La figure 8 présente l'influence du coefficient de corrélation reliant deux variables.

En pratique, la spécification du coefficient de corrélation des rangs est réalisée soit par dire d'experts, soit par une estimation réalisée à partir d'un jeu de données : on dispose d'une mesure des deux variables sur une série d'individus, on calcule le coefficient de corrélation entre ces deux variables, et on utilise cette estimation dans la simulation des valeurs, à l'aide du logiciel d'analyse de risque.

Il est essentiel, lors de toute évaluation des risques, d'étudier les liens entre les différentes variables, ou au moins de discuter l'absence de leur prise en compte, le cas échéant.

2. ANALYSE DE SENSIBILITE DU MODELE

Pour certaines analyses, le résultat de l'appréciation quantitative des risques est moins important que l'analyse de l'influence de chacun des facteurs d'entrée sur le risque. Par exemple, dans le cadre d'une importation d'animaux vivants, il peut être plus important de connaître l'influence de la réalisation d'un test à l'importation sur le risque que le risque lui-même. Un des intérêts de l'appréciation quantitative de risque probabiliste est de pouvoir proposer une quantification approximative de l'influence des paramètres d'entrée sur le risque.

Le principe est le suivant : la simulation de Monte-Carlo permet d'obtenir une valeur de risque pour chaque jeu de paramètre d'entrée, tirés au sort dans les lois de distribution spécifiées par l'évaluateur. L'étude du lien entre chacune des valeurs des paramètres d'entrée et le risque est analysée par une statistique particulière : le coefficient de corrélation. *Nota* : le coefficient de corrélation classiquement utilisé repose sur une hypothèse de distribution normale des paramètres : on préfère donc utiliser un coefficient de corrélation établi sur les rangs des paramètres étudiés, appelé coefficient de corrélation des rangs de Spearman, insensible à cette distribution.

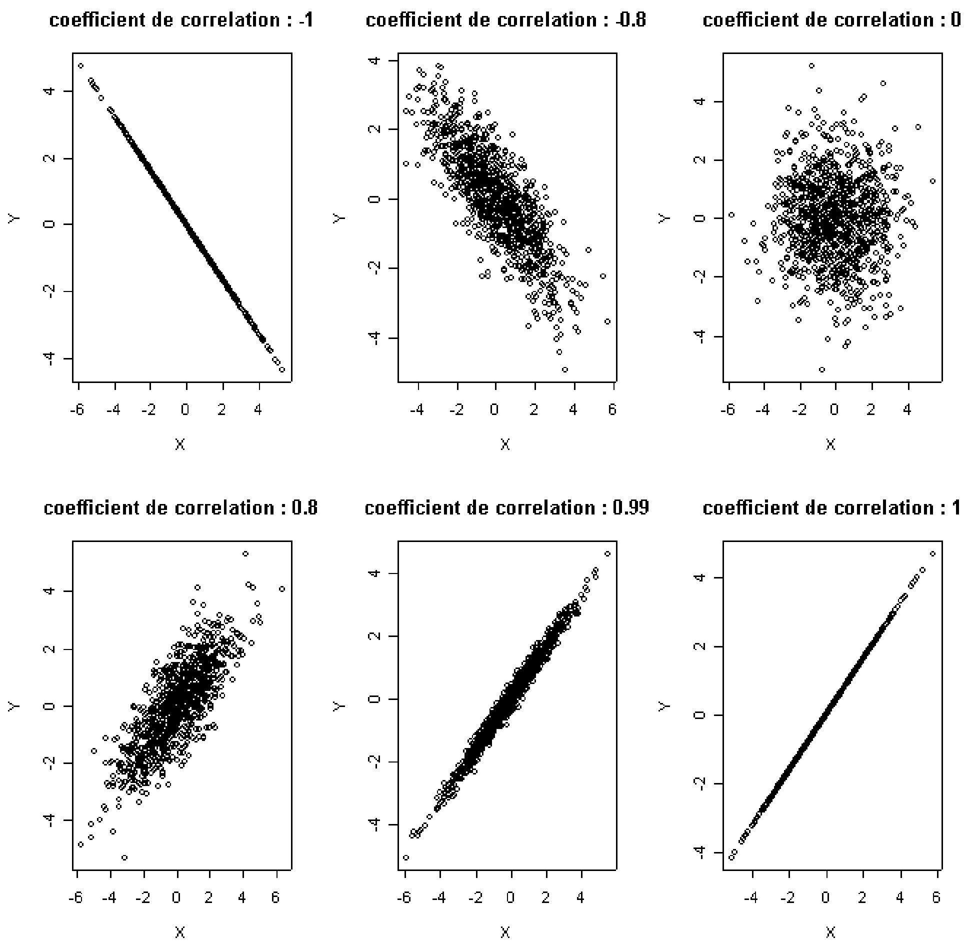
---

[3] car $\Pr(X \text{ et } Y) = \Pr(X) \times \Pr(Y)$ seulement si X et Y sont indépendants.

[4] car $E(X \times Y) = E(X) \times E(Y)$ seulement si X et Y sont indépendants.

FIGURE 8

Exemples de l'influence du coefficient de corrélation entre deux variables. 1 000 échantillons d'une variable aléatoire *X* et 1 000 échantillons d'une variable aléatoire *Y* ont été tirés au sort en assurant un coefficient de corrélation égal à -1, -0,8, 0, 0,8, 0,99 et 1.
On note que la liaison entre les deux variables augmente avec la valeur absolue du coefficient de corrélation. Quand le coefficient de corrélation est égal à 0, les deux variables sont indépendantes.



Une valeur absolue élevée du coefficient de corrélation des rang de Spearman souligne une forte influence du paramètre d'entrée sur le risque ; une valeur positive souligne une corrélation positive : une augmentation du paramètre d'entrée est liée à une augmentation du risque ; une valeur négative souligne une corrélation négative : une augmentation du paramètre d'entrée est liée à une diminution du risque.

La représentation classique de cette analyse, dite analyse de sensibilité, en appréciation quantitative des risques est souvent nommée « tornado chart ». On présente sur un graphique les coefficients de corrélation des rangs de Spearman par ordre de valeur absolue décroissante. Les principaux facteurs d'influence sur le risque sont donc présentés en premier. La figure 9 présente le « tornado chart » du modèle proposé comme exemple. On constate que le risque est fortement corrélé avec la prévalence réelle animale (ce qui était *a priori* évident), puis avec le nombre de foyers déclarés dans la zone. Une autre corrélation importante semble être la sensibilité du test d'achat. Cette corrélation est négative, signifiant que plus la sensibilité est élevée, plus le risque est faible (ce qui est logique). Les autres corrélations sont, dans l'ordre de valeur absolue décroissante : l'effectif des cheptels atteints, le nombre d'animaux achetés, la sensibilité du test de dépistage et enfin, la prévalence intra-cheptel. L'observation de ce « tornado chart » permet de vérifier, pour des modèles simples, la cohérence du sens de la corrélation. Par exemple, il est normal de constater une corrélation négative entre la sensibilité des tests et le risque, une corrélation positive entre le nombre de foyers et le risque, etc.

Pour « l'estimateur du risque », il est intéressant de regarder les paramètres les plus influents : si la distribution utilisée pour ces paramètres servait à modéliser de l'incertitude, il pourrait être intéressant d'obtenir des données supplémentaires spécifiquement sur ces paramètres, afin d'affiner l'estimation du risque. Au contraire, s'il s'avère qu'un paramètre incertain n'a pas d'influence sur le modèle, il ne vaut peut être pas la peine d'affiner son estimation.

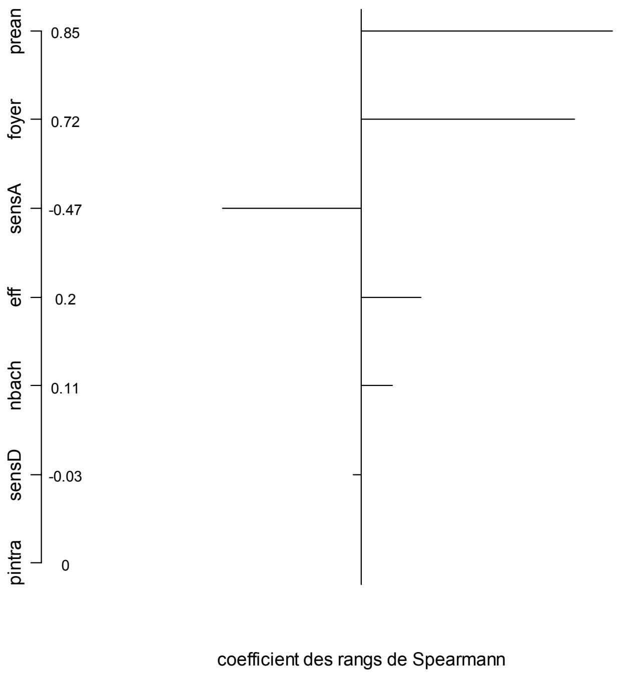
Si le « gestionnaire du risque » peut agir sur des paramètres, il sera plus intéressant d'agir sur les paramètres les plus influents. Dans notre exemple, la sensibilité du test d'achat pourrait par exemple être augmentée en utilisant une série de tests en parallèle : la réduction du risque sera plus importante qu'en agissant sur le nombre d'animaux achetés ou la sensibilité du test de dépistage.

Il est directement possible, en modifiant certaines distributions de paramètres, de mesurer l'influence d'une mesure de gestion sur le risque. Ainsi, il suffirait dans notre exemple d'introduire un test d'achat de plus grande sensibilité, puis de répéter l'analyse, pour mesurer directement la réduction du risque dans le schéma actuel.

FIGURE 9

Graphe de sensibilité (« Tornado Chart ») du modèle.

Les correspondances des variables sont les suivantes : « prean » : taux de prévalence animale réelle, « foyer » : nombre de foyers déclarés, « sensA » : sensibilité du contrôle d'achat, « eff » : taille moyenne des cheptels infectés une année donnée, « nbach » : nombre d'animaux achetés une année donnée, « sensD » : sensibilité du dépistage, « pintra » : prévalence intra-cheptel moyenne dans les cheptels infectés.



3. SEPARATION DE L'INCERTITUDE ET DE LA VARIABILITE

Ce paragraphe pourra n'être abordé qu'en seconde lecture.

L'appréciation des risques telle que présentée précédemment propose une modélisation globale de l'incertitude et de la variabilité. Or, il est de plus en plus recommandé par les instances internationales de séparer, dans l'appréciation du risque, la « variabilité » de « l'incertitude », puisqu'elles ne représentent pas la même grandeur. On réalise à cet effet un type de modélisation particulier, nommé modélisation de second ordre. Le principe général est de dissocier, lors de la simulation de Monte-Carlo, la simulation des paramètres **incertains** des paramètres **variables**, puis de représenter graphiquement, de manière séparée, les fluctuations de l'estimation du risque inhérentes à la variabilité des paramètres de celles inhérentes à l'incertitude des paramètres.

PRINCIPE DE LA MODELISATION DE SECOND ORDRE

La modélisation réalisée est la suivante :

1. on tire au sort de nos lois de distributions un jeu de données pour tous les paramètres reflétant de **l'incertitude** ;
    a. à l'aide de ces paramètres, considérés alors comme fixes, on réalise une appréciation quantitative du risque, en utilisant tous les paramètres reflétant de la **variabilité**. On obtient donc une courbe de distribution de risque cumulée, telle que proposé dans la figure 7 b ;
    b. on calcule différentes statistiques (moyenne, variance, médiane, différents percentiles,…) sur cette courbe de distribution ;
2. on reprend à l'étape 1, un grand nombre de fois.

En fin de simulation, on calcule les percentiles des statistiques obtenues à l'étape 1b. Le 2,5ème et le 97,5ème permettent alors de proposer une intervalle de confiance des estimations, la médiane (50ème percentile) permet de proposer une estimation du risque médian, tel que, toute incertitude prise en compte, 50% des estimations du risque sont supérieures, 50% des estimations sont inférieures.

Cette procédure s'appelle « modélisation de second ordre » car il s'agit de deux modélisations : une modélisation de la variabilité, imbriquée dans une modélisation de l'incertitude.

Dans l'exemple de la figure 2, on peut considérer que seuls les paramètres de sensibilité des tests sont incertains, tous les autres étant variables. On va réaliser 1 000 appréciations quantitatives des risques, chaque appréciation étant définie par un jeu de valeurs de sensibilité du test d'achat et de sensibilité du test de dépistage, tirées au sort selon leur loi de distribution adéquate.

Le tableau II présente un exemple de résultats obtenus : lors de la première « appréciation quantitative des risques », on a tiré au sort une valeur de sensibilité du test de dépistage de 0,964, et une valeur de sensibilité du test d'achat de 0,962. Les résultats de l'appréciation quantitative des risques sont les suivants : le risque moyen estimé est de $9{,}8 \times 10^{-4}$; le 90ème percentile du risque estimé est de $18{,}43 \times 10^{-4}$.

TABLEAU II

Résultats de la simulation de second ordre

N° de simulation	Sensibilité dépistage	Sensibilité achat	Risque moyen estimé ($\times 10^4$)	90ème percentile du risque ($\times 10^4$)	…
1	0,964	0,962	9,80	18,43	…
2	0,950	0,958	10,86	21,66	…
3	0,988	0,968	7,95	15,35	…
…	…	…	…	…	…
1000	0,912	0,948	10,03	19,29	…
Puis on calcule par colonne					
		2,5ème percentile	4,63	9,17	…
		médiane	12,02	23,37	…
		97,5ème percentile	25,22	48,97	…

Lors de la seconde « appréciation quantitative des risques », on a tiré au sort une valeur de sensibilité du test de dépistage de 0,950, et une valeur de sensibilité du test d'achat de 0,958. Les résultats de l'appréciation quantitative des risques sont : risque moyen estimé $10{,}86 \times 10^{-4}$; 90ème percentile du risque estimé $21{,}66 \times 10^{-4}$.

On répète la procédure 1000 fois. Ensuite, on calcule, pour chacune des statistiques issues des analyses de risque la médiane, le 2,5^{ème} et le 97,5^{ème} percentile (par exemple). On obtient pour la moyenne du risque : médiane de la moyenne = $12{,}02 \times 10^{-4}$, 2,5^{ème} et 97,5^{ème} percentiles = 4,63 et $25{,}22 \times 10^{-4}$, respectivement.

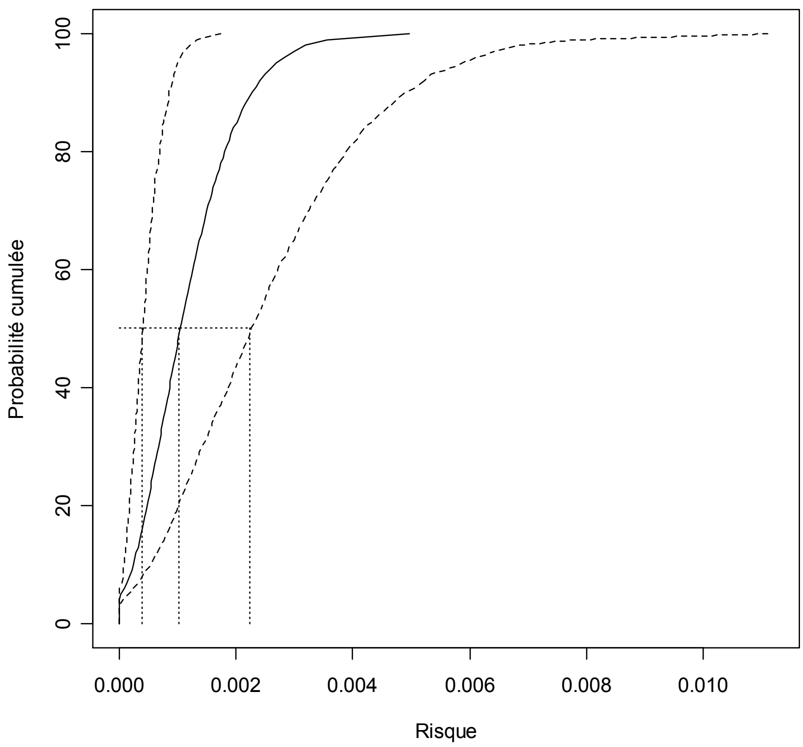
RESULTATS ET INTERPRETATION DE LA MODELISATION DE SECOND ORDRE

Dans notre exemple, la lecture est la suivante : le risque moyen estimé est d'au moins $12{,}02 \times 10^{-4}$ dans 50% des simulations, après prise en compte de l'incertitude. L'intervalle de crédibilité de cette estimation est de $[4{,}63 ; 25{,}22] \times 10^{-4}$.

Le même raisonnement peut être conduit pour le 90ème percentile. On prend alors en compte la variabilité du risque, mais on estime l'incertitude des estimations.

Ces résultats peuvent être proposés sous la forme d'un graphique (figure 10). La lecture se fait comme pour la figure 7. Cependant, on ne dispose pas d'une courbe, mais de trois courbes : la courbe en traits pleins représente la médiane des estimations, les courbes pointillées présentent les 2,5^{ème} et les 97,5^{ème} percentiles. Pour une « variabilité du risque » donnée (un percentile particulier, à lire en ordonnées), on obtient le risque estimé, abscisse du point d'intersection entre la courbe pleine et la parallèle à l'axe de abscisses passant par le percentile désiré. On peut obtenir l'intervalle de crédibilité de cette estimation en réalisant la même procédure avec les courbes pointillées. Dans la figure 10, on peut ainsi estimer le 50ème percentile du risque à environ 10^{-3} avec un intervalle d'incertitude d'environ [$0{,}5\ 10^{-3}$; $2{,}2\ 10^{-3}$].

FIGURE 10
Représentation graphique des résultats d'une modélisation de second ordre



VI - CONCLUSION

Cet article propose des éléments de base de la théorie de l'appréciation quantitative probabiliste des risques. Cette méthode est une méthode complexe, qui ne peut en général se résoudre à l'application de « recettes » piochées dans tel ou tel ouvrage, tel ou tel manuel de logiciel.

L'intérêt de l'appréciation quantitative probabiliste du risque est évident, en comparaison de l'appréciation quantitative ponctuelle du risque, notamment par la possible prise en compte de la variabilité et de l'incertitude des paramètres, et, par le transfert de cette variabilité dans le modèle, par la spécification de la variabilité et l'incertitude du risque estimé.

Cependant, ce type d'analyse implique la connaissance, ou au moins l'appréhension, de la loi de distribution de l'ensemble des paramètres. Cette tâche peut être ardue : elle ne doit pas être l'alibi pour utiliser des données d'origines diverses, sans discussion de la validité de ces données pour estimer un risque particulier.

BIBLIOGRAPHIE

AndersonN E.L., Hatis, D. ~ When and how can you specify a probability distribution when you don't know much. Foundations: A. Uncertainty and Variability. *Risk Anal.*,1999, **19,** 47-49.

Vose D. ~ Risk analysis. A quantitative guide. 418 pages, Wiley. Chichester, 2000.

Murray N. ~ Import risk analysis. Animals and animals products. 183 pages, MAF New Zealand, Wellington (NZ), 2002.

Pouillot R. ~ Caractérisation d'une loi de distribution d'une variable entrant dans un modèle d'analyse probabiliste de risque. *Epidémiol. et santé anim*.**,** 2002, **41**, 111-141.

Pouillot R., Sanaa M. ~ Bases probabilistes et statistiques nécessaires à l'appréciation du risque. *Epidémiol. et santé anim*.**,** 2002, **41**, 65-81.

Toma B. ~ L'appréciation quantitative du risque : notions générales. *Epidémiol. et santé anim*.**,** 2002, **41**, 51-63.

Réalisation pratique : le lecteur est invité à se référer aux manuels des logiciels @Risk (www.palisade.com) et Crystal-Ball (www.decisioneering.com).