Société de Calcul Mathématique, S. A.
Outils d'aide à la décision
depuis 1995

Contrôler un process industriel :

la "hiérarchisation" des paramètres

Société de Calcul Mathématique SA

Novembre 2018

Un process industriel dépend généralement de paramètres en très grand nombre : en entrée, des teneurs, des compositions chimiques, etc. ; dans le cours du process, des pressions, des températures, etc. Tous ces paramètres sont mesurés, mais avec des précisions différentes et des périodicités différentes. Quelquefois le paramètre est mesuré en continu, d'autres fois on ne le connaît que lorsque le process est terminé et que l'on peut accéder à l'intérieur des installations (cas d'un four, par exemple).

Etant donné une variable "de sortie", c'est-à-dire caractérisant la qualité du process en sortie, l'Industriel souhaite savoir quels sont les paramètres qui influent le plus sur cette qualité. La raison est généralement que l'Industriel veut "contrôler" son process. Même si le résultat est de bonne qualité (c'est-à-dire au-dessus du seuil fixé), l'Industriel souhaite que cette qualité soit la plus constante possible.

Dans ce qui suit, nous prendrons l'exemple de la fabrication d'un acier ; la variable de sortie sera la limite d'élasticité (EL), mesurée en MPa (mégapascal), unité de pression.

Sur un ensemble de sites de fabrication, et sur tout un historique, l'Industriel observe pour EL une dispersion, donnée par exemple par l'histogramme ci-dessous :

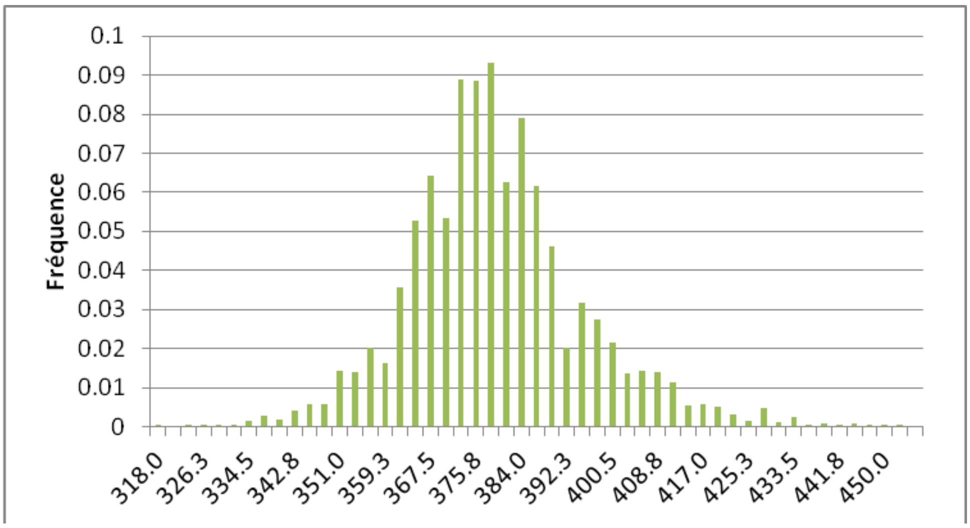

Figure 1 Histogramme de la limite d'élasticité (abscisses en MPa)

Il n'est pas content, parce qu'il considère que la dispersion des valeurs est excessive ; il aimerait par exemple que toutes ou presque soient concentrées entre 360 et 380 MPa. La question qu'il se pose est : sur quels paramètres, dans le process de fabrication, faut-il agir pour réduire cette dispersion ? La réponse est évidente : sur les paramètres les plus influents. Mais comment les déterminer ?

C'est ce que permet la méthode de hiérarchisation des paramètres, conçue par la SCM : elle va permettre un classement de tous les paramètres, par ordre décroissant d'influence. Cette méthode est de nature probabiliste.

Concrètement, nous avons la variable de sortie EL et un certain nombre de paramètres $P_1,...,P_n$ (des températures, des pressions, des compositions, ou ce que l'on voudra) ; un total de N expériences ont été faites et on dispose d'un tableau Excel à N lignes (hors en-têtes) et $n+2$ colonnes :

	EL	P_1	...	P_n
exp 1	el_1	$p_{1,1}$	...	$p_{1,n}$
...	...	...	...	
exp N	el_N	$p_{N,1}$	...	$p_{N,n}$

Les valeurs $p_{i,j}$ peuvent être numériques ou qualitatives (par exemple une couleur).

Commençons par le cas des paramètres à valeurs numériques (qui est le plus fréquent) ; prenons d'abord le premier paramètre, P_1, et déterminons la médiane, notée m_1 : c'est, par définition, le nombre tel qu'il y ait autant de valeurs de P_1 au-dessous et au-dessus. Excel détermine automatiquement la médiane.

Nous allons maintenant grandement simplifier le tableau ci-dessus : nous ne gardons que les deux colonnes EL et P_1 et, parmi les lignes, nous ne gardons que celles où $P_1 > m_1$. Ceci fait, nous construisons à nouveau l'histogramme de EL dans ce cas (cela s'appelle une "loi conditionnelle", puisqu'on a retenu la condition $P_1 > m_1$) et la fonction de répartition :

$$F_{1,1}(x) = P(EL \geq x)$$

Elle commence à 1 et finit à 0 (formellement, c'est 1- fonction de répartition usuelle, mais nous préférons aller de 1 à 0 plutôt que l'inverse).

Nous répétons les mêmes opérations pour la situation $P_1 < m_1$ obtenue en ne gardant que les lignes correspondantes et nous calculons la fonction de répartition de EL dans ce cas, notée $F_{1,2}$. Voici un exemple numérique de résultat :

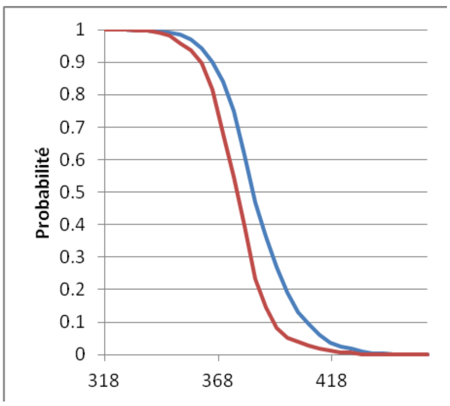

Figure 2 : graphe des deux fonctions de répartition, en abscisse des MPa

La courbe rouge représente la fonction de répartition de EL quand P_1 est supérieur à la médiane m_1 ; la courbe bleue représente la fonction de répartition EL quand P_1 est inférieur à la médiane m_1.

Comme la courbe bleue est systématiquement au-dessus de la courbe rouge, nous en déduisons que le paramètre P_1 influe sur la valeur de EL ; la valeur de EL est plus élevée quand le paramètre P_1 est petit que quand il est grand. Dans ce cas, l'influence est négative : augmenter P_1 diminuera EL.

On adopte comme mesure de cette influence l'aire A_1 séparant les deux courbes ; plus cette aire est grande, plus les courbes sont séparées et plus l'influence est importante. Si les courbes étaient très proches, voire confondues, le paramètre serait sans influence sur EL.

On procède de même pour les autres paramètres $P_2,...,P_n$: on obtient les aires $A_2,...,A_n$; le paramètre le plus influent est celui pour lequel l'aire est la plus importante, et ainsi de suite pour les autres.

Voici un exemple de résultat :

Paramètre	Aire	Sens de variation
P_3	13	+
P_1	7.3	-
P_2	7	+
P_4	6.9	-

Il peut arriver (mais c'est rare en pratique) que les courbes se croisent. Par exemple, dans le cas ci-dessus, on peut imaginer que les courbes se croisent à l'abscisse 380 MPa. En ce cas, il faudrait faire deux études distinctes : les cas où $EL < 380$ et les cas où $EL > 380$, car, selon les situations, le paramètre P_1 agirait dans un sens ou dans un autre.

La méthode possède un avantage considérable : il n'est pas utile de normaliser les paramètres, car on travaille toujours sur l'unique variable EL. Les paramètres peuvent avoir des significations totalement distinctes, être grands ou petits (une longueur, une température, une pression, etc.), on se contente, pour chacun d'eux, de le diviser en deux classes et de fabriquer la fonction de répartition de EL dans chaque cas.

La méthode s'accommode bien de la situation où un paramètre est binaire : 0 ou 1 ; on conditionne par chacune des situations. Elle s'accommode mal des situations où un paramètre est qualitatif à plus de deux valeurs (par exemple ouest, est, nord, sud). En ce cas, il faudrait tester les uns contre les autres tous les doublets possibles (ouest-est, ouest-nord, etc.) et calculer les fonctions de répartition de EL dans chaque cas.

Les résultats sont ainsi obtenus sans aucun test statistique, sans aucun ajustement, linéaire ou non : il s'agit simplement de construire une distribution de probabilité dans les différents cas ; ils sont donc robustes, puisque l'on se contente d'utiliser les données telles qu'elles viennent.

On trouvera les justificatifs théoriques dans le livre [NMP] :

[NMP] Bernard Beauzamy : Nouvelles Méthodes Probabilistes pour l'évaluation des risques. Ouvrage édité et commercialisé par la Société de Calcul Mathématique SA. ISBN 978-2-9521458-4-8. ISSN 1767-1175, avril 2010.