

Société de Calcul Mathématique SA
Outils d'aide à la décision
depuis 1995



Méthodes probabilistes
pour l'analyse des productions d'un process industriel :

Quelles productions sont compatibles entre elles ?

par Chérif Seddik et Guillaume Bartholomeüs

Société de Calcul Mathématique SA

mai 2024

I. Introduction

Un cultivateur peut vouloir utiliser un même champ pour diverses cultures : des arbres fruitiers, des plantes fourragères (luzerne), des légumes, des fleurs, etc. Evidemment, les pratiques agricoles seront différentes d'une culture à l'autre : engrais, traitements insecticides, arrosage, labour, etc. Selon les pratiques choisies, certaines cultures seront favorisées et d'autres non ; on peut ainsi parler de "réglage" des pratiques. Pour un réglage donné, certaines cultures vont être compatibles avec ce réglage, d'autres non. L'agriculteur sait cela de toute éternité.

Il en va différemment d'un process industriel, toujours beaucoup plus récent que les pratiques agricoles. Nous prendrons l'exemple d'un process de traitement de déchets, réalisé par la société Befesa Valera ; il s'agit d'extraire les métaux contenus dans les résidus afin qu'ils soient réintégrés dans le processus de production des clients, plutôt que d'être envoyés en filière d'élimination. A partir des résidus, le process industriel extrait des métaux, que nous noterons A,B,C, etc. Il s'avère (et c'est en soi évident) que l'on ne peut pas régler le process pour obtenir, séparément, telle quantité de A, telle quantité de B, telle quantité de C, etc. : les quantités produites en sortie ne sont pas indépendantes, au sens probabiliste du mot.

Le but de cet article est d'étudier les dépendances entre productions, ce qui est fondamental pour l'industriel. Nous obtenons des résultats du type :

- Quoi que vous fassiez, si vous produisez beaucoup de A, vous produirez beaucoup de B ;
- Quoi que vous fassiez, si vous produisez beaucoup de C, vous produirez peu de D.

Le "quoi que vous fassiez" désigne évidemment l'ensemble des réglages qui ont été réalisés à ce jour, par référence à une base de données existante. Il n'est pas impossible qu'un procédé chimique différent permette de produire beaucoup de A et peu de B, beaucoup de C et beaucoup de D : nous n'avons aucun moyen de le savoir, puisque nous travaillons à partir d'un historique.

Nous remercions la société Befesa Valera pour nous avoir notifié un contrat relatif au réglage des fours (2022) ; le présent article trouve son origine dans ces travaux. Mais, bien entendu, nous respectons les obligations de confidentialité et nous ne présentons ni le process industriel, ni les valeurs numériques des données.

II. Présentation du cadre de travail

Les données représentent des concentrations d'oxydes métalliques présents en sortie de traitement des résidus lors d'une production. Ces concentrations sont exprimées en pourcentage de composition des résidus.

Au total, la base de données comporte plus de mille lignes, correspondant à des campagnes de production, et douze paramètres (colonnes), correspondant chacun à un oxyde métallique : ZnO, MnO, Cr2O3, NiO, PbO, CaO, MgO, SiO2, MoO3, Fe2O3, Al2O3, P2O5.

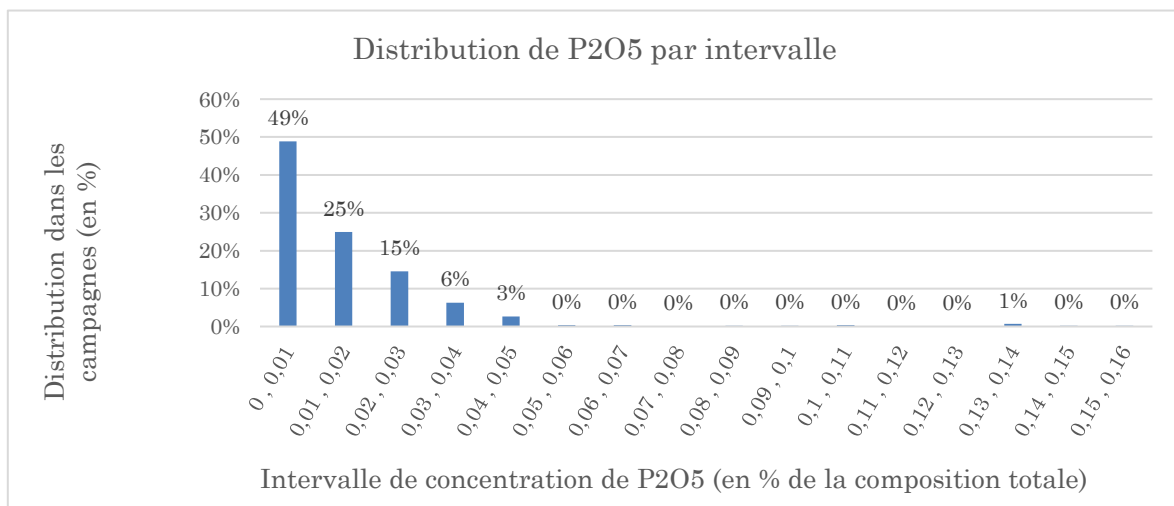
Nous choisissons une variable d'intérêt, à savoir le P2O5, qu'on notera Y , les autres paramètres sont notés $X_i, i=1, \dots, 11$. Comme expliqué plus haut, nous nous posons la question : si un réglage produit beaucoup de Y , quels autres paramètres sont nécessairement forts et quels autres nécessairement faibles ? Le choix de Y est arbitraire et on pourrait recommencer l'étude pour n'importe quel autre paramètre.

III. Mise en œuvre

Comme toujours pour une étude probabiliste d'un process, nous commençons par tracer les histogrammes pour chacune des variables. L'axe des abscisses correspond à des intervalles de concentration de l'oxyde métallique (en pourcentage de composition), et celui des ordonnées au pourcentage de distribution dans les campagnes. Par exemple, 25% des campagnes révèlent un pourcentage en P2O5 entre 0,01 et 0,02% de la composition totale.

A. Histogrammes

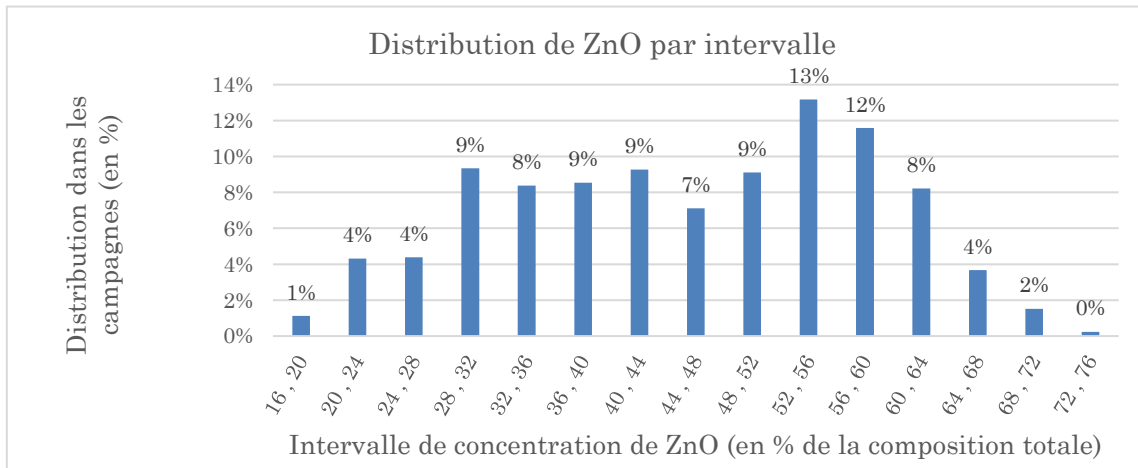
1. Distribution de Y (P2O5)



- L'histogramme est constitué de 16 intervalles de concentration (en % de la composition)
- 89% des campagnes révèlent un pourcentage en P2O5 entre 0 et 0,03% de la composition
- Plus la concentration de P2O5 augmente, moins il y a de valeurs
- La majorité des intervalles révèlent peu de phosphore

De la même façon, voici l'histogramme pour X_1 :

2. Distribution de X_1 (ZnO) :

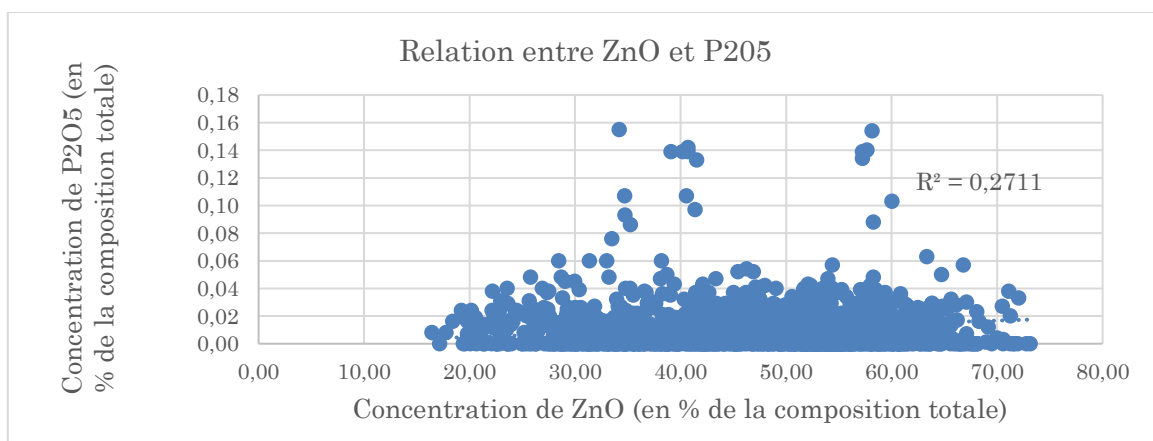


- L'histogramme est constitué de 15 intervalles de concentration (en % de la composition)
- Les valeurs de ZnO sont généralement bien réparties dans l'ensemble des intervalles, à l'exception des classes extrêmes.

Une fois l'ensemble des histogrammes tracés pour tous les paramètres, on aborde les comparaisons deux à deux. Ceci est fait en premier lieu sous forme de nuage de points, que Excel réalise très facilement.

B. Nuage de points

Nous tracerons le nuage conjoint pour Y et X_1 ; les autres se construisent de la même manière.



- Aucune dépendance linéaire n'est observée ; un coefficient de détermination de 0,27 indique une faible signification
- La majorité des résultats montrent des valeurs basses en P2O5, comme vu plus haut
- On observe des valeurs élevées en P2O5 pour des valeurs moyennes en ZnO

C. Analyse des corrélations linéaires

Excel fournit un outil permettant le calcul du coefficient de corrélation linéaire entre deux colonnes dans un tableau. Si le coefficient est proche de 1, il y a une corrélation linéaire entre les deux variables, mais deux variables peuvent être corrélées sans que cette dépendance soit linéaire (par exemple $Y = X^2$) ; ce coefficient, en pratique, est donc de peu d'intérêt.

Couples	Coefficient de détermination R^2
Phosphore vs. Zinc	0,27
Phosphore vs. Manganèse	0,18
Phosphore vs. Chrome	0,26
Phosphore vs. Nickel	0,23
Phosphore vs. Plomb	0,26
Phosphore vs. Calcium	0,26
Phosphore vs. Magnésium	0,25
Phosphore vs. Silicium	0,28
Phosphore vs. Molybdène	0,10
Phosphore vs. Fer	0,27
Phosphore vs. Aluminium	0,38

Les nuages de points ne fournissent que des indications grossières et le calcul des coefficients de corrélation ne donne rien. Nous allons donc utiliser la méthode de hiérarchisation de paramètres.

D. Méthode de hiérarchisation

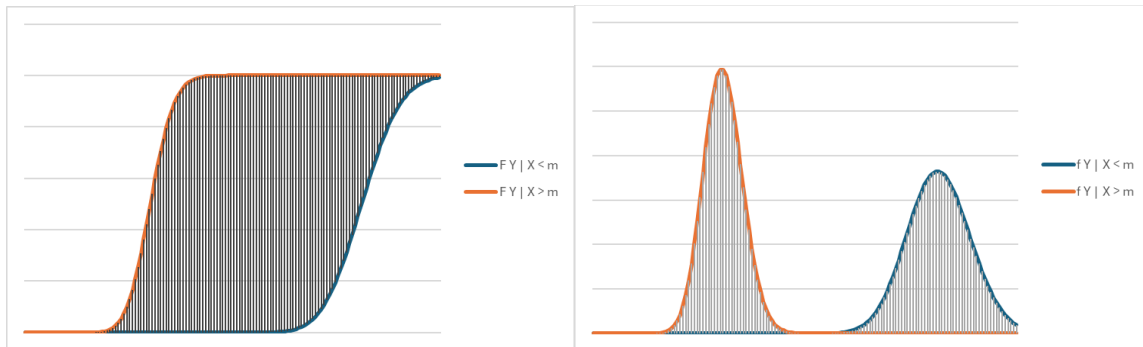
Elle est décrite dans le livre [NMP], chapitre III. Commençons par une description générale de la méthode.

1. Influence d'un paramètre sur un autre

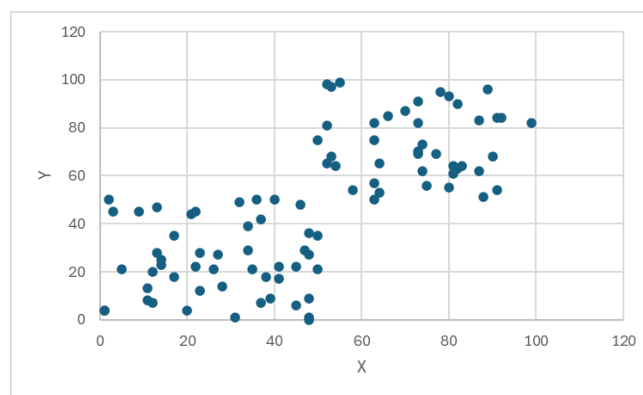
La première question qui se pose est : quand peut-on dire qu'un paramètre X a une influence sur un paramètre Y ? L'approche probabiliste grossière consiste à considérer deux modalités de X , radicalement différentes, à savoir $X > m$ et $X < m$, où m est la médiane de X , et à regarder les lois de probabilité conditionnelles de Y dans chaque cas : $Y | X > m$ et $Y | X < m$; si ces deux lois sont à peu près identiques, alors X n'a aucune influence sur Y . Pour des raisons pratiques, on regarde la fonction de répartition de Y dans chaque cas, et on calcule la différence des aires limitées par chacune des fonctions dont les courbes ne se croisent pas.

Si les courbes se croisent, l'influence n'est pas univoque : il se peut par exemple que la variable Y soit croissante pour X petit mais décroissante pour X grand.

Exemple générique de fonctions de répartition et densités dont les courbes ne se croisent pas :



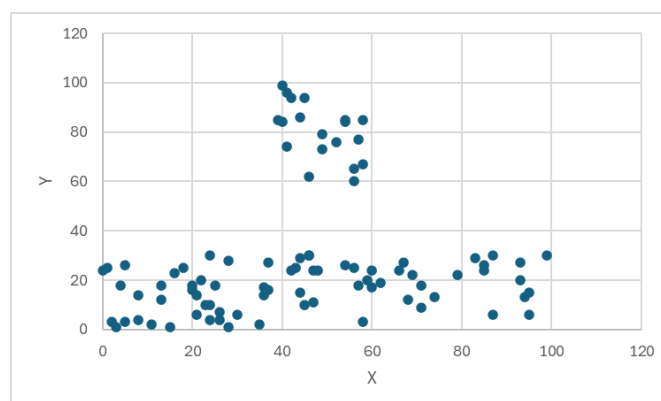
Représentation générique par nuage de points :



Supposons que, comme dans l'exemple ci-dessus, la courbe représentant $F_{Y|X>m}$ soit au-dessus de la courbe représentant $F_{Y|X<m}$. Cela signifie que la densité $f_{Y|X>m}$ est à gauche de la densité $f_{Y|X<m}$. Par conséquent, la condition $X < m$ implique Y petit et la condition $X > m$ implique Y grand. A l'inverse, Y grand implique nécessairement X grand. La différence d'aire entre les deux courbes caractérisera donc le degré d'influence de X sur Y .

Le cas symétrique correspond à une relation inverse : l'un grand implique l'autre petit.

Il est également possible d'avoir les deux implications en même temps : $X < m$ implique Y petit et $X > m$ implique Y petit, tout en ayant Y grand :

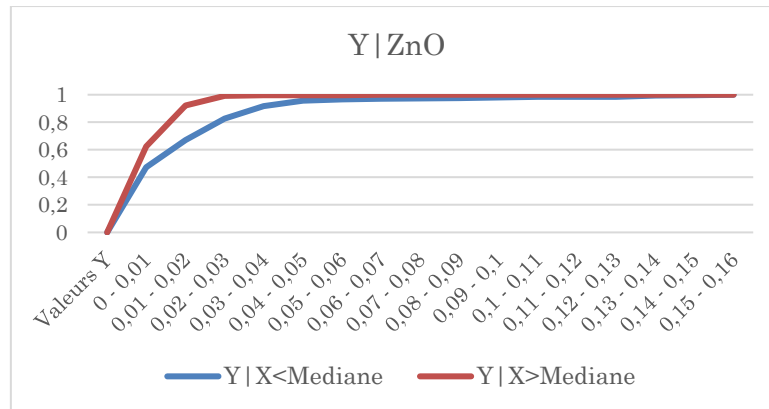


La disposition relative des courbes des fonctions de répartition nous permet ainsi de conclure quant à la nature de l'influence.

2. Application au process industriel

Voici ce qu'on obtient pour le premier couple :

a. Y (phosphore) | X_1 (zinc)



Si les deux courbes se croisent, le paramètre X n'a pas d'influence univoque sur Y et on l'élimine de la comparaison. Si elles ne se croisent pas, on calcule l'aire géométrique entre les deux courbes. Ici, on trouve $A_1 = 0.0089$.

On fait de même pour les autres paramètres ; voici le tableau récapitulatif des aires triées par ordre décroissant ; on a éliminé les paramètres pour lesquels les courbes se croisent.

A_7	0,0090
A_1	0,0086
A_{11}	0,0085
A_2	0,0061
A_8	0,0060
A_6	0,0046
A_5	0,0035

En examinant la position relative des courbes, pour les trois paramètres les plus influents, on constate que :

- X_7 : la fonction de répartition $Y | X_7 < médiane$ est au dessus, cela signifie qu'une forte concentration en phosphore implique une faible concentration en magnésium.
- X_1 : la fonction de répartition $Y | X_1 > médiane$ est au dessus, cela signifie qu'une forte concentration en phosphore implique une forte concentration en zinc.

- X_{11} : la fonction de répartition $Y \mid X_{11} < \text{médiane}$ est au dessus, cela signifie qu'une forte concentration en phosphore implique une faible concentration en aluminium.

IV. Conclusion

La méthode de hiérarchisation permet d'obtenir une réponse claire à la question posée : dépendances éventuelles entre produits en sortie du process.

La production de magnésium, de zinc et d'aluminium est fortement associée à la production de phosphore.

Plus précisément :

- Si l'Industriel veut une faible concentration en phosphore, il aura nécessairement une forte concentration en magnésium et en aluminium.
- S'il veut une forte concentration en phosphore, il aura nécessairement une forte concentration en zinc.
- En revanche, il n'y pas de dépendance univoque entre production de phosphore et productions de chrome, de nickel et de fer ; selon les réglages, il peut y avoir beaucoup de nickel et beaucoup de phosphore, comme peu de nickel et beaucoup de phosphore.

V. Référence

[NMP] Bernard Beuzamy : Nouvelles Méthodes Probabilistes pour l'évaluation des risques. SCM SA, ISBN 978-2-9521458-4-8. ISSN 1767-1175, broché, 272 pages. Avril 2010.