

Analyse de sensibilité globale basée sur
l'entropie

Benjamin Auder

Septembre 2006

**Étude théorique, Algorithmes et
Applications**

encadré par Bertrand Iooss (CEA Cadarache).

Table des matières

1	Introduction	6
2	Mesures d'incertitude	7
2.1	Analyse de sensibilité basée sur la variance	7
2.1.1	Intérêt de la variance	7
2.1.2	Indices de sensibilité	8
2.2	Analyse de sensibilité basée sur l'entropie	9
2.2.1	Pourquoi H a-t-elle cette forme?	9
2.2.2	Information mutuelle	12
2.2.3	Indices de sensibilité	13
2.2.4	Extension au cas multivarié	16
2.3	Calculs d'entropie	17
2.3.1	Entropies de lois continues	17
2.3.2	Entropies de lois discrètes	22
2.3.3	Comparaison entropie discrète - entropie continue	27
2.4	Bornes sur la variance et l'entropie	29
2.4.1	Maximisation de la variance	29
2.4.2	Maximisation de l'entropie	32
2.4.3	Borne inférieure de l'entropie	35
2.5	Comportements de la variance et de l'entropie	36
2.5.1	Évolution qualitative variance - entropie	36
2.5.2	Surfaces "iso-information"	39
2.5.3	Indices basés sur l'entropie continue	43
3	Calcul pratique des indices de sensibilité η	44
3.1	Algorithmes	44
3.1.1	Fonction connue	45
3.1.2	Boîte noire	47
3.1.3	Instabilité des algorithmes	48
3.2	Convergence avec le nombre de sous-intervalles	49
3.2.1	Restriction de l'étude à l'entropie discrète	50
3.2.2	Évolution de l'entropie discrète	52
3.2.3	Évolution de la quantité d'information discrète	53
3.3	Convergence avec la taille de l'échantillon	55
3.3.1	Échantillonnage déterministe pour les indices η	56
3.3.2	Vitesse de convergence de l'entropie	56
3.3.3	Vitesse de convergence de l'information mutuelle	58
3.3.4	Généralisation à des lois non uniformes	58
3.3.5	Bilan sur la convergence	60

4	Applications à des fonctions tests	60
4.1	Fonction d'Ishigami	61
4.1.1	Indices exacts	61
4.1.2	Indices approchés	63
4.1.3	Indices multi-sorties	65
4.1.4	Influences des couples de variables	66
4.2	Fonction g de Sobol	68
4.2.1	Indices théoriques	69
4.2.2	Indices expérimentaux	71
4.2.3	Influences des couples de variables	71
4.3	Limitations sur les indices : non-discrimination	73
4.3.1	Exemple où la variance n'apporte pas d'information	73
4.3.2	Exemple où l'entropie n'apporte pas d'information	74
4.3.3	Les indices de sensibilité ρ	75
5	Conclusion	76

Lexique :

Analyse de sensibilité :

Étude quantifiant les influences de chaque paramètre d'entrée d'un modèle sur sa sortie. Celle-ci diffère de l'analyse des incertitudes, où l'on cherche à donner l'incertitude sur la variable de sortie en fonction de celles sur les variables d'entrées (supposées connues).

Bien sûr ces deux concepts sont fortement reliés, car si un paramètre d'entrée a une grande influence sur la sortie, l'incertitude qu'il propage sur la sortie est d'autant plus grande.

Variance :

Grandeur qui quantifie l'écart quadratique moyen à la moyenne d'une variable aléatoire X . C'est donc un terme de dispersion autour de la moyenne : plus il est grand, plus les points se répartissent loin de la moyenne, et inversement une faible valeur signifie que beaucoup de points sont proches de la moyenne.

La variance s'écrit :

$$V(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Entropie :

Grandeur introduite pour modéliser le manque d'information sur la vraie valeur d'une variable aléatoire ; par exemple si on sait qu'un variable suit une loi normale, on a plus d'indications sur sa valeur que si elle répartit les points uniformément. Cette quantité est d'une certaine manière liée à la précédente car une faible variance semble indiquer beaucoup d'information, mais il y a des différences importantes entre les deux notions.

L'entropie s'écrit :

$$H(X) = - \int f(x) \ln(f(x)) d\mu(x),$$

μ étant une mesure appropriée (discrète ou continue, la plupart du temps), et f la densité de X pour cette mesure.

I : Information mutuelle.

L'information mutuelle mesure la connaissance apportée sur une variable

aléatoire par une autre ; elle indique donc d'une certaine manière la dépendance entre deux variables aléatoires.

$\underline{\eta}$: Indice de sensibilité basé sur l'entropie.

$\underline{\rho}$: Indice de sensibilité utilisant la variance et l'entropie pour de meilleurs résultats.

\underline{S} : Nombre de points dans les échantillons utilisés pour les algorithmes.

\underline{n} : Dans le cas des algorithmes se basant sur l'entropie, nombre de sous-intervalles de taille $\frac{|I|}{n}$ où $|I|$ est la longueur de l'intervalle en question.

$\underline{\psi}$: Modèle liant des entrées X_1, \dots, X_s à une sortie $Y = (Y_1, \dots, Y_d)$ par une relation connue ou non.

\underline{s} : Nombre de paramètres d'entrées pour un modèle ψ .

\underline{d} : Toujours dans le cadre d'un modèle ψ , dimension du vecteur Y en sortie.

1 Introduction

La simulation de phénomènes physiques complexes conduit souvent à des modèles dépendant d'un grand nombre de paramètres, dont certains peuvent être redondants ou d'effet négligeable. Il est alors intéressant d'isoler les facteurs les plus influents. On doit pour cela définir ce qu'est un facteur influent, et donc quantifier sa contribution au modèle.

La plupart du temps les paramètres d'entrée ne sont pas connus précisément. Par exemple si le système dépend de la température ambiante on ne peut donner qu'une estimation de l'évolution de la température au sein du système ; ces paramètres sont donc naturellement modélisés par des variables aléatoires.

On s'intéresse ici à l'analyse de sensibilité globale : les influences des paramètres sont étudiées dans leur globalité, et non localement comme on pourrait le faire en calculant des dérivées au voisinage d'un point. On évite ainsi les erreurs que l'on ferait en discrétisant trop grossièrement une fonction à variations rapides : c'est un peu comme si on effectuait une moyenne des sensibilités locales.

À l'heure actuelle presque toutes les études concernant les indices de sensibilité - globaux, donc - sont basées sur la mesure de la variance. Or nous allons voir que cette dernière quantité n'est pas toujours la mieux adaptée, et n'est pas discriminante dans certains cas. Par exemple, des distributions bimodales peuvent être très différentes mais avoir la même variance. Or ce type de distribution est apparu dans certaines applications (Iooss & Van Dorpe [2], Volkova et al. [10]). Middleton [6] a montré que l'entropie pouvait être une mesure d'information bien plus adaptée que la variance. Krzykacz-Hausmann [5] a utilisé l'entropie pour l'analyse de sensibilité : il a défini de nouveaux indices de sensibilité qu'il estime plus adaptés dans le cas où des paramètres d'entrée déterministes ne sont pas connus exactement (incertitude "épistémique"). La variance cède alors la place à l'entropie. Quels sont les avantages obtenus ? Nous essaierons de préciser ce point autant que possible.

L'objectif de ce travail est donc d'étudier et d'appliquer l'analyse de sensibilité basée sur l'entropie.

Après deux parties donnant respectivement les indices de sensibilité basés sur la variance et l'entropie, on présentera les algorithmes utilisés dans le cas de l'entropie ainsi que les tests et résultats expérimentaux obtenus.

2 Mesures d'incertitude

L'objectif de cette section est de trouver des fonctions pour mesurer l'incertitude sur la valeur d'une variable aléatoire.

2.1 Analyse de sensibilité basée sur la variance

Effectuons tout d'abord quelques rappels sur la variance.

2.1.1 Intérêt de la variance

L'analyse de sensibilité mentionnée en introduction se fait d'habitude à l'aide de la variance. Montrons pourquoi cette quantité est intéressante (paragraphe emprunté à Antoniadis [1]).

Soit un modèle Y dépendant des paramètres X_1, \dots, X_s . Le but est de quantifier l'influence de l'un des X_i (fixé) sur Y . Il apparaît donc naturel de considérer la prédiction que l'on peut faire de Y lorsque X_i est inconnu : on note celle-ci Y^i .

Comparons Y^i à Y , prédiction à partir du modèle complet $X = (X_1, \dots, X_s)$, avec une fonction d'erreur quadratique :

$$L = \mathbb{E}_X[(Y - Y^i)^2] = \int (Y(x) - Y^i(x))^2 f_X(x) dx.$$

Comment choisir Y^i ? Cherchant à minimiser L on s'aperçoit après calculs qu'il s'agit de l'espérance conditionnelle : $E[Y|X_i]$.

Par conséquent :

$$\begin{aligned} L_{min} &= \mathbb{E}_X[(Y - \mathbb{E}[Y|X_i])^2] \\ &= \mathbb{E}_X[((Y - \mathbb{E}[Y]) - (\mathbb{E}[Y|X_i] - \mathbb{E}[Y]))^2] \\ &= V(Y) + V(\mathbb{E}[Y|X_i]) - 2\mathbb{E}[(Y - \mathbb{E}[Y])(\mathbb{E}[Y|X_i] - \mathbb{E}[Y])], \end{aligned}$$

car d'après le théorème de l'espérance totale $\mathbb{E}[\mathbb{E}[Y|X_i]] = \mathbb{E}[Y]$.

Enfin :

$$\begin{aligned} L_{min} &= V(Y) - V(\mathbb{E}[Y|X_i]) \\ &= V(Y) \left(1 - \frac{V(\mathbb{E}[Y|X_i])}{V(Y)} \right) \\ &= V(Y)(1 - h^2), \end{aligned}$$

le coefficient h^2 représentant la fraction de la variance de Y apportée par X_i .

En conclusion les indices $\frac{V(\mathbb{E}[Y|X_i])}{V(Y)}$ sont parfaitement adaptés dans le cas où l'on veut représenter l'impact d'un paramètre d'entrée sur la variabilité de la sortie. Cette situation diffère de celle de l'entropie où on cherche plutôt à modéliser l'information apportée par la connaissance - déterministe - d'un des paramètres d'entrée sur la sortie, ne disposant que des lois des autres paramètres.

2.1.2 Indices de sensibilité

On présente rapidement dans ce paragraphe comment sont construits les indices de sensibilité basés sur la variance, qui sont appelés dans le cas général les *indices de Sobol* (Sobol [9], Jacques [3]).

Supposons (comme on le fait tout au long de ce rapport) que la sortie Y est liée aux paramètres d'entrées par une fonction ψ (connue ou non) : $Y = \psi(X_1, \dots, X_s)$, les X_i ($i = 1 \dots s$) étant des paramètres d'entrée indépendants. ψ est en fait un modèle de la relation physique probablement bien plus complexe reliant X et Y .

Une décomposition de la variance du modèle est alors possible :

$$V(Y) = \sum_{i=1}^s V_i(Y) + \sum_{1 \leq i < j \leq s} V_{ij}(Y) + \sum_{1 \leq i < j < k \leq s} V_{ijk}(Y) + \dots + V_{12\dots s}(Y), \quad (1)$$

où $V_i(Y) = V(\mathbb{E}[Y|X_i])$, $V_{ij}(Y) = V(\mathbb{E}[Y|X_i X_j]) - V(\mathbb{E}[Y|X_i]) - V(\mathbb{E}[Y|X_j])$, ...

On peut ainsi définir des indices de sensibilité de façon générale par :

$$S_i = \frac{V(\mathbb{E}[Y|X_i])}{V(Y)} = \frac{V_i(Y)}{V(Y)}, \quad S_{ij} = \frac{V_{ij}(Y)}{V(Y)}, \quad S_{ijk} = \frac{V_{ijk}(Y)}{V(Y)}, \quad \dots$$

Les indices de sensibilité d'ordre deux (S_{ij}) expriment la sensibilité de la variance de Y à l'interaction des variables X_i et X_j , et ainsi de suite. L'interprétation de ces indices (tous positifs) est naturelle : puisque grâce à (1), leur somme est égale à 1, plus l'indice est grand (proche de 1), plus la variable ou l'ensemble de variables a d'importance.

Des indices de sensibilité totaux ont également été définis afin d'exprimer la sensibilité "totale" de la variance de Y à une variable X_i :

$$S_{T_i} = S_i + \sum_{j \neq i} S_{ij} + \sum_{j \neq i, k \neq i, j < k} S_{ijk} + \dots = \sum_{l \in \#i} S_l,$$

où $\#i$ représente toutes les combinaisons possibles contenant l'indice i . En pratique, lorsque le modèle possède un grand nombre de variables d'entrée, on se contente d'estimer les indices de premier ordre, voire d'ordre deux, et les indices totaux.

2.2 Analyse de sensibilité basée sur l'entropie

L'entropie est une fonction bien connue en théorie de l'information, introduite historiquement par Shannon dans les années 1950, qui pourrait être adéquate dans le cas qui nous intéresse : en effet celle-ci modélise le manque d'information sur un système - et donc l'information, par complémentarité. De là à représenter l'importance d'un paramètre d'entrée dans un modèle, il n'y a qu'un pas.

Définition 1 *L'entropie d'une variable aléatoire discrète X prenant les valeurs x_1, \dots, x_n avec les probabilités respectives p_1, \dots, p_n est, à une constante multiplicative près (avec la convention $0 \ln(0) = 0$),*

$$H(X) = - \sum_{k=1}^n p_k \ln(p_k).$$

Pour une étude mathématique simplifiée on choisit le logarithme népérien, ceci n'ayant aucune influence sur les résultats.

Remarquons que cette grandeur ne dépend pas des valeurs prises par X , mais uniquement de leurs probabilités. C'est là une différence importante avec la variance qui se calcule par la moyenne des écarts à la moyenne, ces écarts dépendants des valeurs prises par la variable aléatoire. Nous reviendrons sur ce point au paragraphe 2.5.2.

2.2.1 Pourquoi H a-t-elle cette forme ?

Ce paragraphe est adapté du livre d'E.T. Jaynes ([4]).

Pourquoi ne pas avoir choisi une autre fonction pour modéliser l'incertitude ? Le choix de H est-il arbitraire ? Le cours raisonnement suivant va éclairer notre lanterne. On suppose qu'une fonction H existe, décrivant la quantité d'information manquante pour arriver au déterminisme, sur un système discret dont on connaît la distribution de probabilité des états : les p_k . Il s'ensuit trois caractéristiques naturelles devant être satisfaites par H , fonction de p_1, \dots, p_k :

1. H doit être continue. Si tel n'était pas le cas, un très léger changement sur les p_k provoquerait un "saut d'incertitude", ce qui n'est pas souhaitable.
2. Le sens commun nous dit que plus il y a d'états possibles équiprobables, moins l'état du système est connu avec précision. $H(\frac{1}{n}, \dots, \frac{1}{n})$ doit donc être une fonction croissante de n .
3. Si plusieurs raisonnements probabilistes sont possibles pour obtenir H , on doit aboutir au même résultat quel que soit la façon d'effectuer le calcul.

Supposons que l'on ait n propositions mutuellement exclusives A_1, \dots, A_n , de probabilités respectives p_1, \dots, p_n .

Groupons les événements A_i comme suit ($r_q = n$) :

$$\begin{aligned}
A_1 \cup \dots \cup A_{r_1} &= W_1, \quad \mathbb{P}(W_1) = \sum_{k=1}^{r_1} p_k = w_1 \\
&\vdots \\
A_{r_{q-1}+1} \cup \dots \cup A_{r_q} &= W_q, \quad \mathbb{P}(W_q) = \sum_{k=r_{q-1}+1}^{r_q} p_k = w_q.
\end{aligned}$$

Posons à présent $p_{r_{k-1}+1}|W_k, \dots, p_{r_k}|W_k$ les probabilités conditionnelles de $A_{r_{k-1}+1}, \dots, A_{r_k}$ sachant W_k , avec $r_0 = 0$. L'incertitude ajoutée par chaque groupe est alors $H(p_{r_{k-1}+1}|W_k, \dots, p_{r_k}|W_k)$, avec la probabilité w_k .

La connaissance des w_k et des $p_i|W_k$ est équivalente à celle des p_i seuls, et on aboutit ainsi au même résultat que l'on calcule l'entropie avec les probabilités conditionnelles ou les probabilités initiales :

$$\begin{aligned}
H(p_1, \dots, p_n) &= H(w_1, \dots, w_q) + w_1 H(p_1|w_1, \dots, p_{r_1}|w_1) \\
&\quad + \dots + w_q H(p_{r_{q-1}+1}|w_q, \dots, p_{r_q}),
\end{aligned} \tag{2}$$

Comme H doit être continue, il suffit de déterminer ses valeurs aux points rationnels $p_i = \frac{n_i}{n}$ avec les n_i entiers.

En imposant aux A_i d'être équiprobables, l'équation (2) devient :

$$h\left(n = \sum_{k=1}^q n_k\right) = H(w_1, \dots, w_q) + \sum_{k=1}^q w_k h(n_k) \tag{3}$$

en posant $h(m) = H(\frac{1}{m}, \dots, \frac{1}{m})$, entropie pour m alternatives équiprobables.

Choisissons alors tous les n_k égaux à r . Cela impose $w_k = \frac{r}{n}$, et l'équation précédente se réduit à :

$$h(n = qr) = h(q) + h(r).$$

La solution générale étant de la forme $h(n) = \alpha \ln(n)$, on injecte ce résultat dans l'équation (3) :

$$\alpha \ln(n) = H(w_1, \dots, w_q) + \alpha \sum_{k=1}^q w_k \ln(n_k).$$

Soit :

$$\begin{aligned} H(w_1, \dots, w_q) &= \ln \frac{n^\alpha}{\prod_{k=1}^q n_k^{\alpha w_k}} \\ &= \ln \frac{n^{\alpha \sum_{k=1}^q w_k}}{\prod_{k=1}^q n_k^{\alpha w_k}} \\ &= \ln \prod_{k=1}^q \left(\frac{n}{n_k} \right)^{\alpha w_k} \\ &= -\alpha \sum_{k=1}^q w_k \ln(w_k), \quad \text{car } \frac{n_k}{n} = w_k. \end{aligned}$$

On retombe ainsi sur la formule donnée en définition 1 en étendant la dernière égalité obtenue par continuité. Quand au cas continu il s'agit simplement d'une copie de la formule discrète, conservant certaines propriétés intéressantes.

□

On vient de démontrer le

Théorème 1 *Une fonction modélisant le manque d'information sur un système prenant des états discrets de probabilités p_k est nécessairement de la forme*

$$H = \alpha \sum p_k \ln(p_k).$$

Plusieurs nouvelles notions sont nécessaires pour parvenir au but des deux paragraphes qui vont suivre, qui est la définition d'indicateurs de la sensibilité du modèle à ses paramètres. Leur chronologie reprend celle du cours de M-J. Rendas ([7]).

2.2.2 Information mutuelle

L'entropie conditionnelle $H(Y|X)$ mesure le manque d'information moyen sur une variable aléatoire Y si on connaît le comportement d'une variable aléatoire X . Intuitivement, plus Y "dépend" de X , plus l'entropie conditionnelle sera faible, et inversement.

La définition est assez naturelle.

Définition 2 L'entropie conditionnelle de Y - à valeurs dans \mathcal{Y} - sachant X est :

$$H(Y|X) = \mathbb{E}_X[H(Y|x)],$$

où

$$H(Y|x) = - \sum_{y \in \mathcal{Y}} p_Y^{X=x}(y) \ln(p_Y^{X=x}(y))$$

dans le cas discret.

On notera plus simplement $p_Y^{X=x}(y) = p(y|x)$ lorsqu'aucune ambiguïté ne sera possible.

Dans le cas continu, il faut d'abord définir l'entropie.

Définition 3 L'entropie continue d'une variable aléatoire X à densité f est :

$$H(X) = - \int_{x \in \mathbb{R}} f(x) \ln(f(x)) dx,$$

avec toujours la convention $0 \ln(0) = 0$.

Il suffit alors de remplacer la somme par une intégrale :

$$H(Y|x) = - \int_{y \in \mathcal{Y}} f_Y^{X=x}(y) \ln(f_Y^{X=x}(y)) dy.$$

Par ailleurs, l'entropie d'une variable aléatoire vectorielle $X = (X_1, \dots, X_s)$ à valeurs dans \mathcal{X} se définit de façon analogue à la dimension 1 :

$$H(X) = - \sum_{x \in \mathcal{X}} f(x) \ln(f(x)) d\mu(x),$$

où f est la densité (resp. loi) de probabilité et μ la mesure sous-jacente.

On sait donc quantifier le manque d'information sur une variable aléatoire. Il est temps à présent d'exprimer son complémentaire. L'*information mutuelle* entre deux variables aléatoires X et Y mesure l'information apportée sur Y par X (resp. sur X par Y).

Définition 4 L'information mutuelle entre X et Y est :

$$I(X, Y) = H(X) + H(Y) - H(X, Y).$$

Remarque 1 Clairement $H(X, X) = H(X)$, et donc d'après la définition $I(X, X) = H(X)$. C'est cohérent avec ce que l'on souhaite représenter à travers I : l'information apportée à X par X est exactement $H(X)$; en effet connaissant X on comble tout le manque d'information sur cette dernière variable aléatoire, égal à $H(X)$.

Remarque 2 L'information mutuelle est symétrique : $I(X, Y) = I(Y, X)$; cette propriété peut se révéler intéressante, ne serait-ce que conceptuellement : tout comme le principe d'action-réaction en physique, il semble logique que l'information apportée par X sur Y soit la même que celle apportée par Y sur X .

En revanche la formule donnant l'influence de X sur Y n'est pas symétrique en X et Y dans le cas de la variance : $V(\mathbb{E}[Y|X]) \neq V(\mathbb{E}[X|Y])$ en général.

Par un calcul élémentaire on obtient la relation $H(X, Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$, qui permet d'écrire I sous une autre forme, non symétrique mais plus facile pour le calcul :

$$I(X, Y) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

Cherchons alors à écrire l'information mutuelle d'une manière encore plus compacte :

$$I(X, Y) = - \int_{y \in \mathcal{Y}} f(y) \ln(f(y)) d\mu(y) + \int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} f(y|x) f(x) \ln(f(y|x)) d\mu(x) d\mu(y).$$

Soit :

$$I(X, Y) = \int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} f(x, y) \ln \frac{f(x, y)}{f(x) f(y)} d\mu(x) d\mu(y), \quad (4)$$

μ étant (souvent) la mesure de Lebesgue (resp. de dénombrement) dans le cas continu (resp. discret).

2.2.3 Indices de sensibilité

L'expression (4) fait penser à la définition de l'entropie relative, également appelée *divergence* ou *distance de Kullback-Leibler*. Elle mesure l'information donnée par une distribution de probabilité relativement à une autre mesure.

Définition 5 *L'entropie relative $D(p : q)$ d'une mesure de probabilité p par rapport à une mesure de probabilité q est :*

$$D(p : q) = \int_{x \in \mathcal{X}} p(x) \ln \frac{p(x)}{q(x)} d\mu(x).$$

Propriété 1 *Positivité de l'entropie relative :*

$D(p : q)$ est une grandeur positive ou nulle.

Preuve :

L'inégalité $\ln(x) \leq x - 1$ permet d'écrire :

$$\begin{aligned} -D(p : q) &= \int_{x \in \mathcal{X}} p(x) \ln \frac{q(x)}{p(x)} d\mu(x) \\ &\leq \int_{x \in \mathcal{X}} p(x) \left(\frac{q(x)}{p(x)} - 1 \right) d\mu(x) \\ &= 0, \end{aligned}$$

car p et q sont des mesures de probabilité.

Donc D est dans \mathbb{R}_+ .

□

Remarque 3 *On peut utiliser cette dernière propriété pour montrer très rapidement que l'entropie est maximale uniquement pour la distribution uniforme U sur $[a, b]$:*

$$\begin{aligned} H(U) - H(f) &= \ln(b - a) + \int_a^b f(x) \ln(f(x)) dx \\ &= \int_a^b f(x) \ln((b - a) f(x)) dx \\ &= D(f : U) \end{aligned}$$

avec les notations évidentes.

Cette dernière quantité étant positive d'après ce qui précède, on a le résultat.

Il est tant alors d'énoncer la

Propriété 2 *Positivité de l'information mutuelle :*

$I(X, Y)$ est une grandeur positive ou nulle.

Preuve :

$I(X, Y)$ peut s'écrire à partir de la divergence D , en prenant p égale à la loi du couple (X, Y) et $q(x, y) = f(x)g(y)$ si les lois (ou densités) de X, Y s'écrivent respectivement f, g . q est bien une loi (ou densité) de probabilité, car

$$\int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} f(x) g(y) d\mu(x) d\mu(y) = 1,$$

et la démonstration est immédiate à partir de la relation (4).

□

Dans le cas discret l'entropie conditionnelle est positive, car tous les termes dans la somme la définissant le sont. On obtient alors en conséquence des considérations précédentes :

$$0 \leq \frac{I(X, Y)}{H(Y)} \leq 1,$$

avec égalité à 0 si X et Y sont indépendants, à 1 si X et Y sont totalement dépendants.

L'écriture de l'information mutuelle à partir de la divergence permet une nouvelle interprétation : $I(X, Y)$ représente en quelque sorte le degré d'indépendance de Y par rapport à X .

Bien sûr si X et Y sont des variables aléatoires continues il n'y a aucune raison pour que la fraction précédente soit dans $[0, 1]$. Mais ce n'est pas grave, car en pratique les lois continues sont approchées par des lois discrètes et on peut donc utiliser les résultats montrés ici.

De tout ce qui précède on déduit une définition possible d'un *indice de sensibilité basé sur l'entropie* à une variable X_i pour un modèle $(X_1, \dots, X_s) \rightarrow Y$ (Krzykacz-Hausmann [5]).

Définition 6 *L'indice*

$$\eta_i = \frac{I(X_i, Y)}{H(Y)} = 1 - \frac{H(Y|X_i)}{H(Y)}$$

représente l'information que l'on apporte sur Y par la connaissance de X .

C'est le calcul de cet indice qui va nous occuper au cours de cet exposé.

2.2.4 Extension au cas multivarié

Contrairement à ce qui se passe pour les indices calculés à partir de la variance, il est assez simple d'étendre la définition des indices de sensibilité au cas d'une sortie multi-dimensionnelle $Y = (Y_1, \dots, Y_d)$. En effet il suffit de reprendre l'écriture de l'information mutuelle sous la forme initiale

$$I(X, Y) = H(Y) - H(Y|X)$$

avec la définition naturelle

$$H(Y|X) = \mathbb{E}_X[H(Y|x)],$$

où

$$H(Y|x) = - \sum_{y \in \mathcal{Y}} p(y|x) \ln(p(y|x))$$

dans le cas discret, et

$$H(Y|x) = - \int_{y \in \mathcal{Y}} f(y|x) \ln(f(y|x)) d\mu(y)$$

dans le cas continu, μ étant par exemple la mesure de Lebesgue sur \mathbb{R}^d .

Après quelques calculs élémentaires :

$$I(X, Y) = \int_{x \in \mathcal{X}} \int_{y \in \mathcal{Y}} f(x, y) \ln \frac{f(x, y)}{f(x)f(y)} d\mu(x) d\hat{\mu}(y),$$

μ et $\hat{\mu}$ étant des mesures adéquates.

Dans le cas qui nous intéresse on calcule l'information mutuelle discrète

$$I(X, Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y_1, \dots, y_d) \ln \frac{p(x, y_1, \dots, y_d)}{p(x)p(y_1, \dots, y_d)}$$

avec les indices de sensibilités qui en découlent, dont la définition est la même que celle donnée au paragraphe 2.2.3.

Avec la variance on ne peut pas donner de cadre théorique rigoureux aux indices de sensibilité multi-sorties : il faut déterminer les indices sur chaque composante, puis, par exemple, en faire la somme.

2.3 Calculs d'entropie

On calcule ici les entropies de diverses lois continues d'usage courant en analyse d'incertitudes ainsi que de certaines lois discrètes, car les algorithmes calculeront de toutes façons des entropies discrètes.

2.3.1 Entropies de lois continues

Loi normale : (sa densité est représentée figure 1)

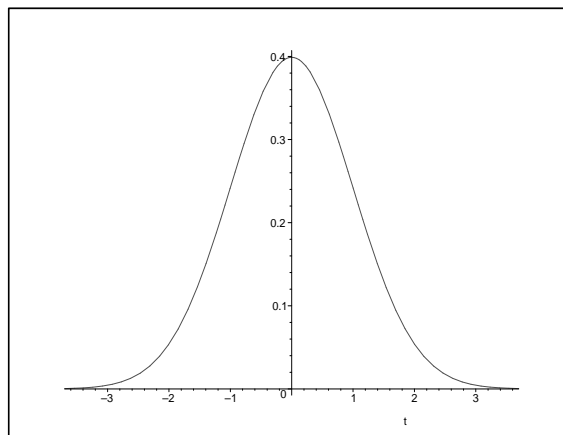


FIG. 1 – Loi normale, $f(x) = \frac{e^{-\frac{(x-m)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}$; ici $(m, \sigma) = (0, 1)$

L'entropie de la loi normale est :

$$\begin{aligned} H &= \frac{1}{2} + \ln(\sqrt{2\pi}\sigma) \\ &= \frac{1}{2} (1 + \ln(2\pi) + \ln(V)), \end{aligned}$$

V étant la variance et σ l'écart-type ($V = \sigma^2$).

Loi uniforme :

L'entropie de la loi uniforme sur $[a, b]$ est :

$$\begin{aligned} H &= \ln(b - a) \\ &= \frac{1}{2} (\ln(12) + \ln(V)), \end{aligned}$$

où $V = \frac{(b-a)^2}{12}$.

Loi trapézoïdale : (sa densité est représentée figure 2)

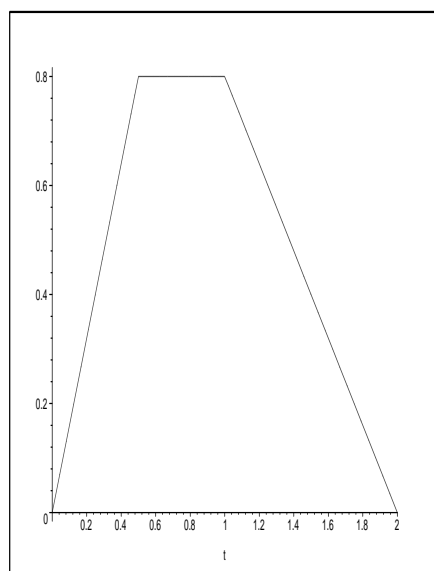


FIG. 2 – Loi trapézoïdale, $f(x) = h \frac{x-a}{b-a} \mathbb{1}_{[a,b]}(x) + h \mathbb{1}_{[b,c]}(x) + h \frac{d-x}{d-c} \mathbb{1}_{[c,d]}(x)$;
ici $(a, b, c, d) = (0, 0.5, 1, 2)$

L'entropie de la loi trapézoïdale sur $[a < b < c < d]$ est :

$$H = h \frac{d - c + b - a}{4} - \ln(h),$$

avec $h = \frac{2}{c+d-a-b}$.

Il est remarquable que l'on puisse réécrire ce résultat uniquement à l'aide des distances $\alpha = d - a$ et $\beta = c - b$:

$$H = -\ln(2) + \frac{1}{2} \frac{\alpha - \beta}{\alpha + \beta} + \ln(\alpha + \beta).$$

La variance peut s'exprimer à l'aide de a, b, c, d , mais après quelques calculs on se rend compte qu'elle n'a rien à voir avec l'expression de l'entropie.

Loi triangulaire : (sa densité est représentée figure 3)

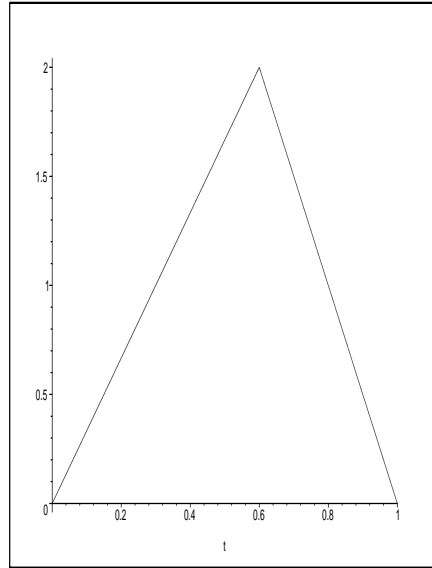


FIG. 3 – Loi triangulaire, $f(x) = h \frac{x-a}{b-a} \mathbb{1}_{[a,b]}(x) + h \frac{c-x}{c-b} \mathbb{1}_{[b,c]}(x)$;
ici $(a, b, c) = (0, 0.6, 1)$

L'entropie de la loi triangulaire est :

$$H = \frac{1}{2} - \ln(h),$$

avec $h = \frac{2}{c-a}$. On peut obtenir ce résultat à la limite $b = c$ du cas précédent. On constate que l'entropie ne dépend que du support de la loi, et pas de la position du mode d'abscisse b .

De même la variance peut être calculée :

$$V = \frac{1}{18} (a^2 + b^2 + c^2 - ab - bc - ca).$$

Loi log-normale : (sa densité est représentée figure 4)

L'entropie de la loi log-normale est :

$$H = \frac{1}{2} + e^{\frac{\sigma^2}{2}+m} \left(\frac{\ln(\sqrt{2\pi} \sigma)}{\sigma} + \sqrt{2\pi} \sigma (\sigma^2 + m) \right).$$

Ce dernier calcul a nécessité une intégration formelle à l'aide de Mathematica en ligne (<http://integrals.wolfram.com/index.jsp>).

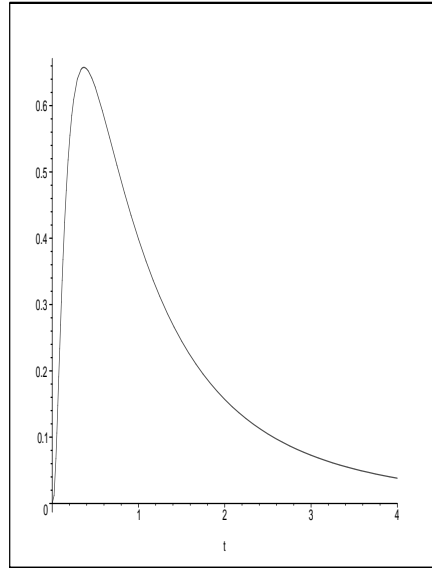


FIG. 4 – Loi log-normale, $f(x) = \frac{e^{-\frac{(\ln(x)-m)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma^2 x} \mathbb{1}_{\mathbb{R}_+}(x)$; ici $(m, \sigma) = (0, 1)$

La variance vaut :

$$V = e^{\sigma^2+2m} (e^{\sigma^2} - 1),$$

m étant la moyenne de $\ln(X)$ et σ son écart-type.

Loi log-uniforme : (sa densité est représentée figure 5)

L'entropie de la loi log-uniforme est :

$$H = \frac{\ln(a) + \ln(b)}{2} + \ln(\ln(b) - \ln(a)).$$

Quant à la variance :

$$V = \frac{1}{2} (4ab - 3a^2 - b^2).$$

Loi exponentielle : (sa densité est représentée figure 6)

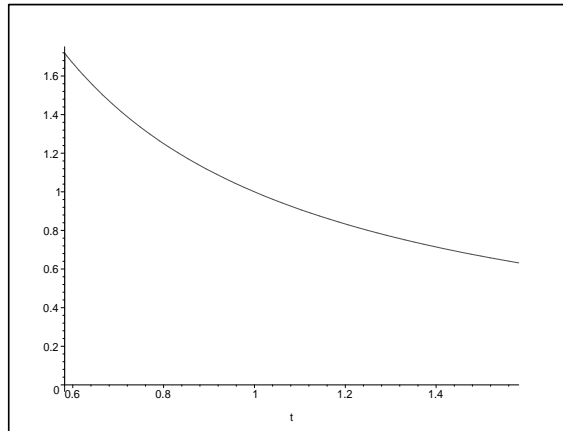


FIG. 5 – Loi log-uniforme, $f(x) = \frac{1}{(\ln(b)-\ln(a))x} \mathbb{1}_{[a,b]}(x)$; ici $(a, b) \simeq (0.582, 1.582)$

L'entropie de la loi exponentielle est :

$$\begin{aligned} H &= 1 - \ln(\lambda) \\ &= 1 + \frac{\ln(V)}{2}, \end{aligned}$$

où $V = \frac{1}{\lambda^2}$.

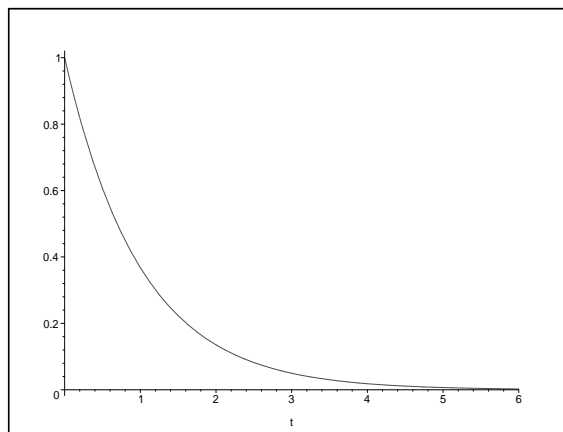


FIG. 6 – Loi exponentielle, $f(x) = \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x)$; ici $\lambda = 1$

Loi de Weibull : (sa densité est représentée figure 7)

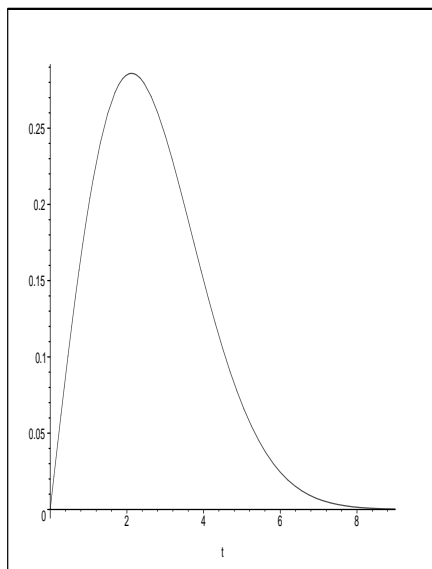


FIG. 7 – Loi de Weibull, $f(x) = \alpha \lambda x^{\alpha-1} e^{-\lambda x^\alpha} \mathbb{1}_{\mathbb{R}_+}(x)$; ici $(\lambda, \alpha) = (2, 3)$

L'entropie de la loi de Weibull est :

$$H = 1 - \ln(\lambda) - \ln(\alpha) - \frac{\alpha - 1}{\alpha} (\ln(\lambda) + \gamma),$$

avec la constante d'Euler-Mascheroni $\gamma = \int_{\mathbb{R}_{+*}} \ln(x) e^{-x} dx$.

Voici finalement la variance :

$$V = \lambda^{-\frac{2}{\alpha}} \left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right) \right).$$

On remarque que dans chaque cas où l'on peut exprimer H fonction de V , l'entropie croît logarithmiquement avec la variance. Plus exactement, on a asymptotiquement $H = \frac{\ln(V)}{2}$. L'entropie est donc en quelque sorte une mesure "moins sensible" que la variance.

2.3.2 Entropies de lois discrètes

Comme on le verra au paragraphe 2.5.3, il est plus intéressant d'utiliser l'entropie discrète à des fins d'analyse de sensibilité. Donnons quelques résultats dans le même esprit qu'en 2.3.1.

Loi uniforme :

C'est une des distributions de probabilités les plus simples ; son entropie se calcule immédiatement pour n états équiprobables :

$$\begin{aligned} H &= \ln(n) \\ &= \ln(\sqrt{12V + 1}), \end{aligned}$$

V étant la variance de cette même loi, valant $\frac{(b-a)^2}{12}$.

La relation entre l'entropie et la variance est similaire à celle du cas continu.

Loi de Bernoulli :

C'est la loi utilisée pour représenter un système à deux états de probabilités respectives p et q ; son entropie vaut :

$$H = -p \ln(p) - q \ln(q).$$

Elle est d'autant plus grande que p et q sont proches, atteignant son maximum pour $p = q = \frac{1}{2}$.

La variance vaut quant à elle :

$$V = p(1-p)(a^2 + b^2) - 2abpq,$$

les valeurs a et b ayant les probabilités respectives p et q .

Loi géométrique :

Il s'agit du temps d'attente du premier succès lors de la répétition de plusieurs épreuves d'une loi de Bernoulli. Une variable aléatoire X suivant la loi géométrique de paramètre $p \in [0, 1]$ vérifie

$$\forall k \in \mathbb{N} \quad , \quad \mathbb{P}(X = k) = p(1-p)^k.$$

Son entropie est égale à :

$$H = \ln(p) + \frac{p-1}{p} \ln(1-p),$$

et est représentée sur la figure 8.

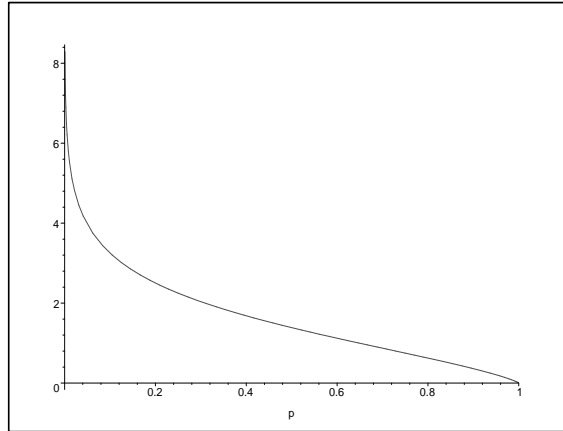


FIG. 8 – Entropie de la loi géométrique en fonction de $p \in [0, 1]$

La variance de cette même loi est :

$$V = \frac{1-p}{p^2}.$$

À titre indicatif on donne la répartition des probabilités en fonction de n pour $p = 0.6$ en figure 9.

Loi de Poisson :

On représente la fonction $n \mapsto \frac{\lambda^n}{\Gamma(n+1)} e^{-\lambda}$ pour $\lambda = 2$, $\Gamma(n+1) = n!$ lorsque n est entier (figure 10).

Une variable aléatoire X suivant la loi de Poisson de paramètre λ vérifie $\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$. Pour calculer l'entropie d'une telle loi il faut donc calculer la série

$$H(\lambda) = - \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!} e^{-\lambda} \ln \left(\frac{\lambda^n}{n!} e^{-\lambda} \right),$$

qui se transforme en

$$H(\lambda) = \lambda(1 - \ln(\lambda)) + e^{-\lambda} \sum_{n=0}^{+\infty} \lambda^n \frac{\ln(n!)}{n!}.$$

Cette dernière somme étant très difficile à expliciter, on évalue H graphiquement à l'aide de Maple; le résultat est visible figure 11.

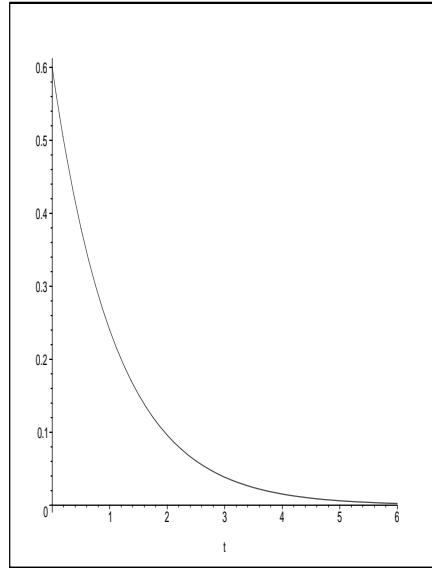


FIG. 9 – Loi géométrique

Remarque 4 *L'entropie croît logarithmiquement en λ ; or la variance est justement égale à λ : il semble que dans le cas discret l'entropie évolue aussi logarithmiquement avec la variance.*

Loi binômiale :

Elle représente les probabilités d'obtenir k succès lors de la répétition de n épreuves.

Traçons cette fonction en fonction de $k \in [0, n]$:

$k \mapsto \binom{n}{k} p^k (1-p)^{n-k}$ pour $p = 0.3$ et $n = 50$ (figure 12).

L'entropie d'une telle loi se calcule avec une série similaire à la précédente. Cette fois c'est la présence des coefficients binômiaux $\binom{n}{k}$ qui rend le calcul exact impossible. On évalue donc H graphiquement pour $n = 50$ (figure 13).

La courbe de l'entropie est symétrique par rapport à la droite $x = \frac{1}{2}$ (figure 13) ; c'est logique car les valeurs p et $1-p$ conduisent aux mêmes distributions de probabilité, en intervertissant les rôles des succès et des échecs. Comme pour l'entropie de la loi de Poisson on remarque un début de croissance logarithmique avec p . Or ici la variance est encore polynômiale en p .

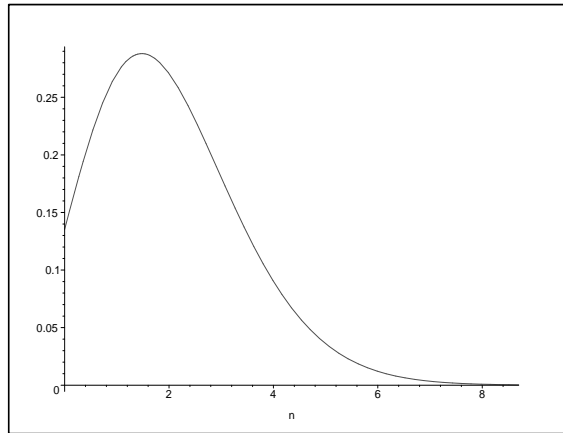


FIG. 10 – Loi de Poisson

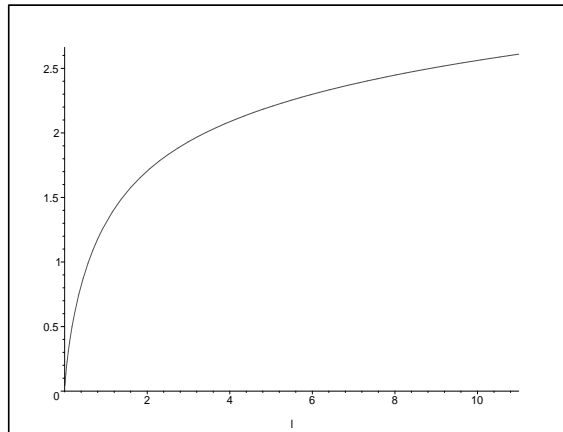


FIG. 11 – Entropie de la loi de Poisson en fonction de $\lambda \in \mathbb{R}_+$

La variance vaut dans ce cas :

$$V = n p (1 - p).$$

En ce qui concerne la loi binômiale négative, Maple ne parvient pas à tracer la courbe de l'entropie en fonction de p en temps raisonnable.

En conclusion on peut dire que l'entropie des lois discrètes croît elle aussi logarithmiquement avec la variance, ce qui laisse présager une convergence (ou divergence..) logarithmique de l'entropie dans le cadre des algorithmes de la section 3.1.

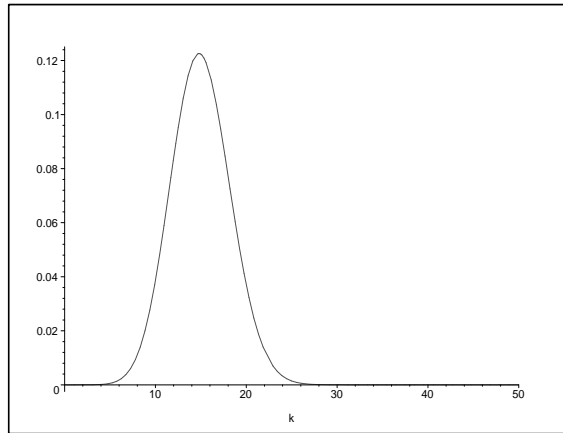


FIG. 12 – Loi binômiale

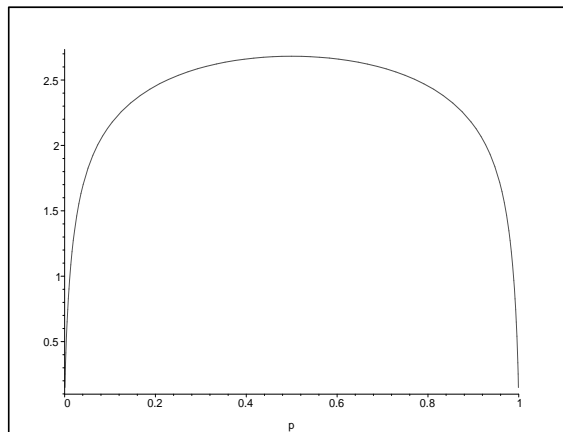


FIG. 13 – Entropie de la loi binômiale en fonction de $p \in [0, 1]$

2.3.3 Comparaison entropie discrète - entropie continue

Ce paragraphe a pour but de comparer numériquement les entropies discrètes et continues sur trois lois simples. On discrétise naturellement une variable aléatoire en divisant son intervalle de variation en n sous-intervalles de même largeur, et on calcule son entropie discrète en sommant les probabilités d'appartenance à chaque sous-intervalle multipliées par leur logarithme.

On choisit des lois à support compact : ainsi l'algorithme ne travaille pas sur des données approchées. Les tests se font sur l'intervalle $[0,1]$ avec $n = 300$, pour 100 répétitions du calcul avec échantillonnage par hypercube latin. La loi triangulaire est symétrique et la loi trapézoïdale a un plateau de

	$S = 1000$	$S = 5000$	$S = 10000$	$S = 50000$	$S = 100000$
<i>i</i>)	$H = 5.538$ $\sigma = 1.23e-2$	$H = 5.673$ $\sigma = 2.32e-3$	$H = 5.688$ $\sigma = 1.17e-3$	$H = 5.701$ $\sigma = 2.41e-4$	$H = 5.702$ $\sigma = 1.23e-4$
<i>ii</i>)	$H = 5.448$ $\sigma = 1.97e-2$	$H = 5.567$ $\sigma = 8.43e-3$	$H = 5.578$ $\sigma = 4.85e-3$	$H = 5.584$ $\sigma = 2.58e-3$	$H = 5.585$ $\sigma = 1.67e-3$
<i>iii</i>)	$H = 5.378$ $\sigma = 2.10e-2$	$H = 5.497$ $\sigma = 9.50e-3$	$H = 5.506$ $\sigma = 5.80e-3$	$H = 5.513$ $\sigma = 3.23e-3$	$H = 5.513$ $\sigma = 2.34e-3$

TAB. 1 – Entropie H en fonction du nombre de points S dans l'échantillon simulé ; *i*) : loi uniforme, *ii*) : loi trapézoïdale, *iii*) : loi triangulaire

$\frac{1}{4}$ à $\frac{3}{4}$.

Le tableau 1 présente les résultats obtenus, " $e^{-\alpha}$ " signifiant $\times 10^{-\alpha}$, σ étant l'écart-type obtenu sur les 100 répétitions.

On remarque globalement une convergence lente : il faut de l'ordre de 50000 points pour obtenir 3 digits dans les cas les plus favorables ci-dessus. En effet, on peut calculer analytiquement les indices discrets dans ces cas simples.

Loi uniforme :

$$\begin{aligned}
 H &= - \sum_{k=1}^{300} \frac{1}{300} \ln \frac{1}{300} \\
 &= \ln(300) \\
 &\simeq 5.703782.
 \end{aligned}$$

Loi trapézoïdale :

$$\begin{aligned}
 H &= -2 \sum_{k=1}^{75} \frac{1}{300} \frac{16}{3} \frac{2k-1}{600} \ln \left(\frac{1}{300} \frac{16}{3} \frac{2k-1}{600} \right) - 150 \frac{4}{3} \frac{1}{300} \ln \left(\frac{4}{3} \frac{1}{300} \right) \\
 &\simeq 5.5828,
 \end{aligned}$$

où $\frac{1}{300}$ est un coefficient de normalisation, $\frac{16}{3}$ étant la valeur absolue de la pente à droite et à gauche du trapèze ; on échantillonne avec $\frac{2k-1}{600}$ au lieu de $\frac{k}{300}$ afin de choisir les points aux centres des intervalles de discrétisation.

Loi triangulaire :

$$H = -2 \sum_{k=1}^{150} \frac{1}{150} \frac{2k-1}{300} \ln\left(\frac{1}{150} \frac{2k-1}{300}\right) \\ \simeq 5.510662,$$

où $\frac{1}{150}$ est un coefficient de normalisation des probabilités.

Pour les mêmes données, on obtient dans le cas continu :

Loi uniforme : $H = 0$.

Loi trapézoïdale : $H = -\ln \frac{4}{3} + \frac{1}{6} \simeq -0.121015$.

Loi triangulaire : $H = \frac{1}{2} - \ln(2) \simeq -0.193147$.

On constate donc comme prévu que les résultats n'ont rien à voir, mais restent tout de même cohérents avec l'intuition : une loi triangulaire apporte plus d'information qu'une loi trapézoïdale, qui elle-même apporte plus d'information qu'une loi uniforme.

2.4 Bornes sur la variance et l'entropie

On cherche ici à borner inférieurement et supérieurement les deux mesures étudiées jusqu'alors.

2.4.1 Maximisation de la variance

En analyse d'incertitudes la distribution uniforme représente souvent le manque total de connaissances : on souhaite donc qu'une mesure de manque d'information soit maximale dans ce cas ; or ce n'est pas vrai pour la variance, pas plus dans le cas discret que dans le cas continu.

Théorème 2 *Parmi toutes les variables aléatoires discrètes à valeurs dans $\{x_1, \dots, x_n\}$, celle qui a la plus grande variance est l'unique variable aléatoire vérifiant $\mathbb{P}(X = \min_{i \in [1, n]} \{x_i\}) = \mathbb{P}(X = \max_{i \in [1, n]} \{x_i\}) = \frac{1}{2}$.*

Preuve :

Soit $n \in \mathbb{N}^*$. Les variables aléatoires sont à valeurs dans $\{x_1, \dots, x_n\}$.

On veut maximiser

$$V(X) = \sum_{k=1}^n x_k^2 p_k - \left(\sum_{k=1}^n x_k p_k \right)^2 = V(p)$$

avec $p_k = \mathbb{P}(X = x_k)$, sous les contraintes $\sum_{k=1}^n p_k = 1$ et $\forall k \in \mathbb{N}_n p_k \geq 0$.
Soit alors le lagrangien

$$L(p, \lambda, \lambda', \mu) = \lambda V(p) + \lambda' \left(\sum_{k=1}^n p_k - 1 \right) + \sum_{k=1}^n \mu_k p_k .$$

Ses dérivées partielles par rapport aux p_k

$$\frac{\partial L}{\partial p_k}(p, \lambda, \lambda', \mu) = \lambda \left(x_k^2 - 2x_k \sum_{i=1}^n x_i p_i \right) + \lambda' + \mu_k$$

doivent être nulles en un optimum d'après le théorème de Karush-Kuhn-Tucker.

Si $\lambda = 0$ on aboutit à une solution déterministe où un des p_k vaut 1, les autres étant égaux à 0. Bien sûr c'est un minimum global, car la variance est toujours positive ; on choisit donc $\lambda = 1$ (possible par homogénéité).

Le système suivant en résulte :

$$\begin{aligned} \forall k \in \mathbb{N}_n \quad 2x_k \sum_{i=1}^n x_i p_i &= \lambda' + \mu_k + \mu'_k + x_k^2 \\ \forall k \in \mathbb{N}_n \quad \mu_k p_k &= 0 \\ \sum_{k=1}^n p_k - 1 &= 0, \end{aligned}$$

à $2n + 1$ équations pour $2n + 1$ inconnues.

Posons alors $m = \sum_{k=1}^n x_k p_k$. On a, pour tout k tel que $p_k > 0$:

$$sx_k^2 - 2x_k m + \lambda' = 0, \tag{5}$$

équation qui ne peut être satisfaite qu'en deux valeurs de k , distinctes si $\lambda' \neq m^2$. Donc exactement deux p_k sont non nuls (s'il n'y en avait qu'un la variance serait nulle). On se ramène au problème plus simple de maximiser

$$x_k^2 p_k + x_l^2 p_l - (x_k p_k + x_l p_l)^2 = x_k^2 (p_k - p_k^2) + x_l^2 (p_l - p_l^2) - 2x_k x_l p_k p_l$$

sous les contraintes $p_k + p_l = 1$ et $p_k, p_l \in [0, 1]$, avec $k \neq l$.

λ' doit être choisi de manière à ce que l'équation (5) soit satisfaite par les valeurs optimales x_k et x_l , qui d'après cette même équation sont symétriques par rapport à m . S'inspirant de cette dernière remarque on pose $x_l = 2m - x_k$, puis il ne reste plus qu'à maximiser une fonction d'une seule variable définie sur $[0, 1]$:

$$\max_{p_k \in [0, 1]} (x_k - m)^2 (p_k - p_k^2),$$

après simplifications.

Cette dernière expression est clairement maximale pour $x_k \in \{\min_{i \in [1, n]} \{x_i\}, \max_{i \in [1, n]} \{x_i\}\}$ à p_k fixé, puis pour $p_k = \frac{1}{2}$ lorsque k est fixé.

□

En revanche dans le cas continu le maximum n'est pas atteint, comme on le montre ci-dessous.

Soient $a < b \in \mathbb{R}$. On cherche à maximiser la fonction variance :

$$\sup_{X \in \mathcal{D}} \int_a^b (x - \mathbb{E}[X])^2 f_X(x) dx$$

sur l'ensemble \mathcal{D} des variables aléatoires réelles X à densité sur $[a, b]$.

Pour une telle variable aléatoire de densité f on note

$$V(f) = \int_a^b x^2 f(x) dx - \left(\int_a^b x f(x) dx \right)^2,$$

expression qui doit donc être maximisée sous la contrainte $\int_a^b f(x) dx = 1$.

On introduit alors le lagrangien

$$L(f, \lambda, \lambda') = \lambda V(f) + \lambda' \left(\int_a^b f(x) dx - 1 \right).$$

Soient f une densité sur $[a, b]$ et δf une "petite" perturbation telle que $f + \delta f$ soit encore une densité.

Alors, en négligeant les termes au delà de l'ordre 1 :

$$L(f + \delta f, \lambda, \lambda') - L(f, \lambda, \lambda') = \int_a^b \left(\lambda x^2 - 2\lambda x \int_a^b t f(t) dt + \lambda' \right) \delta f(x) dx.$$

D'où

$$\frac{\delta L}{\delta f}(x) = \lambda x^2 - 2\lambda x \int_a^b t f(t) dt + \lambda'.$$

Un optimum est atteint si cette dérivée fonctionnelle s'annule sur tout le support de f . Ce n'est possible que si f prend un nombre fini de valeurs (λ ne pouvant s'annuler, car les deux multiplicateurs de Lagrange ne peuvent s'annuler simultanément), et l'optimum est donc atteint pour une variable aléatoire discrète : on est ramené au cas précédent.

Cependant, pour s'approcher le plus possible de l'optimum discret dans le cas continu on peut utiliser la fonction suivante définie sur $[a, b]$, avec $\varepsilon \in]0, \frac{b-a}{2}[$:

$$f_\varepsilon(x) = \begin{cases} \frac{1}{2\varepsilon} & \text{si } x \in [a, a + \varepsilon] \cup [b - \varepsilon, b], \\ 0 & \text{sinon.} \end{cases}$$

Après calculs $V(f_\varepsilon) = \frac{(b-a)^2}{4} + O(\varepsilon)$, et la variance est maximale pour une densité qui présente deux pics de Dirac équiprobables en a et b respectivement.

Ce dernier exemple fait penser à une situation où deux distributions bien différentes du point de vue de l'information apportée ont la même variance, laissant présager que cette dernière quantité n'est pas toujours adaptée : les densités tracées sur la figure 14 en sont une illustration, d'équations respectives $f(x) = \mathbb{1}_{[-1, -\frac{1}{2}]}(x) + \mathbb{1}_{[\frac{1}{2}, 1]}(x)$ (à gauche) et $f(x) = 5 \mathbb{1}_{[-\alpha, -\alpha + \frac{1}{10}]}(x) + 5 \mathbb{1}_{[\alpha - \frac{1}{10}, \alpha]}(x)$; $\alpha \simeq 0.8132$ (à droite).

2.4.2 Maximisation de l'entropie

Contrairement à la variance, l'entropie est maximale pour une distribution uniforme. Elle est donc plus susceptible de modéliser l'incertitude, qui intuitivement devrait être maximale lorsque toutes les valeurs possibles sont équiprobables.

Théorème 3 Soient $x_1, \dots, x_n \in \mathbb{R}$.

Parmi toutes les distributions de probabilités discrètes à valeurs dans $\{x_1, \dots, x_n\}$, celle qui a la plus grande entropie est la loi uniforme.

Preuve :

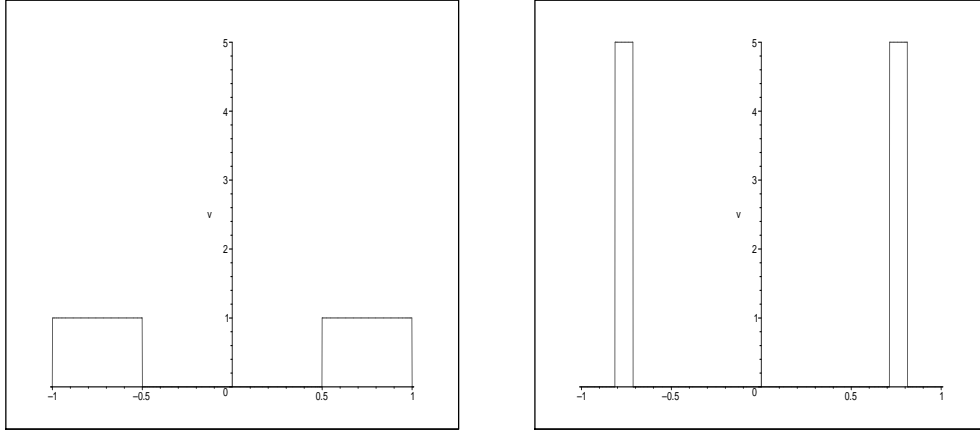


FIG. 14 – Deux distributions de même variance. Information floue à gauche (a), plus précise à droite (b).

On cherche à maximiser $H(p = (p_1, \dots, p_n))$ sous la contrainte $\sum_{k=1}^n p_k = 1$ (il est inutile ici de demander aux p_k d'être positifs, la présence du logarithme impose déjà cette condition).

Une nouvelle fois, un lagrangien apparaît :

$$L(p, \lambda, \lambda') = \lambda H(p) + \lambda' \left(\sum_{k=1}^n p_k - 1 \right),$$

dont les dérivées partielles par rapport aux p_k s'écrivent

$$\frac{\partial H}{\partial p_k}(p) = -\lambda(\ln(p_k) + 1) + \lambda'.$$

On en déduit en cherchant à annuler l'expression précédente qu'à l'optimum tous les p_k doivent avoir la même valeur.

□

Il est remarquable que cette propriété soit conservée dans le cas continu, sans la dégénérescence observée dans le cas de la variance : l'entropie continue est maximale pour une loi uniforme.

Preuve :

On souhaite maximiser la fonction entropie :

$$H(f) = \int_a^b f(x) \ln(f(x)) dx$$

sous la contrainte $\int_a^b f(x) dx = 1$.

On introduit le lagrangien

$$L(f, \lambda, \lambda') = \lambda H(f) + \lambda' \left(\int_a^b f(x) dx - 1 \right).$$

Soit δf une "petite" perturbation telle que $f + \delta f$ soit encore une densité. Ecrivons la différence $\delta L = L(f + \delta f, \lambda, \lambda') - L(f, \lambda, \lambda')$:

$$\delta L = - \int_a^b \lambda f(x) (\ln(f(x) + \delta f(x)) - \ln(f(x))) + \lambda \delta f(x) \ln(f(x) + \delta f(x)) + \lambda' \delta f(x) dx.$$

Il suffit alors d'effectuer un développement limité du logarithme pour aboutir à l'expression suivante, en négligeant les termes au delà de l'ordre 1 :

$$\delta L = - \int_a^b \delta f(x) \left(-\lambda (1 + \ln(f(x))) + \frac{\lambda}{f(x)} + \lambda' \right) dx,$$

puis

$$\frac{\delta L}{\delta f}(x) = -\lambda (1 + \ln(f(x))) + \frac{\lambda}{f(x)} + \lambda'.$$

Un optimum est atteint si cette dérivée fonctionnelle s'annule sur tout le support de f . Il faut donc en particulier que $-1 - \ln(f(x)) + \frac{1}{f(x)}$ soit une constante (comme dans le cas de la variance λ ne peut s'annuler), et par conséquent que f soit constante, la fonction $x \in]0, 1] \mapsto -1 - \ln(x) + \frac{1}{x}$ étant bijective. On conclut grâce à la condition de normalisation que f doit être constante égale à $\frac{1}{b-a}$, ce qui est exactement la définition de la loi uniforme dans le cas continu.

□

On généralise ainsi le théorème 3.

Théorème 4 *L'entropie est maximale pour des variables aléatoires suivant des lois uniformes*

En particulier, l'entropie est toujours négative pour des variables aléatoires dont le support est de largeur ≤ 1 .

Fixons $a < b \in \mathbb{R}$, et soit X_n la variable aléatoire discrète sur $[a, b]$ qui associe la probabilité $\frac{1}{n}$ à chaque élément d'un ensemble de n points choisis au hasard dans $[a, b]$. L'entropie de X_n est alors $\ln(n)$. On pourrait penser que la limite du cas discret est le cas continu, comme souvent en analyse, mais ici se produit un phénomène remarquable : il y a un saut totalement discontinu de la limite du cas uniforme discret ($+\infty$) au cas uniforme continu ($\ln(b-a)$). Ceci donne une idée des écarts existants entre entropie discrète et continue.

2.4.3 Borne inférieure de l'entropie

La valeur maximale de l'entropie est $\ln \frac{1}{b-a}$, mais on ne s'est jusqu'ici pas intéressé à sa borne inférieure sur $[a, b]$ (pour la variance c'est clairement 0). Intuitivement l'entropie devrait être minimale lorsqu'on a une parfaite connaissance de l'information, ce qui se traduit sur une densité par un pic de Dirac. Il suffit pour s'en convaincre de considérer la fonction suivante :

$$f_\varepsilon(x) = \begin{cases} \frac{1-\varepsilon+\varepsilon^2}{\varepsilon} & \text{si } x \in [a, a + \varepsilon[, \\ \varepsilon & \text{si } x \in [a + \varepsilon, b]. \end{cases}$$

On a $H(f_\varepsilon) = (-\varepsilon + \varepsilon^2) \ln(\varepsilon) - (1 - \varepsilon + \varepsilon^2) (\ln(1 - \varepsilon + \varepsilon^2) - \ln(\varepsilon)) \sim \ln(\varepsilon)$, qui tend vers $-\infty$ quand ε tend vers 0. Regardons alors la forme des fonctions de variance minimale.

Si une variable aléatoire X a une variance nulle, cela signifie que l'intégrale suivante est nulle

$$\int (x - \mathbb{E}[X])^2 f(x) dx.$$

La fonction intégrée étant positive ou nulle en tout point, on en déduit par continuité qu'elle est identiquement nulle : $f(x) = 0$ partout sauf en $x = \mathbb{E}[X]$, et donc X est déterministe : il existe un élément r tel que $\mathbb{P}(X = r) = 1$.

En extrapolant sur l'exemple introductif, on remarque que l'entropie continue vaut $-\infty$ seulement pour des combinaisons de pics de Dirac, qui correspondent alors à des distributions discrètes dont on calcule l'entropie en tenant compte des valeurs que la variable ne peut jamais prendre (cela revient à calculer l'entropie d'une variable aléatoire discrète en considérant

qu'elle prend ses valeurs dans un ensemble continu).

Un phénomène remarquable se produit : la variance est maximale sur $[a, b]$ pour la distribution $\frac{1}{2} \delta_a + \frac{1}{2} \delta_b$, alors que l'entropie est minimale pour cette même distribution !

2.5 Comportements de la variance et de l'entropie

On s'intéresse ici aux comportements de la variance et de l'entropie continue sur certaines lois simples, afin de mieux cerner leurs variations.

2.5.1 Évolution qualitative variance - entropie

L'idée ici est de regarder comment la variance (resp. l'entropie) séparent les distributions donnant plus ou moins d'information.

Commençons par la variance en étudiant la famille de densités de probabilité

$$\left\{ f_n(x) = \frac{(2n+1)x^{2n}}{2} \right\}_{n \in \mathbb{N}}$$

définie sur $[-1, 1]$ (figure 15).

Un rapide calcul donne

$$V(f_n) = \frac{2n+1}{2n+3}.$$

C'est cohérent : la variance vaut $\frac{1}{3}$ pour $n = 0$, et dans ce cas f_n est la loi uniforme. Ensuite, on observe que $V(f_n)$ croît très rapidement vers 1, qui est la variance limite obtenue pour deux pics delta de Dirac équiprobables en a et b : c'est logique car alors x^{2n} forme un graphe de plus en plus "aplati" sur $[-1 + \varepsilon, 1 - \varepsilon]$, et prend donc ses valeurs non triviales principalement au voisinage de -1 et 1.

L'entropie vaut quant à elle sur cette même famille :

$$H(f_n) = \ln(2) - \ln(2n+1) + \frac{2n}{2n+1}.$$

On obtient bien $\ln(2)$ en $n = 0$: c'est l'entropie de la distribution uniforme. Ensuite, opposément à la variance l'entropie décroît logarithmiquement, estimant que plus la masse totale de f_n est regroupée plus on a d'information.

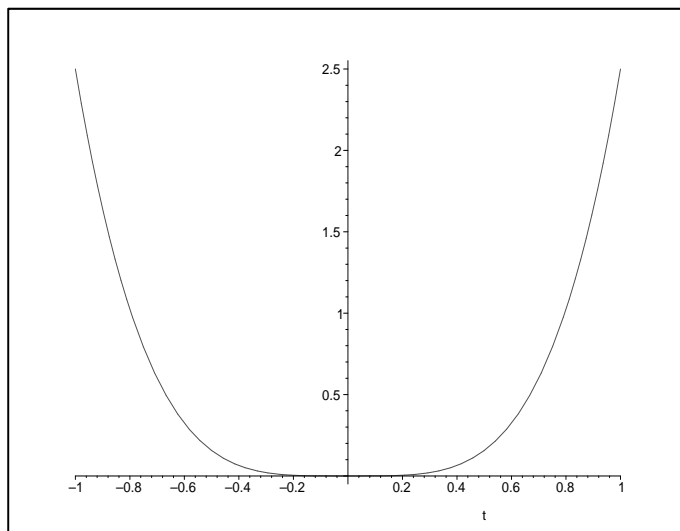


FIG. 15 – Famille polynômiale "aplatie"; ici $n = 2$

Jusque-là tout va bien, nos deux grandeurs se lançant sur des voies différentes mais cohérentes.

Essayons alors la famille suivante :

$$g_n(x) = \begin{cases} (n^2x + n) \mathbb{1}_{]-\frac{1}{n}, 0]}(x) + (-n^2x + n) \mathbb{1}_{]0, \frac{1}{n}]}(x) \}_{n \in \mathbb{N}^*} & \text{si } n > 0, \\ \frac{\mathbb{1}_{[-1, 1]}(x)}{2} & \text{si } n = 0, \end{cases}$$

définie elle aussi sur $[-1, 1]$ (figure 16).

Il s'agit simplement d'une suite de triangles de plus en plus hauts, et de moins en moins larges.

On obtient après calculs :

$$V(g_n) = \begin{cases} \frac{1}{6n^2} & \text{si } n > 0, \\ \frac{1}{3} & \text{si } n = 0. \end{cases}$$

Cette dernière quantité est toujours inférieure à $\frac{1}{3}$, conformément à l'intuition : plus le triangle est haut, plus l'incertitude maximale que l'on peut avoir sur le système diminue ; en effet celle-ci est égale à la largeur du triangle. À la limite, on est sûr que l'on est dans l'état représenté par 0, et la

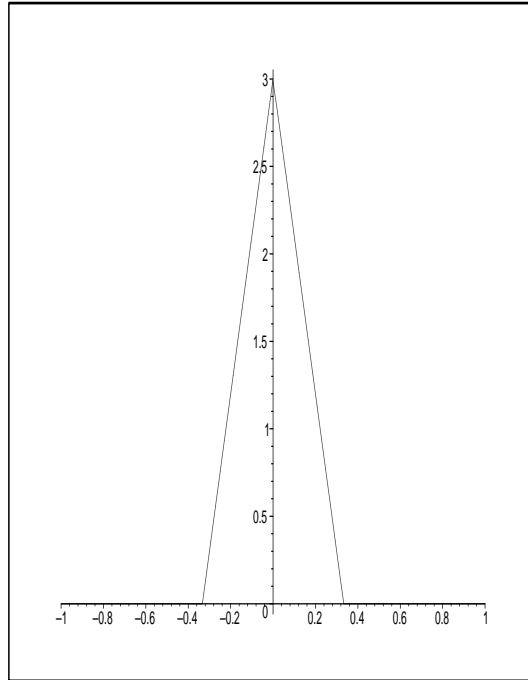


FIG. 16 – Famille des triangles centrés ; ici $n = 3$

variance est alors nulle.

Et dans le cas de l'entropie :

$$H(g_n) = \begin{cases} \frac{1}{2} - \ln(n) & \text{si } n > 0, \\ \ln(2) & \text{si } n = 0. \end{cases}$$

Comme pour la variance, on vérifie la cohérence : l'entropie tend bien vers $-\infty$ lorsque la hauteur du triangle tend vers $+\infty$.

On voit cependant apparaître une anomalie par rapport à la variance, qui n'en est pas une si l'on se place du point de vue de l'entropie mais qui mérite d'être signalée : la fonction $(x, y) \mapsto \ln \frac{2x}{2y+1} + \frac{2y}{2y+1} - \frac{1}{2}$ définie sur \mathbb{R}_{+*}^2 admet clairement une infinité de racines (fixer y quelconque puis ajuster x), donc, intuitivement, "par densité des parties fractionnaires des solutions dans $[0, 1]^2$ " - le raisonnement correct étant en annexe -, pour tout $\varepsilon > 0$ il existe une paire $(n, m) \in \mathbb{N}^2$ telle que $H(f_n) = H(g_m) + \varepsilon$. Ainsi l'entropie ne distingue pas les cas où l'information se regroupe en un point de ceux où elle se "regroupe" en deux points (ou plus : on pourrait généraliser cet exemple).

Ce résultat a déjà été indiqué au paragraphe 2.4.3, et trouve ici une illustration pratique.

Continuons notre exploration avec la famille "gaussienne tronquée" sur $[-1,1]$:

$$\left\{ h_\sigma(x) = \frac{1}{\int_{-1}^1 e^{-\frac{t^2}{2\sigma^2}} dt} e^{-\frac{x^2}{2\sigma^2}} \right\}_{\sigma \in \mathbb{R}_{+*}} .$$

Après calculs :

$$V(h_\sigma) = \sigma^2 - \sigma \sqrt{\frac{2}{\pi}} \frac{e^{-\frac{1}{2\sigma^2}}}{\operatorname{erf}\left(\frac{1}{\sqrt{2}\sigma}\right)},$$

et

$$H(h_\sigma) = \frac{1}{2} + \ln\left(\sqrt{2\pi} \sigma \operatorname{erf}\left(\frac{1}{\sqrt{2}\sigma}\right)\right) - \frac{e^{-\frac{1}{2\sigma^2}}}{\sqrt{2\pi} \sigma \operatorname{erf}\left(\frac{1}{\sqrt{2}\sigma}\right)} .$$

Plus σ augmente, plus h_σ tend vers une loi uniforme ; on vérifie alors graphiquement que la variance tend bien vers $\frac{1}{3}$ (figure 17) et l'entropie vers $\ln(2)$ (figure 18). Si à l'inverse σ diminue, h_σ tend vers un pic de Dirac et on vérifie graphiquement une fois encore que $V \rightarrow 0$ tandis que $H \rightarrow -\infty$.

Conformément au cas similaire du triangle, la variance reste inférieure à $\frac{1}{3}$, et évolue en croissant de $\sigma = 0$ à $\sigma = +\infty$. Dans le cas de l'entropie, on a comme précédemment une répartition dans $]-\infty, \ln(2)]$, les valeurs négatives n'étant cependant prises que sur un voisinage de l'origine très restreint.

On remarque que ces deux courbes sont d'allure très similaire : dans ce cas variance et entropie ont à peu près le même comportement, ce qui n'était pas vrai précédemment (et n'est peut-être qu'une coïncidence).

2.5.2 Surfaces "iso-information"

Ce qui précède laisse à penser que les densités à forte variance sont celles qui répartissent les poids assez équitablement de part et d'autre de la moyenne, le plus éloigné possible de cette dernière. La variance diminue ensuite si l'on rapproche les poids vers la moyenne, ou si on les déséquilibre à un point tel que la "balance probabiliste" penche suffisamment d'un côté pour donner de bons intervalles de confiance.

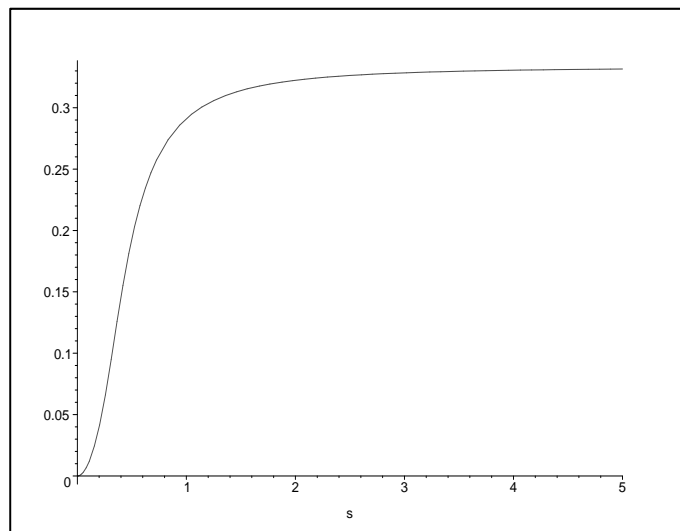


FIG. 17 – Variance de la gaussienne tronquée

À un stade intermédiaire se situe la loi uniforme, correspondant à un rapprochement vers la moyenne, non déséquilibré. Le "déséquilibre" maximum est ensuite obtenu par un pic delta de Dirac sur la moyenne, ce qui correspond à une information déterministe.

L'entropie en revanche est plus souple en ce qui concerne les distributions "à information maximale" : dès que la densité tend vers une loi discrète, l'entropie tend vers $-\infty$. Pour elle, on a le moins d'information lorsque chaque valeur de l'intervalle $[a, b]$ est équiprobable, ce qui est une autre vision des choses.

Peut-on dire quelque chose de précis sur les propriétés que doivent vérifier les densités f vérifiant l'équation $H(f) = h$, ou $V(f) = v$? Idéalement on souhaiterait déterminer l'ensemble des fonctions f solutions ; malheureusement cette quête semble hors de portée, l'ensemble des fonctions densité vérifiant ces contraintes étant beaucoup trop vaste comme on le voit au cours de ce travail.

Qu'en est-il de l'exemple donné en fin de paragraphe 2.4.1 ? On se rend rapidement compte que l'entropie est plus grande dans le cas de la figure 14 a. . Mais celle-ci a un autre défaut : elle ne distingue pas les deux distributions suivantes (figure 19). Ce défaut provient du fait que l'entropie ne dépend que de la distribution de probabilité et non pas des valeurs prises par la variable

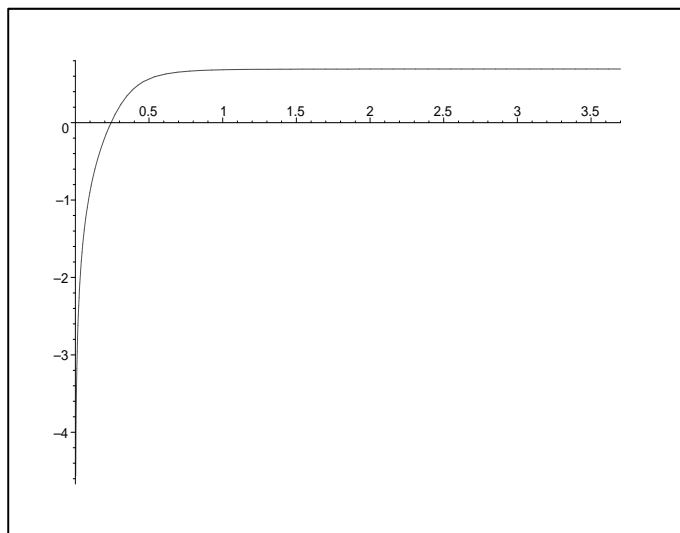


FIG. 18 – Entropie de la gaussienne tronquée

aléatoire sous-jacente, entraînant une insensibilité totale aux écarts entre les deux moitiés de la figure 19 *a.* ; on généralise sans mal à d'autres distributions présentant la même caractéristique.

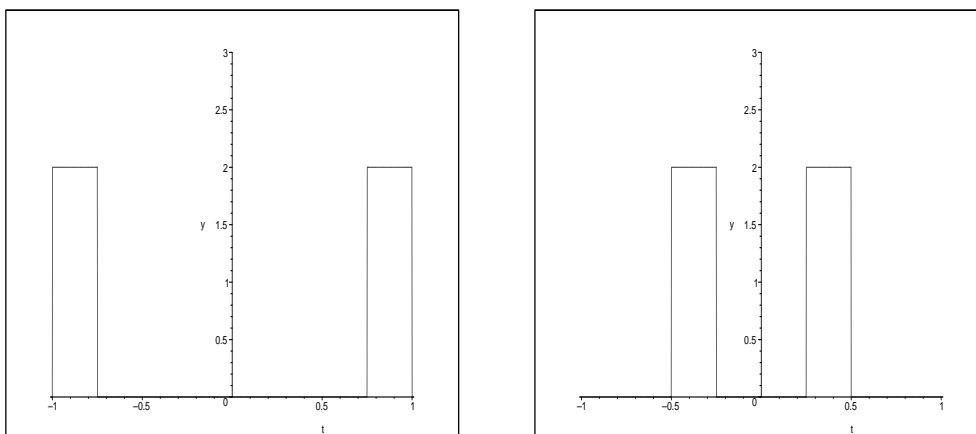


FIG. 19 – Deux distributions de même entropie. Distribution de probabilité excentrée à gauche (*a.*), centrée à droite (*b.*).

Or il est clair qu'on a plus d'information dans le cas de la figure 19 *b.*, où

l'on peut donner un plus petit intervalle de confiance. La variance distinguerait ces deux situations : il semble donc que l'entropie et la variance soient des entités complémentaires. Il resterait à définir une formule de "super-indices" utilisant ces deux notions (voir le paragraphe 2.5.3).

En conclusion, on peut faire le comparatif partiel suivant.

Variance :

- Elle est maximale pour des densités f qui prennent une grande plage de valeurs aux voisinages de deux points distincts, les plus éloignés possible de la moyenne et entre eux ; optimum atteint pour deux pics de Dirac le plus loin possible de part et d'autre de la moyenne.
- Elle diminue ensuite en ramenant les poids des pics de Dirac vers la moyenne de la variable aléatoire X sous-jacente ; la valeur de la densité f en $\mathbb{E}[X]$ semble être un assez bon indicateur en pratique.
- Elle est minimale pour des densités f qui prennent une grande plage de valeurs au voisinage d'un unique point, et qui ont leurs valeurs dans un petit intervalle ailleurs ; optimum atteint pour des densités qui prennent toutes les valeurs réelles au voisinage d'un unique point et sont nulles partout ailleurs (pic de Dirac).

Entropie :

- Elle est maximale pour des densités f dont toutes les valeurs sont dans un petit intervalle, optimum atteint pour un singleton (loi uniforme) : il faut que f varie peu autour de sa moyenne (égale à $\frac{1}{b-a}$).
- Elle diminue ensuite en "étalant" progressivement les pics de Dirac. Elle varie relativement peu sur les densités f dont l'écart moyen à leur moyenne est le même.
- Elle est minimale pour des densités f qui prennent une grande plage de valeurs, optimum atteint pour des densités qui prennent toutes les valeurs réelles (combinaisons de pics de Dirac) : il faut que f varie beaucoup autour de sa moyenne.

2.5.3 Indices basés sur l'entropie continue

Peut-on, d'une façon similaire au cas discret, définir des indices de sensibilité ici ?

Soit un système pouvant théoriquement prendre tous les états entre a et b ($a < b$), suivant la loi de probabilité $\frac{1}{2} \delta_a + \frac{1}{2} \delta_b$. On est alors sûr que ce dernier est dans l'un des deux états a ou b ; par rapport à l'infinité - indénombrable - d'états possibles sur $[a, b]$ on peut considérer que cela constitue une connaissance parfaite de l'état du système. En effet si l'on raisonne localement au voisinage d'un état fixé il n'y a qu'un unique état que le système peut prendre. Bien sûr on pourrait répliquer que les deux valeurs prises étant très différentes, on ne peut pas donner d'intervalle de confiance plus petit que $[a, b]$ pour un système se comportant ainsi, et donc on a le moins d'information possible sur son état global.

Faut-il alors condamner l'utilisation de l'entropie pour le calcul des indices de sensibilité ? - Non, si l'on se restreint à l'entropie discrète (ou discrétisée) qui a le mérite d'être bornée inférieurement et minimale uniquement pour une distribution déterministe.

Indiquons toutefois un bémol à l'utilisation de l'entropie discrète pour le calcul des indices de sensibilité : comme mentionné en introduction celle-ci ne dépend pas des valeurs prises par la variable aléatoire sous-jacente, que ce soit dans le cas discret ou continu, et donc en particulier l'entropie discrète ne distingue pas elle non plus les figures 19 *a.* et *b.*

Si on veut utiliser l'entropie continue il faut l'allier à la variance pour supprimer l'anomalie indiquée. On veut que le nouvel indice soit maximal seulement pour une distribution uniforme, et minimal pour un pic de Dirac uniquement. D'où la définition suivante.

Définition 7 *L'indice*

$$\rho(X : Y) = c^{-H(Y|X)} + e^{\frac{V(\mathbb{E}[Y|X])}{V_u} - 1}$$

mesure l'influence de X sur Y , V_u étant la variance d'une loi uniforme sur l'espace de variation de Y et c un coefficient > 1 .

On divise par V_u dans cette dernière définition pour que le terme dû à la variance soit bien réparti dans $[e^{-1}, e]$. Le point délicat en pratique sera le choix du coefficient c : trop petit, on perd en sensibilité. Trop grand, le terme $c^{-H(Y|X)}$ devient négligeable.

Ce dernier indice n'a aucune raison de rester dans $[0,1]$, mais mesure l'influence d'un paramètre - on l'espère - plus précisément que l'entropie continue ou la variance pris séparément : en effet les figures 14 *a.* et *b.* sont différenciées, ainsi que les figures 19 *a.* et *b.* ; la séparation des densités est donc plus fine que dans les cas de la variance ou de l'entropie pris séparément. De plus il est facile de normaliser cet indice pour qu'il soit dans $[0,1]$, exprimant un pourcentage. Voir le paragraphe 3.1.2 pour un algorithme de calcul, puis le paragraphe 4.3.3 pour quelques applications.

Remarque 5 *Il serait intéressant aussi d'associer les indices de sensibilité basés sur la variance avec ceux basés sur l'entropie discrète pour éliminer les défauts indiqués par les figures précitées. On sent que les deux mesures étudiées doivent être complémentaires.*

3 Calcul pratique des indices de sensibilité η

On trouvera dans cette section quelques algorithmes de calculs des indices, ainsi qu'une étude sur leur convergence.

3.1 Algorithmes

On donne ici une description des algorithmes utilisés ainsi que leur complexité théorique pour le calcul des indices η_i (définition 6), dont la définition discrète unidimensionnelle est la suivante en supposant chaque intervalle de variation découpé en n sous-intervalles :

$$\eta_i = - \left[\sum_{j,k=1}^n F_{j,k}^i \ln \left(\frac{F_{j,k}^i}{\int_{x=x_j^i}^{x_{j+1}^i} f_{X_i}(x) dx \int_{y=y_k}^{y_{k+1}} f_Y(y) dy} \right) \right] H_0^{-1},$$

$$\text{où } H_0 = - \sum_{k=1}^n \int_{y=y_k}^{y_{k+1}} f_Y(y) dy \ln \left(\int_{y=y_k}^{y_{k+1}} f_Y(y) dy \right)$$

$$\text{et } F_{j,k}^i = \int_{x=x_j^i}^{x_{j+1}^i} \int_{y=y_k}^{y_{k+1}} f_{(X_i,Y)}(x,y) dx dy.$$

En ce qui concerne les indices de sensibilité multi-sorties où Y est de dimension d , on cherche à évaluer la somme suivante :

$$\eta_i = - \left[\sum_{j,k=(k_1,\dots,k_d) \in [1,n]^{d+1}} F_{j,k}^i \ln \left(\frac{F_{j,k}^i}{\int_{x=x_j^i}^{x_{j+1}^i} f_{X_i}(x) dx \int_{y=y_k}^{y_{k+1}} f_Y(y) d\mu(y)} \right) \right] H_0^{-1},$$

où $H_0 = - \sum_{k \in [1, n]^d} \int_{y=y_k}^{y_{k+1}} f_Y(y) d\mu(y) \ln \left(\int_{y=y_k}^{y_{k+1}} f_Y(y) d\mu(y) \right)$

et $F_{j,k}^i = \int_{x=x_j^i}^{x_{j+1}^i} \int_{y \in P_k} f_{X_i, Y}(x, y) dx d\mu(y)$,

μ étant la mesure de Lebesgue sur \mathbb{R}^d et P_k un "pavé de discrétisation" de la variable Y .

Lorsque ce sera possible on se placera dans ce dernier cas, plus général, pour la description des algorithmes. Dans le cas des indices de sensibilité basés sur la variance, on dispose d'un algorithme linéaire en le nombre de points dans l'échantillon pour le calculs des indices : voir par exemple Saltelli [8].

3.1.1 Fonction connue

On suppose connue la relation liant les entrées X_1, \dots, X_s avec la sortie $Y : Y = \psi(X_1, \dots, X_s)$. On est alors en mesure de déterminer les fonctions de répartitions empiriques par des intégrales approchées, puis les indices de sensibilité par des sommes discrètes ; il s'agit en quelque sorte d'une discrétisation d'un variable aléatoire dont la densité est connue.

Le calcul des indices η_i s'effectue en trois grandes étapes - décrites ici pour Y de dimension 1 :

1. Détermination du nombre n de sous-intervalles
2. Estimation des fonctions de répartition $F(x)$ et $F(y)$
3. Estimation des indices par une somme sur les sous-intervalles de variation de X_i et Y

On pourrait se dire qu'il est inutile de discrétiser des intervalles de variation, et qu'il serait plus judicieux de calculer directement les intégrales doubles $\int_{x_i=-\infty}^{+\infty} \int_{y \in \mathbb{R}^d} f(x_i, y) \ln \left(\frac{f(x_i, y)}{f(x_i) f(y)} \right) dx_i d\mu(y)$. Le problème est que l'on calcule alors un indice basé sur l'entropie dans le cas continu ; il est positif d'après la positivité de l'information mutuelle (propriété 2), mais rien ne garantit que cet indice soit plus petit que 1 ! Ceci nous force donc à discrétiser les intervalles de variation en un nombre judicieux de points. On donnera dépendant un algorithme pour le calcul des indices ρ (définition 7) au paragraphe 3.1.2.

Le **pseudo-code** est le suivant (supposant n entré par l'utilisateur) :

1. Créer s vecteurs X^i de taille $n + 1$, et un vecteur \hat{Y} pour stocker les points de discrétisation.
 Pour i allant de 1 à s , définir X^i à l'aide des propriétés de leurs lois, puis Y en fonction des bornes obtenues par Monte-Carlo (le calcul direct serait trop coûteux)
2. Définir la fonction $\Phi(u_1, \dots, u_s, a, b)$ qui renvoie 1 si $\psi(u_1, \dots, u_s) \in [a, b]$, 0 sinon.
3. Créer s vecteurs $F X_i$ destinés à contenir les estimations des intégrales des densités des paramètres entre deux bornes définies par les vecteurs X^i , puis un vecteur $F Y$ pour le même usage au niveau de la sortie.
 Pour i allant de 1 à s , calculer chaque intégrale de f_{X_i} sur $[X^i[j], X^i[j + 1]]$ en remplissant $F X_i$.
 Pour i allant de 1 à n faire :

$$F Y[i] \leftarrow \int_{u_1=-\infty}^{+\infty} \dots \int_{u_s=-\infty}^{+\infty} \Phi(u_1, \dots, u_s, \hat{Y}[i], \hat{Y}[i + 1]) f(u_1) \dots f(u_s) du_1 \dots du_s$$
4. Calculer en une boucle $H_0 = - \sum_{k=1}^n F Y[k] \ln(F Y[k])$
 Pour i allant de 1 à s initialiser η_i à 0, puis pour (j, k) parcourant $[1, n]^2$ faire :
 Si $F X_i[j] > 0$ et $F Y[k] > 0$, estimer $F_{j,k}^i$ comme suit :

$$\int_{u_1=-\infty}^{+\infty} \dots \int_{u_{i-1}=-\infty}^{+\infty} \int_{u_i=X^i[j]}^{X^i[j+1]} \int_{u_{i+1}=-\infty}^{+\infty} \dots \int_{u_s=-\infty}^{+\infty} \Phi(u_1, \dots, u_s, \hat{Y}[k], \hat{Y}[k + 1]) f(u_1) \dots f(u_s) du_1 \dots du_s$$
 Si cette dernière quantité est strictement positive, faire :

$$\eta_i \leftarrow \eta_i - F_{j,k}^i \ln \frac{F_{j,k}^i}{F X_i[j] F Y[k]}$$
 On divise alors les η_i par H_0

Complexité :

On ne stocke en mémoire que des quantités linéaires en $s n$, donc la complexité en mémoire est $\Theta(s n)$, s étant le nombre de paramètres d'entrée et n le nombre de pas de discrétisation.

Quant à la complexité en calcul, on remarque que les estimations les plus lourdes se font dans la dernière double boucle avec $\Theta(s n^2)$ intégrales à calculer. Le point délicat pour un logiciel de calcul formel sera certainement ce dernier, les estimations d'intégrales étant difficiles à effectuer.

Quelles seraient les modifications à effectuer pour généraliser cet algorithme à Y de dimension d ? La méthode précédente nous indique qu'il faut

drait alors évaluer $\Theta(n^{d+1})$ intégrales, ce qui compte-tenu de la difficulté à estimer ces dernières précisément deviendrait très vite beaucoup trop lourd. On n'implémente donc pas l'algorithme exact dans ce cas. De plus, les estimations d'intégrales restant des approximations on ne garantit pas la précision du résultat, en particulier si la fonction ψ est compliquée.

3.1.2 Boîte noire

Cette fois on ne dispose pas de la fonction ψ . On en est donc réduit à estimer les probabilités d'appartenance en effectuant des échantillonnages les plus grands possibles (le frein sera bien entendu les capacités de calcul). Tout en laissant moins de possibilités, cette méthode est en fait beaucoup plus simple à mettre en oeuvre et permet de calculer les indices de couples et multi-sorties assez facilement :

- On simule le modèle pour un grand nombre de points : S
- On estime les quantités $F_{j,k}^i$ par le ratio "nombre de points dans le rectangle $[x_j^i, x_{j+1}^i] \times [y_k, y_{k+1}]$ " divisé par S , et de même les intégrales donnant $F(X)$ et $F(Y)$ par le rapport des nombres de points dans l'ensemble de discrétisation par S .
- Une fois ces intégrales estimées, il suffit de faire la somme comme précédemment.

Complexité :

Comme au paragraphe 3.1.1, le gros du calcul se situe dans la boucle effectuant la somme finale si n^{d+1} est plus grand ou de l'ordre de S (nombre de points dans l'échantillon). En revanche ici on n'a pas de lourd calcul d'intégrale à effectuer, donc on peut stocker chaque valeur à additionner (n^{d+1} au total) puis faire efficacement l'addition en parallèle par une instruction `sum(tmp)`. L'affectation étant beaucoup plus rapide que l'addition, on y gagne en temps de calcul, mais on utilise $\Theta(s n^{d+1})$ emplacements mémoire : on peut donc utiliser cette stratégie pour des petites valeurs du couple (n, d) .

La complexité en temps de calcul est donc $\Theta(s S n^{d+1})$, avec une constante multiplicative d'autant plus petite que l'on stocke des valeurs en mémoire.

Code R correspondant à la dernière boucle :

```
tmp<-matrix(0,nrow=n,ncol=(n+1)dimy)
# dans cet ordre, car les affectations sont plus rapides que les additions
for (j in 1 : n)
  for (k in 1 : (n+1)dimy) {
    if (estimationsXY[j,k]>0)
```

```

      tmp[j,k] <- estimationsXY[j,k,i]*log(estimationsXY[j,k,i]
      / (estimationsX[j,i]*estimationsY[k]))
    }
H[i] <- sum(tmp)

```

Remarque 6 Dans un cas comme dans l'autre on peut aussi calculer les indices d'influence mutuelle de couples de variables sur le modèle, définis naturellement à partir de la formule donnant les indices dans le cas d'une variable :

$$\eta_{i,j} = \frac{I((X_i, X_j), Y)}{H(Y)},$$

avec

$$I((X_i, X_j), Y) = \int_{x_i \in \mathcal{X}_i} \int_{x_j \in \mathcal{X}_j} \int_{y \in \mathcal{Y}} f(x_i, x_j, y) \ln \frac{f(x_i, x_j, y)}{f(x_i, x_j) f(y)} d\mu_i(x_i) d\mu_j(x_j) d\mu(y).$$

Il suffit alors de remplacer n^{d+1} par n^{d+2} dans les calculs de complexité.

En ce qui concerne les indices ρ , voici la marche à suivre pour déterminer $H(Y|X_i)$ (en fait le même schéma de programme) :

- On simule le modèle pour un grand nombre de points : S
- On estime les quantités $f(x_j^i, y_k)$ par le ratio "nombre de points dans le rectangle $[x_j^i, x_{j+1}^i] \times [y_k, y_{k+1}]$ " divisé par $pX_i pY S$ où pX_i (resp. pY) est le pas de discrétisation de l'intervalle de variation de X_i (resp. Y), égal respectivement à $\frac{b_X^i - a_X^i}{n}$ et $\frac{b_Y - a_Y}{n}$ (notations évidentes). De même on estime $f(x_j^i)$ et $f(y_k)$ par le rapport des nombres de points dans l'ensemble de discrétisation respectivement par $pX_i S$ et $pY S$.
- Il suffit alors de sommer les quantités $f(x_j^i, y_k) \ln \frac{f(x_j^i, y_k)}{f(x_j^i) f(y_k)}$ pour (j, k) décrivant $[1, n]^2$, puis de multiplier le tout par $pX_i pY$

On se concentre par la suite sur les indices η . Quelques tests sur les indices ρ seront tout de même effectués bien que ceux-ci ne soient pas l'objet de ce travail.

3.1.3 Instabilité des algorithmes

L'avantage d'un calcul effectuant formellement l'intégrale double est que le résultat est unique, ne dépendant de rien d'autre que des distributions de probabilité en entrée et de la fonction ψ . En revanche calculer des indices

discrets (ou discrétisés) impose de choisir un nombre de sous-intervalles de discrétisation : n .

Tout irait pour le mieux s'il y avait une convergence globale vers les "bons indices" lorsque n et S tendent vers $+\infty$. Malheureusement il n'en est rien comme on le justifiera rigoureusement en 3.2.3. En attendant, contentons-nous d'un raisonnement intuitif :

Augmenter le nombre de points S dans l'échantillon revient à améliorer l'approximation des intégrales intervenant dans le calcul des indices η , et entraîne donc une convergence vers un certain réel α_n . Ce dernier est dans $[0,1]$ car à la limite les intégrales sus-citées représentent une distribution de probabilité discrète, les variables aléatoires prenant leurs valeurs dans l'ensemble des points formant les bornes des sous-intervalles. Jusque là tout va bien.

Tout se complique lorsqu'on constate expérimentalement qu'augmenter n fait croître les indices de sensibilité, et donc que les α_n forment a priori une suite croissante. Voici une tentative d'explication :

Comme on l'a vu lors de la justification de la forme de la fonction entropie, plus on a d'alternatives différentes plus ou moins équiprobables, plus l'entropie est grande ; or augmenter n revient à augmenter le nombre d'alternatives (un point peut se trouver à plus d'endroits différents).

Bien sûr les densités de probabilité n'ont aucune raison de représenter des lois uniformes, et on ne peut pas conclure si vite. Mais les densités étant néanmoins continues et admettant dans tous les cas pratiques un extrémum local sur $]a, b[$ comme illustré sur la figure 20 (ce qui n'est pas vrai en général : prendre le cas d'un segment de droite oblique), il arrive un moment où le pas de discrétisation est si petit que l'on obtient quasiment une loi uniforme sur $[\alpha - \varepsilon, \alpha + \varepsilon]$: à partir de cet instant, l'entropie croît quasi-logarithmiquement en le nombre de points dans $[\alpha - \varepsilon, \alpha + \varepsilon]$; on suppose que l'information mutuelle croît plus vite (ce qui se révélera faux plus tard, mais c'est pourtant ce que l'on observe pour $n \leq 2000$). "D'où" la monotonie observée.

3.2 Convergence avec le nombre de sous-intervalles

n étant le nombre de sous-intervalles dans la discrétisation des lois de chacune des variables aléatoires impliquées, on va déterminer la nature de la

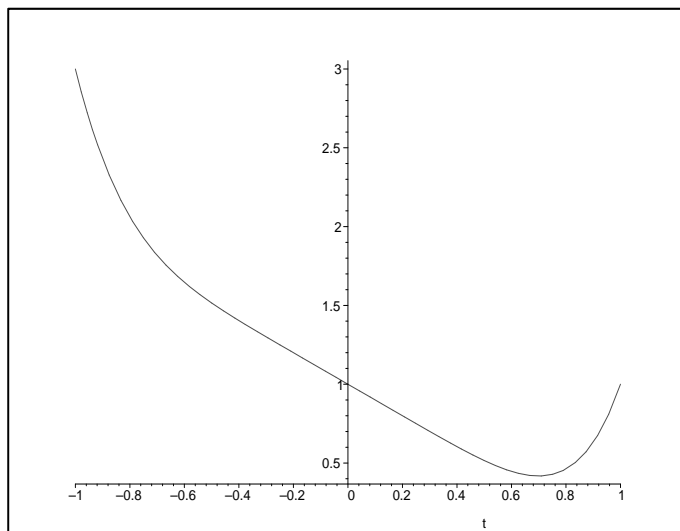


FIG. 20 – Exemple typique avec minimum global

convergence des indices quand n croît.

3.2.1 Restriction de l'étude à l'entropie discrète

Le résultat suivant justifie que l'on ne s'intéresse qu'à l'une des deux entropies (continue ou discrétisée) dans le cas des variables aléatoires continues ; ainsi l'étude théorique discrète - plus naturelle, car tout algorithme ne faisant pas de calcul formel nécessite une discrétisation - suffira à couvrir les cas des indices η et ρ .

Lemme :

Dans le cas d'une variable aléatoire continue X de support $[a, b]$,

$$H(X) = - \int_a^b f(x) \ln(f(x)) dx \text{ et on pose } H_n(X) = - \sum_{k=1}^n \frac{\hat{f}(k)}{Z_n^f} \ln \frac{\hat{f}(k)}{Z_n^f}$$

où $\hat{f}(k) = f\left(a + (2k - 1) \frac{b-a}{2n}\right)$ et $Z_n^f = \sum_{k=1}^n \hat{f}(k)$, quasi-conformément à la définition de l'entropie discrète donnée en introduction de 2.2 (on a juste approximé les intégrales par des sommes de Riemann).

On a alors, $[a, b]$ et $[c, d]$ étant les supports respectifs de X et Y :

$$i) \quad H(X) \leq H(Y) + \ln \frac{b-a}{d-c} \quad \Rightarrow \quad \forall \varepsilon > 0 \exists N_\varepsilon \in \mathbb{N} / \forall n \geq N, \\ H_n(X) \leq H_n(Y) + \varepsilon$$

$$ii) \quad \exists N \in \mathbb{N} / \forall n \geq N, H_n(X) \leq H_n(Y) \quad \Rightarrow \quad H(X) \leq H(Y) + \ln \frac{b-a}{d-c}$$

Preuve :

On montre tout d'abord le second point.

Supposons

$$\exists N \in \mathbb{N} / \forall n \geq N, H_n(X) \leq H_n(Y).$$

Écrivons cette dernière inégalité avec la définition :

$$-\sum_{k=1}^n \frac{\hat{f}(k)}{Z_n^f} \ln \frac{\hat{f}(k)}{Z_n^f} \leq -\sum_{k=1}^n \frac{\hat{g}(k)}{Z_n^g} \ln \frac{\hat{g}(k)}{Z_n^g},$$

f et g étant les densités respectives de X et Y .

Développons :

$$\ln(Z_n^f) - \frac{1}{Z_n^f} \sum_{k=1}^n \hat{f}(k) \ln(\hat{f}(k)) \leq \ln(Z_n^g) - \frac{1}{Z_n^g} \sum_{k=1}^n \hat{g}(k) \ln(\hat{g}(k)).$$

Faisons ensuite apparaître des sommes de Riemann :

$$\ln \frac{Z_n^f}{n} - \frac{n}{Z_n^f} \frac{1}{n} \sum_{k=1}^n \hat{f}(k) \ln(\hat{f}(k)) \leq \ln \frac{Z_n^g}{n} - \frac{n}{Z_n^g} \frac{1}{n} \sum_{k=1}^n \hat{g}(k) \ln(\hat{g}(k)).$$

On sait d'après la théorie de l'intégration des fonctions continues par morceaux que $\lim_{n \rightarrow +\infty} \frac{b-a}{n} \sum_{k=1}^n u(a + (2k-1) \frac{b-a}{2n}) = \int_a^b u(x) dx$, u étant définie sur $[a, b]$.

D'où, par prolongement des inégalités à la limite :

$$-\ln(b-a) - \int_a^b f(x) \ln(f(x)) dx \leq -\ln(d-c) - \int_c^d g(x) \ln(g(x)) dx.$$

L'inégalité souhaitée en découle.

On obtient le premier point en remontant la démonstration du second :
 Supposons $H(X) \leq H(Y) + \ln \frac{b-a}{d-c}$.
 C'est équivalent à

$$-\ln(b-a) + H(X) \leq -\ln(d-c) + H(Y).$$

Ainsi d'après les calculs précédents

$$\ln \frac{Z_n^f}{n} - \frac{n}{Z_n^f} \frac{1}{n} \sum_{k=1}^n \hat{f}(k) \ln(\hat{f}(k)) \leq \ln \frac{Z_n^g}{n} - \frac{n}{Z_n^g} \frac{1}{n} \sum_{k=1}^n \hat{g}(k) \ln(\hat{g}(k)) + \varepsilon_n,$$

avec $(\varepsilon_n)^+ \rightarrow 0$.

Soit

$$H_n(X) \leq H_n(Y) + \varepsilon_n.$$

ε étant fixé, on sait qu'il existe un rang N_ε à partir duquel $\varepsilon_n \leq \varepsilon$. D'où le résultat.

□

3.2.2 Évolution de l'entropie discrète

Comment évolue l'entropie discrète quand le nombre n de sous-intervalles de discrétisation augmente ?

À travers la preuve précédente émerge l'idée de normaliser l'entropie discrète par $\ln(n)$, et on obtient alors la formule

$$\begin{aligned} \frac{H_n(X)}{\ln(n)} &= \frac{1}{\ln(n)} \ln \frac{Z_n^f}{n} + 1 - \frac{1}{\ln(n)} \frac{n}{Z_n^f} \frac{1}{n} \sum_{k=1}^n \hat{f}(k) \ln(\hat{f}(k)) \\ &\rightarrow 1. \end{aligned}$$

Il apparaît donc que quelle que soit X , l'entropie discrète $H_n(X)$ est équivalente à $\ln(n)$ quand n tend vers $+\infty$.

Ainsi si l'on se donne deux variables aléatoires X et Y , pour tout ε strictement positif il existe un rang N_ε à partir duquel $|H_n(X) - \ln(n)| \leq \frac{\varepsilon}{2}$ et $|H_n(Y) - \ln(n)| \leq \frac{\varepsilon}{2}$, et donc par inégalité triangulaire $|H_n(X) - H_n(Y)| \leq \varepsilon$.

Cette entropie discrète H_n coïncide à la limite avec l'expression donnée en introduction de la section 2.2 ; on a en fait approximé les intégrales par la méthode des rectangles.

D'où le

Théorème 5 *Pour toutes variables aléatoires réelles à densité Riemann-intégrable X et Y à support dans $[a, b]$,*

$$\lim_{n \rightarrow +\infty} H_n(X) - H_n(Y) = 0.$$

En guise de conclusion provisoire on peut dire qu'à la limite lorsque le pas de subdivision tend vers 0, toutes les distributions ont la même entropie discrète ; il faut donc choisir n une bonne fois pour toutes : le choix est arbitraire, mais nécessaire.

Il faut maintenant se demander dans quelle mesure ces résultats affectent le calcul des indices de sensibilité. Celui-ci utilise l'entropie comme facteur normalisateur : à terme quand $n \rightarrow +\infty$ tous les indices sont divisés par la même quantité, mais qu'en est-il de la quantité d'information ?

3.2.3 Évolution de la quantité d'information discrète

Ce paragraphe poursuit l'étude précédente dans le cas de l'information mutuelle.

On transforme l'écriture de la quantité d'information avec la même idée que précédemment pour l'entropie (approximation d'intégrales) :

$$I_n(X, Y) = \sum_{k,l=1}^n \frac{\hat{h}(k, l)}{Z_{n^2}^h} \ln \frac{\hat{h}(k, l) Z_n^f Z_n^g}{\hat{f}(k) \hat{g}(l) Z_{n^2}^h}$$

avec $Z_n^f = \sum_{k=1}^n \hat{f}(k)$, $Z_n^g = \sum_{k=1}^n \hat{g}(k)$, $Z_{n^2}^h = \sum_{k,l=1}^n \hat{h}(k, l)$, f , g et h étant les densités respectives de X , Y , (X, Y) , et \hat{f} , \hat{g} , \hat{h} se définissant comme au lemme du paragraphe 3.2.2.

On obtient immédiatement :

$$\begin{aligned} I_n(X, Y) &= \frac{n^2}{Z_{n^2}^h} \frac{1}{n^2} \sum_{k,l=1}^n \hat{h}(k, l) \{ \ln(\hat{h}(k, l)) \\ &\quad + \ln(Z_n^f) + \ln(Z_n^g) - \ln(Z_{n^2}^h) - \ln(\hat{f}(k)) - \ln(\hat{g}(l)) \}. \end{aligned}$$

Puis après quelques calculs

$$I_n(X, Y) = \frac{n^2}{Z_{n^2}^h} \left(\frac{1}{n^2} \sum_{k,l=1}^n \hat{h}(k, l) \ln(\hat{h}(k, l)) \right. \\ \left. + \frac{Z_{n^2}^h}{n^2} \left(\ln \frac{Z_n^f}{n} + \ln \frac{Z_n^g}{n} - \ln \frac{Z_{n^2}^h}{n^2} \right. \right. \\ \left. \left. - \frac{n}{Z_n^f} \frac{1}{n} \sum_{k=1}^n \hat{f}(k) \ln(\hat{f}(k)) - \frac{n}{Z_n^g} \frac{1}{n} \sum_{k=1}^n \hat{g}(k) \ln(\hat{g}(k)) \right) \right),$$

avec $\frac{\sum_{j=1}^n \hat{h}(i, j)}{Z_{n^2}^{f, g}} = \frac{\hat{f}(i)}{Z_n^f}$, analogue discretisé de la formule de la loi conjointe :

$\int_{y=c}^d h(i, y) dy = f(i)$. L'égalité n'est valable qu'à un ε près, mais ce dernier tend vers 0 quand n augmente. Il est alors clair que $I_n(X, Y)$ converge d'après les propriétés des sommes de Riemann.

On peut expliciter la limite :

$$\lim_{n \rightarrow +\infty} I_n(X, Y) = H(X, Y) + \frac{1}{(b-a)(d-c)} (H(X) + H(Y)).$$

Cette dernière expression ressemble à l'information mutuelle "continue", mais n'a pourtant rien à voir avec comme le montrent le signe changé et l'influence des supports des lois.

On en déduit le

Théorème 6

$$\forall i, \quad \lim_{n \rightarrow +\infty} \eta_i^n(X, Y) = 0,$$

η_i^n étant le $i^{\text{ème}}$ indice de sensibilité théorique obtenu avec $S \rightarrow +\infty$, pour n sous-intervalles de discrétisation.

Plus exactement, on a :

$$\forall i, \quad \lim_{n \rightarrow +\infty} \ln(n) \eta_i^n(X, Y) = H(X, Y) + \frac{1}{(b-a)(d-c)} (H(X) + H(Y)),$$

H désignant l'entropie continue.

La démonstration est immédiate en regroupant les résultats qui précèdent.

□

Un phénomène assez désagréable rend ainsi le choix de n encore plus crucial : lorsque le pas de subdivision des intervalles tend vers zéro, les indices de sensibilité tendent tous vers zéro. Cette convergence est toutefois très lente, due à la croissance logarithmique de l'entropie. Le choix de n définit donc à la fois la précision des indices et leur échelle; cette difficulté peut être contournée en ne divisant pas par $H(Y)$ puis en normalisant les résultats en pourcentage d'influence, obtenant ainsi des indices dans $[0,1]$ qui convergent non trivialement avec n et S .

D'où la définition modifiée :

Définition 8 *Indice de sensibilité relatif, exprimé en pourcentages :*

$$\eta_{i,n}^R = \left[\sum_{j,k=1}^n F_{j,k}^i \ln \left(\frac{F_{j,k}^i}{\int_{x=x_j^i}^{x_{j+1}^i} f_{X_i}(x) dx \int_{y=y_k}^{y_{k+1}} f_Y(y) dy} \right) \right] T_n^{-1},$$

où

$$T_n = \sum_{i=1}^s \sum_{j,k=1}^n F_{j,k}^i \ln \left(\frac{F_{j,k}^i}{\int_{x=x_j^i}^{x_{j+1}^i} f_{X_i}(x) dx \int_{y=y_k}^{y_{k+1}} f_Y(y) dy} \right).$$

A-t-on alors réussi à éliminer la convergence vers 0 tout en conservant les propriétés des indices? Hélas non, car si ces nouveaux indices ont en plus le mérite d'exprimer directement les contributions de chaque variable en pourcentages, ils ont le gros défaut de ne pas faire de différence entre des "vrais" indices qui vaudraient (0.1,0.1,0.1) pour un modèle, et (0.5,0.5,0.5) pour un autre. On a perdu l'absolu pour le relatif : ces indices ne peuvent être utilisés en pratique dès que l'on a besoin de connaître l'intensité des influences. De plus les algorithmes sont non linéaires en n , donc savoir qu'il y a convergence aussi pour $n \rightarrow +\infty$ ne donne pas d'applications pratiques.

Maintenant que l'on sait qu'il faut fixer n , étudions le comportement en fonction du nombre du nombre de points dans l'échantillon.

3.3 Convergence avec la taille de l'échantillon

Expérimentalement on constate une convergence lente lorsque le nombre S de points dans l'échantillon augmente. Essayons de préciser cela.

3.3.1 Échantillonnage déterministe pour les indices η

Jusqu'ici les échantillons dans les algorithmes ont été obtenus par un générateur aléatoire. Mais a-t-on toujours besoin de hasard dans le cas des indices basés sur l'entropie ? N'ayant aucune information sur la fonction ψ (ce qui en pratique sera presque toujours le cas) on souhaite juste avoir le plus de points possible, les mieux répartis possible. Dans le cas d'une loi uniforme, il semble plus intéressant de répartir les points régulièrement dans l'intervalle de définition de la variable aléatoire. On se place alors pour la suite dans le cadre d'un paramètre d'entrée suivant une loi uniforme : c'est le cas le plus simple.

Donc si X suit la loi uniforme $\mathcal{U}[a, b]$, il suffit d'échantillonner en

$$X_i = \left\{ a + \frac{(2k+1)(b-a)}{2S} \right\}_{k \in [0, \dots, S-1]} \quad ? - \text{Pas si sûr, car alors tous les p-uplets } (X_{i_1}, \dots, X_{i_p}) \text{ sont multiples du vecteur } (1, \dots, 1) \text{ et ne sont pas du tout représentatifs de l'espace p-dimensionnel qu'ils sont censés parcourir.}$$

Nous allons donc représenter le plus régulièrement possible l'espace sus-cité en espaçant les points non pas de $\frac{b-a}{S}$, mais $\frac{b-a}{\sqrt{S}}$ où s est le nombre de paramètres en entrée, effectuant ainsi de l'ordre de $S^{\frac{s-1}{s}}$ répétitions du même ensemble de points, à chaque fois dans un ordre différent pour avoir le plus de s-uplets possibles ; cela revient simplement à prendre des points sur les noeuds d'une grille régulière de pas $\frac{b-a}{\sqrt{S}}$ à s dimensions.

Ce nouvel échantillonnage est alors mieux représentatif de l'espace des variations qu'un échantillonnage aléatoire (dans le cas d'une loi uniforme). C'est celui-ci que l'on va utiliser pour l'analyse suivante.

3.3.2 Vitesse de convergence de l'entropie

On souhaite préciser la vitesse de convergence de l'entropie en fonction du nombre de points S .

Un rapide coup d'oeil à la définition discrète des indices basés sur l'entropie nous indique qu'il faut - et qu'il suffit - de savoir résoudre les équations du type $\psi(x) \in J$ où x est dans l'espace approprié de dimension s et J un intervalle (de largeur $\frac{|J|}{n}$). C'est impossible en général, d'où la difficulté d'obtenir les indices analytiquement. On se concentre ici sur le calcul d'une seule intégrale type $\mathcal{I}_k = \int_{y_k}^{y_{k+1}} f(y) dy$; il faudra ensuite sommer n telles intégrales.

Si l'on suppose ψ continue (ce qui risque de se révéler faux parfois, mais ψ sera tout de même suffisamment "lisse" la plupart du temps), on sait que l'on peut approcher l'ensemble solution \mathcal{S} aussi près que l'on veut par une réunion finie de pavés produits d'intervalles compacts.

Plus exactement :

$$\forall \varepsilon > 0, \exists P_1, \dots, P_{N_\varepsilon} \quad / \quad \mu(\mathcal{S} - \cup_{i \in [1, \dots, N_\varepsilon]} P_i) \leq \varepsilon,$$

où $P_i = I_1^i \times \dots \times I_s^i \subset \mathcal{S}$, chaque I_k^i étant dans le domaine de définition de X_k , et μ est la mesure de Lebesgue sur \mathbb{R}^s . Cette approximation n'est pas totalement nécessaire, mais aide à la compréhension intuitive du phénomène.

En utilisant l'échantillonnage du paragraphe 3.3.1, il est facile de voir que l'on aura alors environ $\prod_{k=1}^s \left[\frac{\sqrt[s]{S}}{b_k - a_k} |I_k^i| \right]$ s-uplets solutions dans l'ensemble des points échantillonnés (a_k et b_k étant les bornes des intervalles de variations des X_k), que l'on divisera ensuite par S pour avoir une valeur approchée de la probabilité exacte. Or le rapport que l'on cherche à approcher est égal à la mesure de l'ensemble des solutions : c'est donc bien cette valeur qui est obtenue à la limite pour $S \rightarrow +\infty$.

En guise de conclusion provisoire, multiplier S par 10 devrait faire gagner environ 1 digit sur la précision du résultat (modulo ε , mais ce dernier est aussi petit que l'on veut); ce n'est toutefois qu'une approximation de plus, car le résultat exact est que l'on gagne s décimales en élevant S à la puissance s . On en déduit que chaque intégrale est avec une bonne probabilité dans l'intervalle $\left[v_k - \frac{\alpha}{S}, v_k + \frac{\alpha}{S} \right]$ à mesure que S augmente, v_k étant la valeur exacte de l'intégrale et α une constante égale à la précision sur le résultat pour $S = 1$.

De ce qui précède découle un encadrement de notre approximation de l'entropie (en supposant $\frac{\alpha}{S} < 1$) :

$$\sum_{k=1}^n \left(v_k - \frac{\alpha}{S} \right) \ln \left(v_k - \frac{\alpha}{S} \right) \leq \tilde{H}_n \leq \sum_{k=1}^n \left(v_k + \frac{\alpha}{S} \right) \ln \left(v_k + \frac{\alpha}{S} \right).$$

On utilise ensuite le développement limité du logarithme sous la forme $\ln(1 - \varepsilon) = -\varepsilon + o(\varepsilon)$ pour obtenir après quelques lignes :

$$H_n - \frac{\alpha}{S} \left| n + \sum_{k=1}^n \ln(v_k) \right| + o\left(\frac{1}{S}\right) \leq \tilde{H}_n \leq H_n + \frac{\alpha}{S} \left| n + \sum_{k=1}^n \ln(v_k) \right| + o\left(\frac{1}{S}\right),$$

H_n étant la valeur exacte de l'entropie discrète.

n étant fixé (à cause de la divergence de l'entropie en $+\infty$), on en déduit que la convergence de l'entropie (avec S cette fois) s'effectue avec une précision de l'ordre de $\frac{1}{S}$; en pratique on n'observe la convergence que lorsque S dépasse $\left|n + \sum_{k=1}^n \ln(v_k)\right|$, ce qui arrive toutefois assez vite si n est raisonnablement choisi.

3.3.3 Vitesse de convergence de l'information mutuelle

On cherche à présent à savoir comment converge l'information mutuelle avec S .

On refait la même analyse dans le cas du numérateur de la fraction donnant les indices de sensibilité jusqu'à arriver à l'encadrement suivant :

$$\sum_{k,l=1}^n \left(v_{k,l} - \frac{\alpha}{S}\right) \ln \frac{v_{k,l} - \frac{\alpha}{S}}{\left(v_k^X + \frac{\alpha}{S}\right)\left(v_l^Y + \frac{\alpha}{S}\right)} \leq \tilde{I}_n \leq \sum_{k,l=1}^n \left(v_{k,l} + \frac{\alpha}{S}\right) \ln \frac{v_{k,l} + \frac{\alpha}{S}}{\left(v_k^X - \frac{\alpha}{S}\right)\left(v_l^Y - \frac{\alpha}{S}\right)}$$

avec $v_k^X = \int_{x_k}^{x_{k+1}} f(x) dx$, $v_l^Y = \int_{y_l}^{y_{l+1}} f(y) dy$, et $v_{k,l} = \int_{x=x_k}^{x_{k+1}} \int_{y=y_l}^{y_{l+1}} f(x, y) dx dy$.

On transforme cette dernière inégalité jusqu'à obtenir, à l'aide du développement limité du logarithme et de $\frac{1}{1-x}$:

$$I_n - \frac{\alpha}{S} \sum_{k,l=1}^n \left(1 + v_{k,l} \frac{v_k^X + v_l^Y}{v_k^X v_l^Y} + \ln \frac{v_{k,l}}{v_k^X v_l^Y}\right) + o\left(\frac{1}{S}\right) \leq \tilde{I}_n,$$

$$\tilde{I}_n \leq I_n + \frac{\alpha}{S} \sum_{k,l=1}^n \left(1 + v_{k,l} \frac{v_k^X + v_l^Y}{v_k^X v_l^Y} + \ln \frac{v_{k,l}}{v_k^X v_l^Y}\right) + o\left(\frac{1}{S}\right),$$

$v_{k,l}$ et $v_k^X v_l^Y$ étant du même ordre de grandeur, on en déduit que la somme intervenant dans cette dernière inégalité est de l'ordre de n^2 . Ceci a des conséquences dramatiques sur la convergence des algorithmes : il faudra que S dépasse βn^2 où β est souvent de l'ordre de 1 pour observer une convergence nette des indices !

3.3.4 Généralisation à des lois non uniformes

Dans le cas général où X , de densité f , ne suit pas une loi uniforme, on peut reprendre la même idée : plus la densité est élevée plus la concentration de points échantillonnés doit être grande. Voici alors la méthode (toujours

déterministe) à appliquer :

1. Diviser l'intervalle en n morceaux, et calculer (numériquement si nécessaire) les intégrales de f sur chaque morceau : on note ces intégrales α_k , et on notera N_k le nombre de points dans chaque sous-intervalle.
2. Initialisation : Classer les α_n par ordre décroissant, puis mettre un point dans chaque case en respectant cet ordre. S'il n'y a pas assez de points pour remplir toutes les cases, on s'arrête. Sinon, $i \leftarrow n$
3. Boucle : tant que $i < S$ faire
Pour k allant de 1 à $n - 1$ faire
Si $\frac{\alpha_k}{\alpha_{k+1}} > \frac{N_k}{N_{k+1}}$ alors : $\begin{cases} \alpha_k \leftarrow \alpha_k + 1 \\ i \leftarrow i + 1. \end{cases}$
4. On souhaite au final comme ci-dessus de l'ordre de $\sqrt[S]{S}$ points se répétant de manière appropriée dans le vecteur X : on normalise donc chaque N_k pour satisfaire cette contrainte, puis on place les points obtenus dans X

Il est facile de constater qu'avec cet algorithme on tend vers la répartition de points idéale, à n fixé. De plus l'étude qui précède nous donne une borne supérieure sur le nombre de points nécessaires pour gagner une décimale : il faut en moyenne moins de 10 fois plus de points pour gagner un digit sur le résultat, la convergence la plus lente s'obtenant dans le cas de distributions uniformes en entrée.

En pratique on n'implémente pas cet algorithme tel quel pour les lois à fonction de répartition non explicites car cela nécessiterait de calculer énormément d'intégrales numériquement. On tabule d'abord les valeurs qui nous intéressent à n fixé avant de lancer l'algorithme qui n'a pas à faire ces calculs. Compte-tenu du temps perdu à tabuler les valeurs, l'ancien échantillonnage est préférable pour n grand et/ou en présence de paramètres d'entrée donc les intégrales sont difficiles à calculer. De plus une fois que l'on connaît le nombre de points à placer dans chaque sous-intervalle on a aucun renseignement sur la façon dont ceux-ci doivent être agencés à l'intérieur de cet intervalle ; pour n grand cela n'a pas tellement d'importance, mais dans les applications pratiques ce ne sera pas le cas, simplement à cause du temps de calcul en $O(n^{d+\alpha})$.

Attention, cette méthode n'est pas applicable dans le cas de la variance car il faut alors deux échantillons non corrélés pour l'algorithme.

3.3.5 Bilan sur la convergence

On peut à présent donner la vitesse de convergence théorique des algorithmes utilisant l'entropie. Dans le cas du calcul des indices d'ordre 1 avec une sortie unidimensionnelle, la convergence est extrêmement mauvaise tant que S (nombre de points dans l'échantillon) n'a pas dépassé la valeur critique $\sum_{k,l=1}^n \left(1 + v_{k,l} \frac{v_k^X + v_l^Y}{v_k^X v_l^Y} + \ln \frac{v_{k,l}}{v_k^X v_l^Y} \right)$, qui est de l'ordre de n^2 . Pour les indices d'ordre 2 et/ou lorsqu'on a une sortie multidimensionnelle, on généralise sans mal ce résultat : S doit être au moins du même ordre de grandeur que $n^{d+\alpha} - \alpha$ étant l'ordre des indices - pour commencer à observer une convergence nette.

Ainsi dans les tests avec $n = 100$ il aurait fallu commencer à $S = 10000$ pour espérer voir quelque chose de similaire à la convergence des indices basés sur la variance : il faut donc en général (n pas trop petit) beaucoup plus de points que dans le cas des indices de sensibilité basés sur la variance.

On en déduit la méthodologie suivante :

- Fixer n le plus grand possible tel que des simulations à $n^{d+\alpha}$ boucles soient possibles en temps raisonnable, $\alpha \in \{1, 2\}$ suivant l'ordre des indices
- Commencer à faire tourner l'algorithme à $S = \beta n^{d+\alpha}$, $\beta \in [0, 1]$ étant d'autant plus proche de 1 que les lois des variables aléatoires sont de grande entropie. On peut choisir par exemple $\beta = \frac{1}{n \ln(n)} \sum_{k=1}^n F_k^n$, où F_k^n est l'entropie discrétisée du k^{eme} paramètre d'entrée.
- Augmenter ensuite S jusqu'à avoir un temps de calcul trop long ou une stabilisation nette des indices.

4 Applications à des fonctions tests

On présente ici plusieurs cas d'application des algorithmes décrits dans la section précédente.

4.1 Fonction d'Ishigami

Cette fonction constitue un test classique d'efficacité des algorithmes calculant les indices basés sur la variance : elle n'est ni trop compliquée, ni trop simple, représentant un modèle possible physiquement.

Voici sa définition :

$$is_{a,b} : \mathbb{R}^3 \longrightarrow \mathbb{R}, (x, y, z) \longrightarrow \sin(x) + a \sin^2(y) + b z^4 \sin(x),$$

x, y, z étant les résultats de variables aléatoires toutes trois de loi uniforme sur l'intervalle $[-\pi, \pi]$.

L'allure de la loi de $Y = is(X, Y, Z)$ est donnée sur la figure 21.

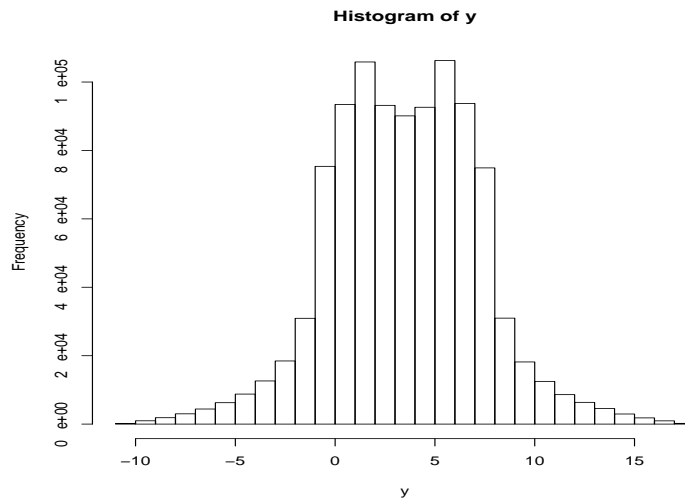


FIG. 21 – Histogramme de la distribution de sortie

4.1.1 Indices exacts

Pour les indices basés sur l'entropie les calculs théoriques sont délicats à cause de la discrétisation effectuée et de la présence du logarithme. On va donc tenter un calcul d'intégrales à l'aide de Maple et de l'algorithme décrit en 3.1.1.

Pseudo-code Maple adapté à la fonction d'Ishigami :

```

hs :=(x,y,z)->sin(x)+a*sin(y)^2+b*z^4*sin(x) :
Phi :=(x,y,z,u,v)->(1/(8*PI^3))*piecewise(is(x,y,z)<u,0,
      is(x,y,z)<=v,1,0) :
FY :=vector(n) :
FX :=2 PI/n :
for j in 1..n do FY[j] :=intégrale de Phi(x,y,z,y_j,y_{j+1})dx dy dz
      sur [-PI,PI]^3 end do :
H0 :=0 :
for j in 1 :100 do H0 :=H0-FY[j]*ln(FY[j]) end do :
for i in 1..3 do :
  η[i] :=0 :
  for j in 1..100 do :
    for k in 1..100 do :
      if FY[k]>0
        aux :=intégrale de Phi(x,y,z,y_k,y_{k+1})dx dy dz sur P_{i,j} :
        if aux>0 η[i] :=η[i]+aux*ln(aux/(FX*FY[k])) end if :
      end if :
    end do :
  end do :
  η[i] :=η[i]/H0 :
end do :

```

où les points y_k sont espacés de $\frac{2\pi}{n}$ régulièrement répartis dans $[-\pi, \pi]$, et $P_{i,j} = [-\pi, \pi]^{i-1} \times [x_j, x_{j+1}] \times [-\pi, \pi]^{3-i}$.

Alternative aux deux dernières lignes pour calculer les indices modifiés du paragraphe 3.2.3, définition 8 :

```

end do :
sum :=0 :
for i in 1..3 do sum :=sum+η[i] : end do :
for i in 1..3 do η[i] :=η[i]/sum : end do :

```

En pratique hélas on se rend compte que Maple ne sait pas intégrer les fonctions définies par morceaux ; or quel que soit la façon de mener le calcul de telles fonctions interviendront forcément. On en est donc réduit au calcul approché par Monte-Carlo.

S	η_1	η_2	η_3
100 (<1s)	0.5289	0.5624	0.5115
500 (<1s)	0.4565	0.4457	0.3693
1000 (<1s)	0.5180	0.3996	0.4100
5000 (<1s)	0.1324	0.1537	$7.924e-2$
10000 (<1s)	0.1220	0.1591	$8.609e-2$
50000 (<1s)	$9.815e-2$	0.1332	$6.409e-2$
100000 (<1s)	$9.254e-2$	0.1238	$6.093e-2$
500000 (1s)	$8.967e-2$	0.1198	$5.821e-2$
1000000 (3s)	$9.189e-2$ (33.9%)	0.1210 (44.7%)	$5.778e-2$ (21.4%)

TAB. 2 – Indices de sensibilité basés sur l'entropie

4.1.2 Indices approchés

Faisons maintenant tourner l'algorithme de Monte-Carlo pour le calcul des indices de la définition 6. On fixe (arbitrairement) le nombre de sous-intervalles de discrétisation $n = 100$, et les paramètres $a = 7$, $b = 0.1$.

Les résultats sont dans le tableau 2.

Dans le cas de l'entropie, tous les calculs ont été effectués avec l'algorithme écrit en C++ utilisant l'échantillonnage déterministe donnée au paragraphe 3.3.1 pour des variables aléatoires uniformes : il est inutile alors de répéter plusieurs fois l'algorithme, ce qui permet des gains de temps considérables. Avec l'algorithme écrit en R utilisant un échantillonnage par hypercube latin on arrive à des temps de calcul beaucoup plus longs, et il faut en plus répéter au moins 10 fois l'algorithme.

Dans le cas de la variance on utilise le générateur "aléatoire" `rand()` du langage C, on pourrait donc améliorer les résultats en utilisant un meilleur échantillonnage ; de plus le gain de temps réalisé est alors minime, car le programme R correspondant utilise bien le calcul matriciel du langage R. Il faudra donc probablement conserver le programme R dans le cas de la variance, et utiliser l'autre programme lorsque l'on souhaite des indices basés sur l'entropie

Il est amusant de constater que les temps de calcul sont "super-linéaires" en S (ie. le temps de calcul pour $2S$ points peut être plus petit que celui pour S points) : cela s'explique par la présence des boucles "parasites" sur n

S	S_1	S_2	S_3	St_1	St_2	St_3
100 (<1s)	0.2622 0.136	0.4482 0.131	$-4.418e-2$ 0.180	0.5779 0.165	0.4663 0.110	0.2715 0.113
500 (<1s)	0.3168 $5.90e-2$	0.4463 $6.87e-2$	$-5.696e-4$ $8.94e-2$	0.5555 $8.18e-2$	0.4451 $4.92e-2$	0.2380 $4.91e-2$
1000 (1s)	0.3102 $4.65e-2$	0.4433 $3.76e-2$	$7.931e-3$ $5.66e-2$	0.5459 $4.48e-2$	0.4462 $4.27e-2$	0.2436 $3.49e-2$
5000 (2s)	0.3192 $2.14e-2$	0.4485 $1.82e-2$	$1.683e-2$ $2.37e-2$	0.5470 $1.95e-2$	0.4361 $1.67e-2$	0.2447 $1.62e-2$
10000 (4s)	0.3115 $1.36e-2$	0.4394 $1.41e-2$	$-7.434e-3$ $2.16e-2$	0.5615 $1.80e-2$	0.4460 $1.20e-2$	0.2426 $1.09e-2$
50000 (20s)	0.3139 $6.07e-3$	0.4411 $5.73e-3$	$-3.307e-3$ $7.45e-3$	0.5604 $6.95e-3$	0.4429 $4.89e-3$	0.2432 $4.62e-3$
100000 (40s)	0.3133 $4.37e-3$	0.4443 $3.76e-3$	$2.406e-3$ $5.49e-3$	0.5549 $4.58e-3$	0.4427 $3.80e-3$	0.2439 $2.78e-3$
500000 (3min10s)	0.3145 $2.06e-3$	0.4421 $1.72e-3$	$3.883e-4$ $2.52e-3$	0.5578 $2.22e-3$	0.4418 $1.56e-3$	0.2437 $1.57e-3$
$S \rightarrow +\infty$ (théorie)	0.3138	0.4424	0	0.5574 (44.9%)	0.4442 (35.7%)	0.2410 (19.4%)

TAB. 3 – Indices de sensibilité basés sur la variance

calculant les indices. Rappelons que la complexité est en $O(s S n^{d+1})$: pour $S \leq 10000$ le gros du calcul se fait dans les boucles sur n ; ensuite à partir de 50000 ces dernières boucles s’effectuent en temps négligeable devant les opérations exécutées S fois, d’où la quasi-linéarité observée alors.

Le tableau 3 présente les résultats obtenus dans le cas des indices de sensibilité basés sur la variance. Les temps de calculs pour les 100 répétitions sont entre parenthèses et les écarts-type sont en dessous des valeurs des indices.

On remarque que dans le cas des indices basés sur l’entropie la convergence est très lente, comme on l’a justifié à la section 3.3. On obtient en revanche très vite deux à trois chiffres significatifs pour les indices basés sur la variance. Les temps de calcul étant plus long, les avantages et inconvénients des algorithmes se compensent en pratique.

Les indices du premier ordre basés sur la variance n’indiquent que la contribution de la variable seule, sans ses interactions. En revanche, les indices basés sur l’entropie expriment toute l’information apportée par la va-

riable. Les indices η_i sont donc globaux : leur homologues sont les indices totaux St_i , auxquels on les compare dans le paragraphe suivant.

Les importances des deux premiers paramètres s'inversent d'un type d'analyse à l'autre : l'un est plus influent que l'autre suivant la mesure choisie, mais les proportions restent les mêmes (34% et 45%). Ceci est peut-être dû à la relativement faible valeur de n (100), mais pour observer de grands changements il faudrait choisir n tel que $\ln(n) \gg \ln(100)$, ce qui est impossible en pratique compte-tenu de la complexité non linéaire en n . Le troisième paramètre a quant à lui la même valeur d'indice dans les deux cas : $\eta_3 \simeq St_3$.

4.1.3 Indices multi-sorties

On peut calculer les indices théoriques et expérimentaux dans le cas multivarié appliqué à la fonction suivante de la même manière :

$$\hat{is}_{a,b} : (x, y, z) \longrightarrow (is_{a,b}(x, y, z), is_{a,b}(y, x, z)).$$

Voici le **pseudo-code Maple** correspondant :

```

hs2 :=(x,y,z)->sin(x)+a*sin(y)^2+b*z^4*sin(x) :
Phi :=(x,y,z,u,v,u',v')->(1/(8*PI^3))*piecewise(is2(x,y,z)<u,0,
    is2(x,y,z)>v,0,is2(y,x,z)<u',0,is2(y,x,z)>v',0,1) :
FY :=matrix(n,n) :
FX :=2 PI/n :
for j in 1..n do
    for k in 1..n do :
        FY[j,k] :=intégrale de Phi(x,y,z,y_j,y_{j+1},y'_k,y'_{k+1}) dx dy dz
            sur [-PI,PI]^3 :
    end do :
end do :
H0 :=0 :
for j in 1 :100 do
    for k in 1..n do H0 :=H0-FY[j,k]*ln(FY[j,k]) end do :
end do :
for i in 1..3 do :
    η[i] :=0 :
    for j in 1..100 do :
        for k in 1..100 do :
            for l in 1..100 do :
                if FY[k,l]>0
                    aux :=intégrale de Phi(x,y,z,y_k,y_{k+1},y'_l,y'_{l+1}) dx dy dz

```

```

        sur  $P_{i,j}$  :
        if aux>0  $\eta[i] := \eta[i] + \text{aux} \ln(\text{aux}/(\text{FX} * \text{FY}[k,1]))$  :
        end if :
    end do :
end do :
 $\eta[i] := \eta[i] / \text{HO}$  :
end do :
```

Les deux dernières lignes pouvant être modifiées comme au paragraphe 4.1.1.

Malheureusement la remarque alors mentionnée est encore valable : il est impossible pour Maple (et donc probablement pour tout autre logiciel de calcul formel) de calculer des intégrales de fonctions définies par morceaux.

Dans le cas discret, il suffit d'adapter légèrement le programme pour qu'il fasse des histogrammes à plusieurs dimensions sur Y ; les modifications sont assez simples. Compte-tenu du rôle symétrique joué par X_1 et X_2 , η_1 et η_2 sont identiques. On constate que ces indices sont environ deux fois plus grands que leurs homologues unidimensionnels : cela devrait pouvoir s'expliquer à l'aide d'une étude théorique plus poussée.

Comme prévu lors de l'étude de la convergence, la stabilisation des indices nécessite plus d'un million de points. Cela nécessiterait d'attendre plusieurs semaines avec le programme R , mais s'exécute en temps très raisonnable avec l'algorithme C++.

Le tableau 4 présente les résultats obtenus (toujours $n = 100$, $a = 7$ et $b = 0.1$).

Remarque 7 *On aurait obtenu le même type de résultat en effectuant la somme des indices basés sur la variance pour Y_1 et Y_2 . Ce n'est cependant qu'un cas particulier, car en général seule l'entropie se prête à un calcul théorique multivarié.*

4.1.4 Influences des couples de variables

Il s'agit à présent de mesurer l'influence qu'a une paire de variables sur la sortie (on pourrait bien entendu généraliser à k variables, $k \leq s$ où s est le nombre de paramètres d'entrée, mais il y aurait très vite à un nombre

S	η_1	η_2	η_3
1000 (<1s)	0.5219	0.5243	0.4899
5000 (<1s)	0.3373	0.3381	0.2327
10000 (<1s)	0.3327	0.3322	0.2404
50000 (<1s)	0.2763	0.2768	0.1600
100000 (1s)	0.2448	0.2448	0.1358
500000 (2s)	0.2182	0.2182	0.1132
1000000 (4s)	0.2213	0.2212	0.1081
5000000 (15s)	0.1878	0.1879	8.400e-2
10000000 (30s)	0.1856	0.1856	8.507e-2

TAB. 4 – Indices de sensibilité pour une sortie bidimensionnelle

exponentiel d'indices à calculer).

On effectue donc ici les calculs des indices de sensibilité $\eta_{i,j}$ selon la formule donnée à la remarque 6 du paragraphe 3.1.2, rappelée ici :

$$\eta_{i,j} = \frac{I((X_i, X_j), Y)}{H(Y)},$$

avec

$$I((X_i, X_j), Y) = \int_{x_i \in \mathcal{X}_i} \int_{x_j \in \mathcal{X}_j} \int_{y \in \mathcal{Y}} f(x_i, x_j, y) \ln \frac{f(x_i, x_j, y)}{f(x_i, x_j) f(y)} d\mu_i(x_i) d\mu_j(x_j) d\mu(y).$$

Au niveau discret où l'on travaille cela se traduit par

$$\eta_{i,j} = - \left[\sum_{k,l=1}^n F_{k,l,m}^{i,j} \ln \left(\frac{F_{k,l,m}^{i,j}}{\int_{x=x_k^i}^{x_{k+1}^i} f_{X_i}(x) dx \int_{x=x_l^j}^{x_{l+1}^j} f_{X_j}(x) dx \int_{y=y_m}^{y_{m+1}} f_Y(y) dy} \right) \right] H_0^{-1},$$

$$\text{où } H_0 = - \sum_{k=1}^n \int_{y=y_k}^{y_{k+1}} f_Y(y) dy \ln \left(\int_{y=y_k}^{y_{k+1}} f_Y(y) dy \right)$$

$$\text{et } F_{k,l,m}^{i,j} = \int_{x=x_k^i}^{x_{k+1}^i} \int_{x=x_l^j}^{x_{l+1}^j} \int_{y=y_k}^{y_{k+1}} f_{(X_i, X_j, Y)}(x_i, x_j, y) dx_i dx_j dy.$$

Les résultats sont donnés dans le tableau 5, toujours pour 100 sous-intervalles de discrétisation. Les temps comprennent le calcul des deux types d'indices.

Les valeurs obtenues sont logiques par rapport aux résultats du cas unidimensionnel.

S	$\eta_{1,2}$	$\eta_{1,3}$	$\eta_{2,3}$
1000 (<1s)	0.7927	0.8089	0.6941
5000 (<1s)	0.5719	0.3692	0.4081
10000 (<1s)	0.5448	0.3365	0.4343
50000 (<1s)	0.5170	0.2794	0.3862
100000 (1s)	0.4707	0.2362	0.3543
500000 (2s)	0.4464	0.2172	0.3338
1000000 (3s)	0.4528	0.2490	0.3312
5000000 (12s)	0.4128	0.2023	0.3049
10000000 (24s)	0.4098	0.2030	0.3033

TAB. 5 – Indices de sensibilité η d'ordre 2

Par comparaison à ces derniers indices, les influences des couples sont deux à trois fois plus grandes que les influences simples : c'est logique, deux paramètres réunis perturbent plus la sortie qu'un seul.

On justifie ceci proprement en revenant à la définition :

$$\begin{aligned}
I((X_i, X_j), Y) &= H(X_i, X_j|Y) - H(Y) \\
&\geq \max \{H(X_i|Y) - H(Y), H(X_j|Y) - H(Y)\} \\
&= \max \{I(X_i, Y), I(X_j, Y)\},
\end{aligned}$$

car $H(X_i, X_j) \geq \max \{H(X_i), H(X_j)\}$ comme on le vérifie facilement.

Remarque 8 *Dans le cadre des indices de sensibilité basés sur la variance, les indices de couples représentent uniquement les interactions entre deux paramètres d'entrée. En revanche ici ces indices donnent l'information globale apportée par deux paramètres : on pourrait alors se dire que ces calculs sont inutiles.*

Avant de pouvoir donner une justification théorique on déterminera néanmoins ces indices.

4.2 Fonction g de Sobol

Il s'agit de la fonction suivante :

$$g_a : \mathbb{R}^p \longrightarrow \mathbb{R}, \quad x = (x_1, \dots, x_p) \longrightarrow \prod_{k=1}^p g_k(x_k),$$

où $g_k(x_k) = \frac{|4x_k-2|+a_k}{1+a_k}$, x_1, \dots, x_p étant les résultats de variables aléatoires toutes de loi uniforme sur l'intervalle $[0,1]$.

On choisit $n = 8$, $a_1 = a_2 = 3$, $a_3 = \dots = a_8 = 0$: ce sont des valeurs retenues pour des tests sur les indices de sensibilité basés sur la variance. L'allure de la loi de $Y = g(X_1, \dots, X_8)$ est donnée sur la figure 22.

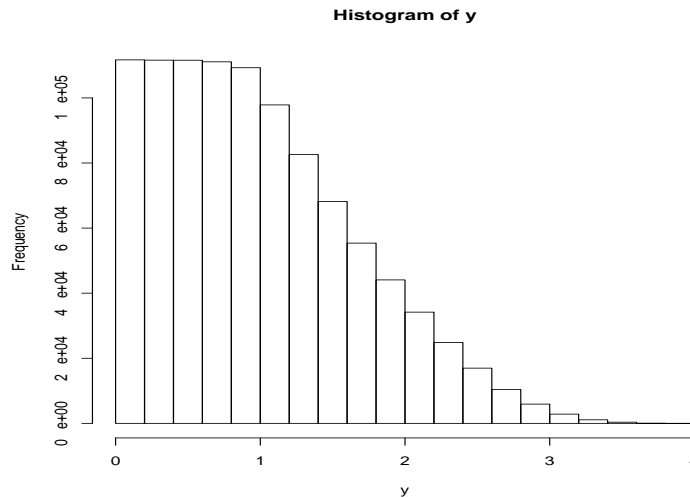


FIG. 22 – Histogramme de la distribution de sortie

4.2.1 Indices théoriques

On essaye ici d'expliciter les indices de sensibilité en fonction des a_k .

Commençons par une remarque simple : les variables aléatoires $|4X - 2|$ et $2X$ ont même fonction de répartition si X suit une loi uniforme sur $[0,1]$:

$$\begin{aligned}
 \mathbb{P}(|4X - 2| \leq t) &= \mathbb{P}(-t \leq 4X - 2 \leq t) \\
 &= \mathbb{P}\left(\frac{-t+2}{4} \leq X \leq \frac{t+2}{4}\right) \\
 &= \min\left(\frac{t+}{2}, 1\right) \\
 &= \mathbb{P}(2X \leq t).
 \end{aligned}$$

On peut donc réécrire la fonction g de Sobol comme suit :

$$g_a : \mathbb{R}^p \longrightarrow \mathbb{R}, \quad x = (x_1, \dots, x_p) \longrightarrow \prod_{k=1}^p g_k(x_k),$$

$g_k(x_k) = \frac{2x_k + a_k}{1 + a_k}$, x_1, \dots, x_p étant définis comme en introduction.

Par un argument de symétrie (n et les a_k étant fixés comme indiqué en introduction) il suffit alors de calculer les indices η_1 et η_3 . On cherche à évaluer l'expression suivante pour $i \in \{1, 3\}$:

$$\eta_i = \sum_{j,k \in [1..n]} \int_{x_j}^{x_{j+1}} \int_{y_k}^{y_{k+1}} f_i(x, y) dx dy \ln \frac{\int_{x_j}^{x_{j+1}} \int_{y_k}^{y_{k+1}} f_i(x, y) dx dy}{\int_{x_j}^{x_{j+1}} f_i(x) dx \int_{y_k}^{y_{k+1}} f_i(y) dy}.$$

Il suffit en fait de parvenir à calculer $f(y)$: la méthode utilisée fonctionnera aussi pour $f_i(x, y)$, et on pourra effectuer les intégrales puis la somme. Le produit fait penser à une relation de récurrence ; essayons :

$$\begin{aligned} f_n(y) &= \int_{x_n=0}^1 f_n^{X_n=x}(y) f(x_n) dx_n \\ &= \int_{x_n=0}^1 f_{n-1}\left(\frac{(1+a_n)y}{2x_n+a_n}\right) dx_n, \\ &\dots \end{aligned}$$

En calculant les premières intégrales précédentes, on se rend compte que les expressions obtenues deviennent très vite trop lourdes pour pouvoir être intégrées : le calcul des indices est donc impossible par cette méthode.

Une autre idée serait de calculer directement les intégrales présentes dans la somme par

$$\int_{y_k}^{y_{k+1}} f(y) dy = \int_{x \in \mathcal{X} / \psi(x) \in [y_k, y_{k+1}]} f(x_1) \dots f(x_s) dx_1 \dots dx_s, \text{ mais une nouvelle difficulté se dresse alors devant nous : la résolution (quasi impossible) de l'équation } \psi(x) \in [y_k, y_{k+1}].$$

Échec total : il ne reste plus qu'à lancer l'algorithme pour trouver les indices numériques. Ce n'est pas si grave car la nature de la convergence est bien connue à présent.

S	η_1	η_3
100 (<1s)	0.8294	0.8291
500 (<1s)	0.5146	0.4594
1000 (<1s)	0.4770	0.4230
5000 (<1s)	0.2253	0.1610
10000 (<1s)	0.1776	0.1023
50000 (<1s)	0.1272	3.290e-2
100000 (1s)	0.1234	3.745e-2
500000 (3s)	0.1084	2.252e-2
1000000 (7s)	0.1114 (36.5%)	1.441e-2 (4.5%)

TAB. 6 – Indices de sensibilité basés sur l’entropie

4.2.2 Indices expérimentaux

Comme précédemment on calcule les indices de sensibilité avec la définition 6, à l’aide de l’algorithme C++ déjà utilisé.

Les résultats se trouvent sur le tableau 6, où seuls η_1 et η_3 sont calculés compte-tenu de la symétrie des indices. Le tableau 7 donne les résultats obtenus pour les indices basés sur la variance, toujours pour 100 répétitions du programme.

Les proportions des indices sont quasiment les mêmes dans le cas de l’entropie et des indices totaux basés sur la variance. Ainsi dans ce cas, malgré les différences entre les deux approches on arrive au même résultat en utilisant les indices relatifs.

4.2.3 Influences des couples de variables

L’objectif est le même que dans le paragraphe analogue dans le cas de la fonction d’Ishigami. On effectue les calculs sur les couples (X_1, X_2) , (X_1, X_3) et (X_3, X_4) uniquement grâce à la symétrie des variables.

Ici encore on constate que chaque indice $\eta_{i,j}$ est nettement plus grand que le maximum des indices simples η_i et η_j , mais plus petit que la somme $\eta_i + \eta_j$ pour les cas de $\eta_{2,3}$ et $\eta_{3,4}$. On n’a donc pas de formule analogue à celle des indices d’ordre 2 basés sur la variance en retranchant $S_i + S_j$: $\eta_{i,j}$ représente l’influence globale du couple, d’autant plus grande que la connaissance

S	S_1	S_3	St_1	St_3
100 (<1s)	0.3618 0.212	$5.048e-2$ 0.178	0.4752 0.176	$5.072e-2$ $8.65e-2$
500 (<1s)	0.3409 $8.47e-2$	$2.220e-2$ $7.99e-2$	0.5084 $8.12e-2$	$4.270e-2$ $4.15e-2$
1000 (<1s)	0.3416 0.0586323	$2.317e-2$ 0.0583069	0.4981 0.0646852	$3.894e-2$ 0.0274674
5000 (2s)	0.3284 0.0273321	$2.222e-2$ 0.0274958	0.4913 0.0284823	$4.047e-2$ 0.0115594
10000 (4s)	0.3324 $1.76e-2$	$2.149e-2$ $1.540e-2$	0.5004 $1.78e-3$	$4.065e-2$ $8.65e-3$
50000 (19s)	0.3297 $7.78e-3$	$2.011e-2$ $7.50e-3$	0.4985 $8.15e-3$	$4.041e-2$ $3.68e-3$
100000 (35s)	0.3297 $5.84e-3$	$2.073e-2$ $5.54e-3$	0.4973 $5.63e-3$	$4.068e-2$ $2.85e-3$
500000 (3min)	0.3289 $2.58e-3$	$2.029e-2$ $2.35e-3$	0.4970 (40.2%) $2.63e-3$	$4.050e-2$ (3.3%) $1.15e-3$

TAB. 7 – Indices de sensibilité basés sur la variance

des deux paramètres apporte de l'information sur la sortie. Cette étude est donc intéressante en complément du calcul des indices simples, permettant de conserver des paramètres qui ont une influence faible mais un indice $\eta_{i,j}$ élevé.

Les résultats sont indiqués dans le tableau 8.

Le temps de calcul pour 10 millions de points est bien plus grand que le double du temps nécessaire pour moitié moins de points ; en effet, l'ordinateur manque alors de mémoire vive et doit stocker des données sur le disque, d'accès beaucoup plus lent.

On remarque que seul $\eta_{1,2}$ se démarque des autres indices ; c'est assez logique : les paramètres 1 et 2 sont les deux plus influents d'après l'étude effectuée.

À titre de comparaison, on donne aussi les résultats obtenus dans le cas de la variance pour les indices $S_{i,j}$ dans le tableau 9

On constate que l'entropie est plus souple que la variance, accordant le quart de l'influence totale au couple (2,3) tandis que la variance ne lui octroie que 6% d'importance, quasiment tout revenant au couple (1,2).

S	$\eta_{1,2}$	$\eta_{2,3}$	$\eta_{3,4}$
1000 (<1s)	0.8798	0.8378	0.8183
5000 (<1s)	0.9117	0.8968	0.8899
10000 (<1s)	0.5388	0.4044	0.3515
50000 (<1s)	0.6301	0.5474	0.5103
100000 (1s)	0.4411	0.2867	0.2196
500000 (4s)	0.3431	0.1707	9.089e-2
1000000 (10s)	0.3662	0.1665	8.043e-2
5000000 (40s)	0.3531	0.1307	2.943e-2
10000000 (2min08s)	0.3320 (69.8%)	0.1205 (25.3%)	2.331e-2 (4.9%)

TAB. 8 – Indices de sensibilité $\eta_{i,j}$ basés sur l'entropie

4.3 Limitations sur les indices : non-discrimination

On montre ici quelques situations où les indices de Sobol ou basés sur l'entropie ne permettent pas de conclure.

4.3.1 Exemple où la variance n'apporte pas d'information

L'étude réalisée à la section 2.5 nous indique que pour construire un exemple où les indices de sensibilité basés sur l'entropie sont les seuls intéressants il suffit de prendre deux paramètres d'entrée : X_1 et X_2 , tels que $\mathbb{E}[Y|X_1]$ et $\mathbb{E}[Y|X_2]$ suivent respectivement une loi semblable à celle tracée figure 14 *a.* et une semblable à celle représentée sur la figure 14 *b.* En effet dans ce cas les deux variances sont identiques, mais pas les entropies.

Par intuition on pose alors X_1 et X_2 variables aléatoires sur $[-1,1]$ de densités respectives $f(x) = \mathbb{1}_{[-1,-\frac{1}{2}]}(x) + \mathbb{1}_{[\frac{1}{2},1]}(x)$ et $f(x) = 5\mathbb{1}_{[-\alpha,-\alpha+\frac{1}{10}]}(x) + 5\mathbb{1}_{[\alpha-\frac{1}{10},\alpha]}(x)$, où $\alpha \simeq 0.8132$ comme au paragraphe 2.4.1. La sortie Y vérifie $Y = \psi(X_1, X_2)$ avec

$$\begin{aligned} \psi : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\mapsto x + y. \end{aligned}$$

Après quelques calculs élémentaires on obtient comme souhaité :

$$\begin{aligned} \mathbb{E}[Y|X_1] &= X_1 \\ \mathbb{E}[Y|X_2] &= X_2, \end{aligned}$$

S	$S_{1,2}$	$S_{2,3}$	$S_{3,4}$
100 (<1s)	$6.895e-2$ 0.231	$-1.372e-2$ 0.194	$-2.395e-2$ 0.186
500 (<1s)	0.105505 0.103	0.00150859 $8.39e-2$	-0.00232615 $8.05e-2$
1000 (1s)	0.1021 $7.24e-2$	$3.258e-3$ $5.76e-2$	$-2.644e-3$ $5.66e-2$
5000 (3s)	0.1052 $3.41e-2$	$4.262e-3$ $2.70e-2$	$-1.871e-3$ $2.66e-2$
10000 (5s)	0.1105 $2.31e-2$	$5.861e-3$ $1.58e-2$	$-6.258e-4$ $1.57e-2$
50000 (27s)	0.1103 $9.90e-3$	$7.097e-3$ $7.42e-3$	$7.299e-4$ $7.51e-3$
100000 (52s)	0.1098 $7.33e-3$	$6.895e-3$ $5.70e-3$	$2.472e-4$ $5.57e-3$
500000 (4min17s)	0.1102 $3.11e-3$ (93.4%)	$7.135e-3$ $2.40e-3$ (6%)	$7.054e-4$ $2.35e-3$ (0.6%)

TAB. 9 – Indices de sensibilité $S_{i,j}$ de Sobol

ce qui est normal car X_1 et X_2 sont indépendantes.

Donc l'analyse de sensibilité basée sur la variance donne une égale influence aux deux paramètres X_1 et X_2 .

En revanche en calculant les entropies conditionnelles on trouve

$$H(Y|X_1) = -\ln(5)$$

$$H(Y|X_2) = 0.$$

Ainsi pour cet exemple $\eta_1 > \eta_2$, ce qui correspond bien à l'intuition : X_1 est un paramètre plus influent que X_2 car sa distribution de probabilités est moins précise : c'est X_1 qui apporte le plus d'incertitude, son entropie étant plus grande que celle de X_2 .

4.3.2 Exemple où l'entropie n'apporte pas d'information

S'inspirant de l'exemple du paragraphe 4.3.1, on construit un modèle où l'entropie n'est pas discriminante.

Soient X_1 et X_2 des variables aléatoires sur $[-1,1]$, de densités respectives $f(x) = \mathbb{1}_{[-1,-\frac{1}{2}]}(x) + \mathbb{1}_{[\frac{1}{2},1]}(x)$ et $f(x) = \mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]}(x)$. La sortie Y vérifie $Y = \psi(X_1, X_2)$ avec

$$\begin{aligned}\psi : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\mapsto x + y.\end{aligned}$$

Alors, comme au paragraphe analogue dans le cas de la variance on trouve :

$$\begin{aligned}\mathbb{E}[Y|X_1] &= X_1 \\ \mathbb{E}[Y|X_2] &= X_2.\end{aligned}$$

Ainsi l'indice de sensibilité basé sur la variance est plus élevé pour X_1 que pour X_2 . C'est logique, car X_1 apporte plus de variabilité que X_2 sur la sortie.

En ce qui concerne les indices η , un rapide calcul donne $H(Y|X_1) = H(Y|X_2) = 0$: l'entropie n'est donc pas discriminante dans ce cas. Les indices ρ définis au paragraphe 2.5.3 différencieraient aussi X_1 et X_2 . Ceux-ci ont-ils un autre intérêt ?

4.3.3 Les indices de sensibilité ρ

Les indices ρ (définition 7) devraient être concluant dans les deux cas mentionnés aux deux premiers paragraphes. Essayons de les calculer sur nos deux fonctions test.

Rappel de la définition :

$$\rho(X : Y) = c^{-H(Y|X)} + e^{\frac{V(\mathbb{E}[Y|X])}{V_u} - 1},$$

V_u étant la variance d'une loi uniforme sur l'espace de variation de Y et c un coefficient >1 . On choisit ici $c = 1.1$.

L'algorithme est un mélange de celui calculant les indices de Sobol et ceux basés sur l'entropie. On le répète 100 fois, comme précédemment, avec les mêmes valeurs des paramètres.

Les valeurs obtenues pour les indices normalisés des fonctions hs et g sont indiqués respectivement dans les tableaux 10 et 11.

On remarque que les valeurs des indices respectent bien le classement donné par l'analyse basée sur la variance ou l'entropie, mais sont très rapprochées ; ceci est sans doute dû aux exponentiations.

S	ρ_1	ρ_2	ρ_3
100 (<1s)	0.3332 $5.29e-3$	0.3408 $4.88e-3$	0.3260 $5.09e-3$
500 (<1s)	0.3373 $1.53e-3$	0.3395 $1.62e-3$	0.3233 $1.48e-3$
1000 (<1s)	0.3400 $1.15e-3$	0.3342 $1.06e-3$	0.325886 $1.07e-3$
5000 (2s)	0.3355 $4.08e-4$	0.3405 $3.88e-4$	0.3240 $4.46e-4$
10000 (5s)	0.3345 $2.36e-4$	0.3409 $2.77e-4$	0.3246 $2.46e-4$
50000 (25s)	0.3346 $1.10e-4$	0.3406 $1.27e-4$	0.3248 $1.25e-4$
100000 (46s)	0.3346 $8.54e-5$	0.3403 $8.43e-5$	0.3250 $8.96e-5$
500000 (3min50s)	0.3347 $3.65e-5$	0.3402 $3.17e-5$	0.3251 $3.68e-5$

TAB. 10 – Indices de sensibilité ρ_i appliqués à la fonction d’Ishigami

L’échelle des indices est donc respectée, mais les proportions ne le sont pas du tout ; ce sont cependant les seuls indices parmi ceux présentés ici qui différencient à la fois les figures 14 *a.* et *b.* ainsi que les figures 19 *a.* et *b.*

En pratique il sera beaucoup plus intéressant de sommer les indices de Sobol avec ceux basés sur l’entropie : ainsi on différencie les figures 14 et 19 comme pour les indices ρ . Le seul défaut notable est que cette somme est minimale pour une combinaison de pics de Dirac, pas uniquement pour un pic seul. Il resterait à étudier le cas où la somme est maximale, qui doit se situer assez proche (visuellement parlant) de la distribution uniforme.

5 Conclusion

Même si l’intérêt des indices de sensibilité utilisant l’entropie est indéniable : symétrie des rôles de l’entrée et de la sortie, nouveau point de vue par rapport à la variance ..etc, on doit tenir compte de la non-uniformité de leur répartition dans $[0,1]$ à cause du caractère logarithmique de l’entropie.

De plus l’entropie n’est pas toujours discriminante comme le montre le dernier exemple. La tentative d’unifier variance et entropie échoue à cause

S	ρ_1	ρ_3
100 (<1s)	0.5079 $6.84e-3$	0.4921 $6.84e-3$
500 (1s)	0.5085 $2.91e-3$	0.4915 $2.91e-3$
1000 (2s)	0.5079 $1.77e-3$	0.4921 $1.77e-3$
5000 (6s)	0.5081 $9.84e-4$	0.4919 $9.84e-4$
10000 (8s)	0.5086 $8.10e-4$	0.4914 $8.10e-4$
50000 (1min)	0.5085 $4.84e-4$	0.4915 $4.84e-4$
100000 (1min23s)	0.5080 $4.14e-4$	0.4920 $4.14e-4$
500000 (6min)	0.5074 $2.84e-4$	0.4926 $2.84e-4$

TAB. 11 – Indices de sensibilité ρ_i appliqués à la fonction g de Sobol

de la trop grande différence entre ces deux notions, et il faut donc calculer les indices séparément.

Que nous apporte le calcul de ces indices? Ceux basés sur la variance nous renseignent sur la part de la variabilité de la sortie expliquée par un paramètre, tandis que ceux basés sur l'entropie nous indique l'information globale sur la sortie apportée par un paramètre.

L'utilisation de l'entropie en pratique apparaît donc plus comme un complément à l'étude basée sur la variance qu'un substitut : les indices calculés ici peuvent servir à affiner une analyse utilisant la variance en cas d'ex-æquo par exemple, ou même précéder l'analyse basée sur la variance. Les calculs pratiques peuvent s'effectuer très rapidement grâce à l'échantillonnage déterministe, mais en revanche la convergence est très lente (apparaissant uniquement pour $S > n^{\alpha+d}$ où α est l'ordre des indices) lorsque l'on veut calculer des indices d'ordre > 2 ou multi-sorties.

Finalement, les indices η sont aux indices de Sobol ce que la variance est à l'entropie : deux points de vue totalement différents, se complétant.

Annexe 1 : recherche de racines

On revient ici sur la fonction suivante

$$\begin{aligned}\psi &: \mathbb{R}_{+*} \longrightarrow \mathbb{R} \\ (x, y) &\mapsto \ln \frac{2x}{2y+1} + \frac{2y}{2y+1} - \frac{1}{2}.\end{aligned}$$

mentionnée au paragraphe 2.5.1.

On veut montrer que quel que soit $\varepsilon > 0$, il existe un couple d'entiers (m, n) tel que $|\psi(m, n)| \leq \varepsilon$.

Commençons par réécrire ψ :

$$\psi(x, y) = \ln(x) - \theta(y),$$

où $\theta(y) = \ln(2y+1) - \frac{2y}{2y+1} + \frac{1}{2} - \ln(2)$.

Soient alors $\varepsilon > 0$, et $N \in \mathbb{N}^* / \frac{1}{N} \leq \varepsilon$. Soit aussi $m \in \mathbb{N}^*$ tel que $\theta(m) \geq N + 1$. $\theta(m)$ est alors compris dans l'intervalle $I = [e^{\theta(m)-\frac{1}{2}}, e^{\theta(m)+\frac{1}{2}}]$ (à condition que m ne soit pas trop petit, ce que l'on peut supposer sans restriction). De plus, pour tout n entier dans cet intervalle on a d'après l'inégalité des accroissements finis :

$$\begin{aligned}|\ln(n+1) - \ln(n)| &\leq \frac{1}{n} \\ &\leq \frac{1}{N} \\ &\leq \varepsilon.\end{aligned}$$

On en déduit alors qu'il existe un entier n dans l'intervalle I vérifiant $|\ln(n) - \theta(m)| \leq \varepsilon$, ce qui prouve le résultat annoncé.

□

Annexe 2 : amélioration du programme

Nous allons ici disséquer les boucles du programme calculant les indices basés sur l'entropie morceau par morceau afin d'effectuer le plus de calculs possibles en parallèle et/ou plus rapidement - R utilise des algorithmes très rapides écrits en C, mais non parallélisés. Par ailleurs, on constate expérimentalement que certaines instructions sont assez, voire très lentes. L'implémentation en langage C++ est donnée en fichier annexe.

```
[...]  
# on repete le programme R fois  
for(z in 1:R) {
```

L'idée la plus simple tout en étant très efficace dès que $R > 1$ est d'effectuer cette boucle en parallèle : il suffit ensuite de regrouper les résultats pour calculer moyennes, écarts-type ..etc. En revanche lorsque $R = 1$ (dans le cas de l'entropie avec échantillonnage déterministe), il faut chercher ailleurs.

```
#Tirage de l'échantillon  
if (LHS) X<-samplingLHS(dim_x,N=S,lois=lois,paramlois=paramlois)  
else X<-samplingSimple(dim_x,N=S,lois=lois,paramlois=paramlois)  
  
#Transformation logarithmique éventuelle  
for(i in 1:dim_x)  
  if (logx[i]) X[,i]<-log(X[,i])
```

Il peut être intéressant d'effectuer l'échantillonnage en parallèle si celui-ci est coûteux. Les boucles C/C++ correspondant à la seconde partie ne sont ni plus rapides ni plus lentes sur une machine monoprocesseur, on se contente donc de les réécrire dans un premier temps.

```
#-----  
#Calcul de la reponse Y |  
#-----  
  
if (methode>0) Y<-predictSR(modele,X,methode)  
else Y<-f(X)  
if (logy) Y<-exp(Y)
```

Même remarque.

```
#transformation eventuelle si Y est de dimension 1
if (NCOL(Y)==1) {
  aux<-array(Y,dim=c(S,1))
  Y<-aux
}

#precalculs: bornes sur les echantillons et sur Y
infX<-vector(length=dim_x)
supX<-vector(length=dim_x)
for (i in 1:dim_x) infX[i]<-min(X[,i])
for (i in 1:dim_x) supX[i]<-max(X[,i])
dim_y<-NCOL(Y)
infY<-vector(length=dim_y)
supY<-vector(length=dim_y)
for (i in 1:dim_y) infY[i]<-min(Y[,i])
for (i in 1:dim_y) supY[i]<-max(Y[,i])
```

Les boucles impliquées ici comportent seulement des comparaisons et des affectations, donc créer des threads pour ces tâches risque de se révéler plus long que l'algorithme séquentiel pour S pas trop grand : on déroule ici aussi les boucles cachées par la syntaxe de R.

Ensuite bien sûr on affecte quelques centaines de calculs à chaque thread pour aller plus vite.

```
[...]
#precalcul: nombres en base n+1 = indices de estimationsY
casesY<-rep(0,S)
for (i in 1:S)
  for (j in 1:dim_y) casesY[i]<-casesY[i]+(floor((Y[i,j]
    -infY[j])/pasY[j]))*(n+1)^(j-1)
casesY<-casesY+1
```

La situation est intéressante ici : paralléliser la boucle extérieure devrait donner un gain de temps appréciable étant donné la lourdeur de l'instruction exécutée à chaque passage (si $\text{dim}_y > 3$, par exemple), d'autant plus que les instructions sont indépendantes d'un passage à l'autre. En revanche la boucle intérieure agit dim_y fois sur la même case mémoire, et ne se prête

donc pas à une programmation EREW (Exclusive Read Exclusive Write), ce qui complique passablement les choses. On se contentera donc de paralléliser la première boucle lorsque la dimension de Y dépassera 3, considérant que dim_y est négligeable devant S .

Pour une première approche on écrit simplement ces boucles en séquentiel : déjà on observe un énorme gain de temps par rapport au programme R.

```
#precalcul: tableau de repartition spatiale de Y
for (i in 1:S) estimationsY[casesY[i]]<-estimationsY[casesY[i]]+1
```

Inutile de réécrire cela dans un premier temps : l'addition et l'affectation réunies ne sont pas assez coûteuses en temps pour justifier l'utilisation de plusieurs threads.

```
[...]
if (!rel) {
  tmp<-rep(0,(n+1)^dim_y)
  for (j in 1:(n+1)^dim_y)
    if (estimationsY[j]>0) tmp[j]<-
      estimationsY[j]*log(estimationsY[j])
  H0<-sum(tmp)
}
```

Même remarque.

```
[...]
#precalcul: tableau de repartition spatiale de Xi, puis de (Xi,Y)
for (j in 1:S) {
  aux<-floor((X[j,i]-infX[i])/pasX)+1
  estimationsX[aux]<-estimationsX[aux]+1
  casesX[j,i]<-aux
  if (Si) estimationsXY[aux,casesY[j]]<-
    estimationsXY[aux,casesY[j]]+1
}
```

On peut être tenté de paralléliser cette boucle dont la première instruction est assez coûteuse, avec un accès concurrent en mémoire car les variables `aux` successives peuvent reprendre les mêmes valeurs d'une itération sur l'autre.

Chaque thread effectue ainsi les instructions présentes dans la boucle for, avec les variables `estimationsX` et `estimationsXY` partagées.

Ici encore pour une première approche on en reste à la boucle séquentielle.

```
[...]
tmp<-matrix(0,nrow=n,ncol=(n+1)^dim_y)
# dans cet ordre, car les affectations sont plus rapides que
# les additions
for (j in 1:n)
  for (k in 1:(n+1)^dim_y) {
    if (estimationsX[j]>0 && estimationsY[k]>0
        && estimationsXY[j,k]>0)
      tmp[j,k]<-estimationsXY[j,k]*log(estimationsXY[j,k]/
        (estimationsX[j]*estimationsY[k]))
  }
HK[z,i]<-sum(tmp)
```

Même analyse que pour la situation analogue dans le cas du calcul de $H(Y)$.

```
[...]
# precalcul: tableau de repartition spatiale de (Xi,Xj,Y)
# et (Xi,Xj):
for (k in 1:S) {
  estimationsXX[casesX[k,i],casesX[k,j]]<-
  estimationsXX[casesX[k,i],casesX[k,j]]+1
  estimationsXXY[casesX[k,i],casesX[k,j],casesY[k]]<-
  estimationsXXY[casesX[k,i],casesX[k,j],casesY[k]]+1
}
```

De même ici, les instructions sont déjà très rapides.

```
[...]
tmp<-array(0,dim=c(n,n,(n+1)^dim_y))
for (k in 1:n)
  for (l in 1:n)
    for (m in 1:(n+1)^dim_y) {
      if (estimationsY[m]>0 && estimationsXX[k,l]>0
          && estimationsXXY[k,l,m]>0)
```

```
        tmp[k,l,m]<-
        estimationsXXY[k,l,m]*log(estimationsXXY[k,l,m]/
        (estimationsXX[k,l]*estimationsY[m]))
    }
HK2[z,i,j]<-sum(tmp)
```

Et on conclut comme ci-dessus, avec encore n fois plus de calculs cependant.

Références

- [1] A. Antoniadis. Analyse de la variance fonctionnelle pour l'analyse de sensibilité globale. In *GDR MOMAS, CIRM, Marseille*, Novembre 2003.
- [2] B. Iooss and F. Van Dorpe. *Analyses de sensibilité du code de calcul d'impact dosimétrique GASCON*. Note Technique CEA DEN/CAD/DER/STR/LCFR 2002/0032 et DEN/CAD/DED/SAMRA 02/053, 2002.
- [3] J. Jacques. *Contributions à l'analyse de sensibilité et à l'analyse discriminante généralisée*. Thèse de l'Université Joseph Fourier, Grenoble 1, 2005.
- [4] E.T. Jaynes. *Probability Theory : The Logic of Science*, chapter 11 - Discrete prior probabilities : the entropy principle, pages 346–351. Cambridge University Press, 2003.
- [5] B. Krzykacz-Hausmann. Epistemic sensitivity analysis based on the concept of entropy. In *Proceedings of SAMO 2001*, pages 31–35, Madrid, 2001. CIEMAT.
- [6] B.D. Middleton. *Use of Information Theory Techniques with System Dynamics Models*. Doctor of Philosophy in Nuclear Science and Engineering at the Massachusetts Institute of Technology, 2005.
- [7] M-J. Rendas. *Théorie de l'Information, Notes de Cours 2005-2006*. SIC-SICOM; Université de Nice.
- [8] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communication*, 145 :280–297, 2002.
- [9] I.M. Sobol. Sensitivity estimates for non linear mathematical models. *Mathematical Modelling and Computational Experiments*, 1 :407–414, 1993.
- [10] E. Volkova, F. Van Dorpe, and B. Iooss. *Modélisation du transport de ^{90}Sr en milieu poreux saturé et analyse de sensibilité du modèle : application sur un site de stockage temporaire de déchets radioactifs (CRR Kurchatov Institute, Russie)*. Note Technique CEA DEN/CAD/DTN/SMTM/LMTE 2005/63, 2005.

Je tiens à remercier Bertrand Iooss pour tous ses conseils, et Sébastien Da-
Veiga pour la méthode d'optimisation continue utilisée en 2.4.1 et 2.4.2.