

# Mathématique et statistique en science de l'information et en science de la communication:

## Infométrie mathématique et infométrie statistique des revues scientifiques

Yves F. Le Coadic

CNAM – ICST

### Résumé

L'application de la mathématique et de la statistique à l'étude des phénomènes informationnels a entraîné la naissance en science de l'information d'un nouvel axe de recherche et de développement, l'infométrie. Après avoir montré l'intérêt de cette application mais aussi avoir mis en garde contre certains abus et contre certains mauvais usages, nous présentons quelques exemples d'infométrie mathématique et d'infométrie statistique appliquées aux revues scientifiques. Ils illustrent l'étendue et l'efficacité des analyses qui peuvent être faites sur une ou plusieurs variables informationnelles.

### Mots-clés

Mathématique, statistique, infométrie mathématique, infométrie statistique, bibliométrie, scientométrie, médiométrie, muséométrie, webométrie, nombre, mots, mesure, documents, cartes, LOTKA, BRADFORD.

### Matemática e estatística na ciência da informação e na ciência da comunicação: Infometria matemática e infometria estatística científicas dos periódicos

### Resumo

A aplicação da matemática e da estatística ao estudo dos fenômenos informacionais trouxe à ciência da informação um novo eixo de pesquisa e desenvolvimento - a infometria. Após demonstrar o interesse desta aplicação, mas também alertar contra certos abusos e contra determinados maus usos, são apresentados alguns exemplos de infometria matemática e de infometria estatística aplicados aos periódicos científicos. Eles ilustram a extensão e a eficácia das análises que podem ser feitas utilizando uma ou mais variáveis informacionais.

### Abstract

Infometrics, the mathematical and statistical study of information processes, is a new promising field of research in information science. Advantages but also pitfalls and misuses of mathematics and statistics in information science are presented. A selection of applications to scientific journals coming from mathematical infometrics and statistical infometrics (mono and bidimensional) illustrate the efficiency of these methods.

### Keywords

Mathematics, statistics, mathematical infometrics, statistical infometrics, bibliometrics, scientometrics, mediometrics, museometrics, webometrics, number, words, measure, documents, maps, LOTKA, BRADFORD.

## INTRODUCTION

On a cherché très tôt à décrire quantitativement les phénomènes informationnels. L'entrée de la statistique et de la mathématique en science de l'information date des années 1920. Les premières lois scientifiques sont ainsi apparues, lois au sens de relations quantitatives relativement constantes et exprimables sous la forme de fonctions mathématiques.

Ainsi la loi de Lotka (figure 1) qui explique les différences de productivité des chercheurs scientifiques dans un domaine donné, nous apprend que la relation auteurs-articles est inversement proportionnelle au carré du nombre d'articles publiés:

$$G_i = \frac{k}{U_i^2}$$

où :

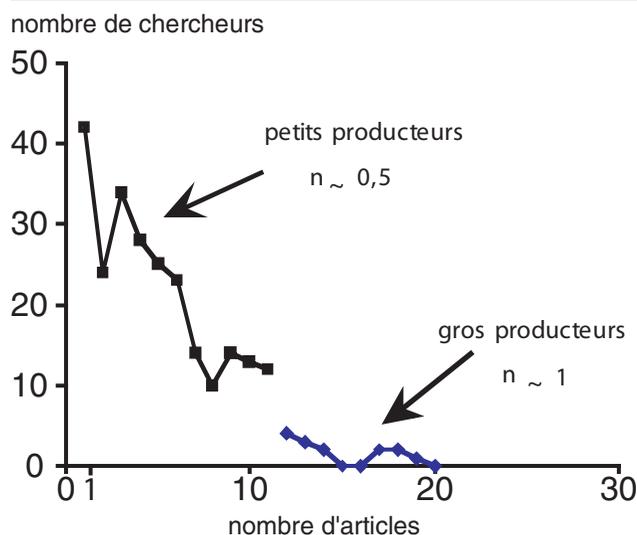
$G_i$  = nombre d'auteurs ayant publié  $U_i$  articles

$U_i$  = nombre d'articles

$k$  = constante

FIGURE 1

La productivité scientifique des chercheurs en chimie (pendant une période de trois ans)



De nombreux travaux ont confirmé depuis l'existence, dans le domaine de l'information, de régularités, de distributions et de rapports mesurables, universels, qui autorisent de ce fait la prévision et peuvent être mobilisés aussi bien dans le secteur culturel que dans le secteur marchand.

Aujourd'hui, dans le secteur de la culture, de l'éducation, de la recherche, une bonne gestion des services publics comme une bonne gestion d'une revue scientifique nécessite de plus en plus l'utilisation d'une large gamme d'outils de gestion adaptés aux contextes culturels, éducatifs, scientifiques, à la taille et au caractère du service. Ce sont des outils d'analyse des besoins d'information de la communauté desservie, des outils de pilotage et d'évaluation et des outils de mesure des performances.

À l'autre extrême, la démarche marchande particulièrement envahissante sur Internet entraîne la réalisation d'analyses statistiques élaborées de la "relation client":

- tableaux de bord de suivi de l'activité des sites ; audience par heure, date (jour, semaine, mois), nombre de sessions, de machines, de pages vues, de clics, etc.
- identification des comportements types, questionnaires, sondages réguliers faits auprès des usagers pour connaître leurs usages, le but de ces méthodes étant d'arriver à convertir un visiteur en un acheteur.

À ceux ou celles que cette mathématisation pourrait étonner, signalons que l'application de la mathématique à l'analyse des phénomènes sociaux et humains ne date pas d'hier : Georges L.L. Buffon sur l'Arithmétique morale, Marie-Jean A. Condorcet sur le problème de l'intérêt général, le dépouillement des scrutins, etc., ont ouvert la voie. Mais c'est l'économie (à la fin du XIXème siècle), la démographie, la psychologie (1910) et plus tardivement la sociologie (1950) qui ont les premières donné lieu à des recherches mathématiques. D'où l'économétrie, la psychométrie, la sociométrie. Mais cela ne s'est pas fait facilement. Ainsi, en biologie, les travaux quantitatifs ont mis quelque temps à avoir droit de citer. Pour pouvoir publier et donc légitimer des travaux quantitatifs en biologie, au début de ce siècle, Karl Pearson, un statisticien (bien connu par son coefficient que nous utiliserons plus loin) a été conduit à créer une nouvelle revue scientifique, la revue *Biometrika*.

## I - LE NOMBRE

Pour dénombrer, classer, distribuer, on utilise des outils et des objets statistiques et mathématiques, les premiers d'entre eux, mais non les moindres, étant les nombres.

L'usage des nombres est omniprésent dans la société en général et dans la société scientifique en particulier: pourcentages, taux en tout genre, enquêtes sur tout et n'importe quoi, audiences des revues, des colloques,... cotes (notoriété) des revues, cotes des femmes et des hommes scientifiques, etc.: pourcentages, taux en tout genre, enquêtes sur tout et n'importe quoi, audiences des revues, des colloques,...cotes (notoriété des revues, cotes des femmes et des hommes scientifiques, etc.

Mais ATTENTION, méfions nous du pouvoir de fascination qu'ont les nombres et les chiffres simples dans les organisations. La complexité de ces organisations étant grande, ramener les jugements à quelques chiffres simples répond à de puissantes attentes. Le langage secret de la statistique, si attractif dans une société qui vit beaucoup de faits et de chiffres, peut être employé pour faire du sensationnel, pour gonfler les résultats ou pour simplifier à l'extrême.

Face à cette religion du nombre, à cette quantofrénésie, on trouve en général des hommes et des femmes toujours aussi désarmés dans leur compréhension des nombres et leur maniement, des hommes et des femmes qui par ailleurs peuvent être fort cultivés mais aussi tout à fait capables de perdre le sens des réalités ou celui des mots quand il s'agit de nombres. Tout se jouerait-il au cours des premiers apprentissages non seulement de la numération mais aussi - et peut-être surtout - de la notation ? Si l'étiologie n'est pas simple à mettre en évidence, le résultat est facile à illustrer. Les exemples des nombreux mésusages du nombre abondent : perte du sens des réalités dans le débat public, erreurs d'interprétation dans le débat scientifique, etc.

## II . LA MESURE

Le nombre étant présent dans les systèmes d'information, il est possible de mesurer, c'est-à-dire de représenter par un nombre certaines caractéristiques d'objets, d'événements, de documents, d'informations présents dans ces systèmes, leurs quantités par exemple.

Dans le cas de la revue, document textuel, la mesure portera par exemple sur les mots, forme linguistique

minimale indépendante douée de sens, mots qui apparaissent dans le paratexte ou le texte des articles. C'est surtout le paratexte qui a été mis à contribution.

Le paratexte est le texte qui enveloppe le texte de l'article, du rapport, de la thèse, etc. Il comprend le titre de l'article, le nom de (ou des) l'auteur(e), l'adresse, les dates de réception, d'acceptation, le résumé, les remerciements, les références,...

C'est un texte non aléatoire, obligatoire, d'étendue variable. Il dévoile l'armature de l'article et assure l'adaptation réciproque de cet article et de son public.

La littérature scientifique se prête en effet à une quantification en ce sens que chaque article de revue, chaque rapport, livre, note, etc., contenant ces éléments bibliographiques paratextuels, l'agencement correct de ces éléments correctement permet de constituer un ensemble de variables textuelles, personnelles et temporelles sur lesquelles on peut effectuer des mesures et des analyses quantitatives puis qualitatives des thématiques de recherche, des relations entre les disciplines et plus généralement de l'information.

Mais on ne parvient à mesurer qu'en s'éloignant des perceptions singulières au profit de procédures réglées qu'il a bien fallu déterminer un jour, et qui conditionnent généralement notre perception du réel. C'est dire s'il est important de comprendre les fondements des techniques de mesure aujourd'hui en vigueur, les pratiques de mesure et leurs limites dans l'ordre de la science de l'information et de la science de la communication comme dans l'analyse des comportements informationnels et de comprendre sur quelles mathématiques elles s'appuient.

## 2.1 BIBLIOMÉTRIE, SCIENTOMÉTRIE, INFOMÉTRIE, MÉDIAMÉTRIE, MUSÉOMÉTRIE, WEBOMÉTRIE

Les mesures dans le secteur de l'information ont commencé par se faire sur l'objet-fétiche: le livre. On effectuait des comptages des livres et autres médias de l'écrit et on appliquait aux chiffres obtenus des techniques mathématiques simples, principalement statistiques. La bibliométrie naissait alors, avec pour objectif la mesure des activités de gestion des livres et des collections de livres et de documents.

Puis, la science et la technologie croissant, il fallait répondre aux nécessités de l'analyse du développement des activités scientifiques et techniques. Apparaissaient

alors, en Union soviétique, entre les deux guerres, la " science de la science " (nauchno y nauchno) et la scientométrie (ainsi que la technométrie), mesure des activités de recherche scientifique et technologique à travers les inputs et les outputs. Si les principales mesures des " entrées ", des ressources, consistaient à mesurer les dépenses et la main d'oeuvre, les mesures des sorties ne pouvaient être que les produits principaux de ces activités, à savoir les publications : les articles pour les activités scientifiques, c'est-à-dire l'information scientifique, les brevets constituant les produits indicateurs des activités technologiques, c'est-à-dire l'information technologique. Articles et brevets sont en effet des ressources de base pour l'évaluation des laboratoires, des pays.

On passait ensuite de l'article à l'information qu'il contient : débutait alors, dans les années 1980, l'infométrie, mesure des activités d'information, en particulier les activités d'information scientifique et technique (IST). C'est d'abord une quantification des informations bibliographiques, du paratexte, à des fins d'analyse (d'où l'usage continué du terme bibliométrie).

Puis sont arrivées depuis la médiamétrie très (trop) connue pour ses mesures d'audience, la muséométrie et enfin la webométrie.

## 2.2 LES ERREURS DE MESURE

**ATTENTION** aux erreurs de mesure. Il n'y a pas de mesure sans erreurs. Les valeurs obtenues sont souvent approximatives. Elles sont dues soit à l'imprécision des outils, soit à l'action même d'observer et sont sans commune mesure avec les erreurs rencontrées en sciences de la matière.

- 1) Erreurs provoquées par les outils bibliographiques classiques  
Ce sont les erreurs de mesure dans le paratexte (au niveau du nom de la revue, du livre, du titre, de (ou des) l'auteur, du volume, de la date d'édition, etc.) : de l'ordre de 10 à 20%.
- 2) Erreurs provoquées par les outils informatiques  
Ce sont par exemple les erreurs dans la mesure du nombre des visiteurs (au sens d'utilisateur) des sites Web : les passages fréquents des robots chargés d'actualiser les index de recherche contribuent à augmenter d'environ 10 % la consultation des sites Web.
- 3) Erreurs provoquées par les méthodes d'observation

Les sociologues connaissent depuis longtemps les problèmes liés à la collecte d'informations faite auprès des usagers pour évaluer leur satisfaction et ont développé de nombreuses techniques et savoir-faire pour collecter ces informations en minimisant les biais.

### III . LA MESURE DES REVUES

Le recours au nombre, à la mesure et à la statistique et à la mathématique peut nous apporter quelques lueurs sur le fonctionnement des revues et aider d'une certaine façon à leur promotion : améliorer leur visibilité et leur impact, élargir leur public.

Voici deux exemples d'études bibliométriques et infométriques effectuées récemment sur une revue danoise et une revue finlandaise :

- étude des autorats (nombre, productivité, localisation géographique et disciplines scientifiques des auteurs)
- étude des lectorats
- étude de la forme des articles (standardisation), de leur profil de citation
- analyse synchrone and diachrone de l'impact des revues
- mise en évidence des 10 articles phares
- calcul de la demi-vie des articles et de l'obsolescence de la littérature de la discipline
- analyse du système dévaluation (taux de rejet) , du comité de lecture.

Avant de voir plus en détail deux types de mesure des revues, évoquons les imperfections des outils disponibles actuellement.

#### 3.1 DES DIFFICULTÉS

Les outils infométriques que nous utilisons en premier sont les banques d'informations électroniques : banques bibliographiques et banques de citations. Elles sont bien développées dans les pays anglo-saxons et signalent et analysent en priorité les revues de langue anglaise, allemande, russe. La couverture des revues en langue française , portugaise, espagnole, italienne, etc.est faible. Les revues en langue française, portugaise, espagnole, italien, etc. n'ont donc qu'une très petite visibilité.

Actuellement, nous commençons à utiliser les revues électroniques et les serveurs de pré-publications. En science de l'information, on compte en effet actuellement près de 400 revues électroniques. Qui plus est, le travail est aussi possible sur le texte intégral (paratexte et texte) grâce à l'index Morganagus: 86 revues électroniques sont ainsi disponibles.

#### 3.2 UNE MESURE QUANTITATIVE : L'AUTORAT

Mots du titre et noms d'auteurs sont, avec les mots-clés, les mots les plus utilisés lors des comptages. Le rôle des mots en tant qu'opérateurs de l'auto-structuration des domaines scientifique et techniques a en effet été mis en évidence.

Les mots-clés utilisés lors de l'indexation d'un article indiquent par exemple quels sont les sujets intéressants dans un domaine de recherche donné à un moment donné. Lorsque deux mots-clés apparaissent ensemble dans une banque d'informations, les sujets qu'ils représentent sont associés. Les schémas d'association des mots-clés permet donc de mettre en évidence les tendances de la recherche ainsi que les principaux centres d'intérêt des chercheurs.

Grâce au comptage des noms d'auteurs, on mesure aussi très facilement le nombre des auteurs de la revue, leur productivité, le co-autorat et les auteurs les plus prolifiques.

Nous nous sommes livrés à une rapide comparaison de la revue française de science de la communication "RÉSEAUX" et de la revue anglaise de science de l'information "JOURNAL of INFORMATION SCIENCE":

- Si la production annuelle moyenne d'articles est sensiblement la même (environ 40 articles par an), le co-autorat est plus important pour la revue anglaise.
- Quant aux auteurs les prolifiques, ils appartiennent pour les 2/3 au comité de rédaction dans la revue française alors qu'ils ne sont qu'un tiers pour la revue anglaise.

Des revues tribales? Est-ce l'effet "tribu" qui est à l'oeuvre ici et qui semblerait caractériser d'autres revues françaises en science de la communication ? Créées par un individu, à l'image d'un groupe restreint, d'un laboratoire, elles se construisent en restant fidèle à la ligne de pensée du créateur, laissant peu de place

à une polyphonie énonciative, à l'expression des controverses qui animent d'ordinaire les communautés scientifiques.

### 3.3 UNE MESURE QUALITATIVE : LES RÉFÉRENCES (CITATIONS)<sup>1</sup>

Élément paratextuel, la citation est un concept simple. Presque tous les articles, notes, mises au point citent, font référence à des documents plus anciens qui servent de support et de précédent à l'élaboration de ce que l'auteur écrit:

En science de l'information, quand un document (A) fait mention, se réfère à un document (B), on dit que le document (B) a été cité par le document (A). Dans ce cas, référence et citation sont interchangeables. Alors qu'en littérature, la citation, c'est un extrait, un emprunt: la phrase, le passage, le texte empruntés à un auteur que l'on reproduit textuellement pour illustrer, appuyer ce que l'on veut dire. Et la référence est une évocation: la simple indication d'un document auquel on se réfère.

A travers, ces citations, un auteur identifie les relations sémantiques entre son article et les documents cités. De plus, les nouveaux articles qui citent les mêmes documents anciens ont en général des relations sémantiques entre eux. Les citations fournissent la lignée historique du savoir et sont le reflet d'une dette intellectuelle. Elles vont servir à décrire la configuration des audiences qui font usage des textes scientifiques dans un domaine donné. Mais elles obéissent aussi à d'autres motivations: citation-récompense pour remercier son supérieur, citation-politique pour que l'article soit accepté, pour susciter des renvois d'ascenseur, citation-alibi destinée à dissiper les scepticismes, citation-persuasion., auto-citation. Base d'une nouvelle méthode d'indexation, les citations ont donné naissance en 1960 aux index de citations et à la méthode d'analyse par les co-citations.

## IV - INFOMÉTRIE STATISTIQUE ET INFOMÉTRIE MATHÉMATIQUE DES REVUES SCIENTIFIQUES

### 4.1 EXEMPLES DE MESURES MATHÉMATIQUES D'UNE REVUE

#### 1) Facteur d'impact d'une revue scientifique

<sup>1</sup> Les exemples qui suivent sont tirés de l'ouvrage "Statistique et Mathématique de l'information - Infométrie, bibliométrie, médiométrie, scientométrie, muséométrie, webométrie" - LAFOUGE Th, LE COADIC Y.F., MICHEL C. - Les presses de l'ENSSIB, Villeurbanne, 2002.

Le facteur d'impact IF d'une revue scientifique mesure la fréquence avec laquelle, pendant une année donnée, l'article "moyen" d'une revue est cité dans les articles d'autres revues. C'est le rapport entre le nombre de citations reçues par une revue pendant une année et le nombre d'articles publiés pendant les deux années précédentes:

$$I_F(t) = \frac{U_{t-1}(t) + U_{t-2}(t)}{A(t-1) + A(t-2)}$$

où:

- t est l'année considérée,
- $U_{t-1}(t)$  est le nombre de citations reçues l'année t par les articles publiés dans la revue l'année précédente t-1,
- $A(t-1)$  est le nombre d'articles publiés par la revue la même année.

On trouvera dans le tableau 1 ci-dessous les valeurs du facteur d'impact des principales revues en Science de l'information. Pour la revue "Journal of the AMERICAN SOCIETY for INFORMATION SCIENCE & TECHNOLOGY", cela signifie qu'en moyenne un article qu'elle publie sera cité 1,32 fois les deux années qui vont suivre.

TABLEAU 1  
Facteur d'impact de revues de science de l'information (source JCR®) (Année 1999)

J AM SOC INFORM SCI	1,325
ONLINE CDROM REVIEW	0,293
SCIENTOMETRICS	0,931
J INFORM SCI	0,659
DATA KNOWLEDGE	0,293
INFO PROCESS MANAG	0,773
SOC STUD SCI	0,296

2) Citations reçues par une revue scientifique et usages de cette revue

Existe-t-il une corrélation entre le nombre de citations qu'une revue scientifique a reçue en une année et son usage, mesuré par le nombre d'articles demandés durant une année chez un fournisseur de documents? La question est de savoir si une revue très citée est plus souvent photocopiée qu'une revue peu citée. Pour les 10 revues les plus demandées, les résultats sont les suivants:

TABLEAU 2

Citations reçues et articles photocopiés des 10 revues les plus demandées (Les citations reçues sont tirées du Journal Citation Report publié par l'Institute for Scientific Information (ISI) à Philadelphie)

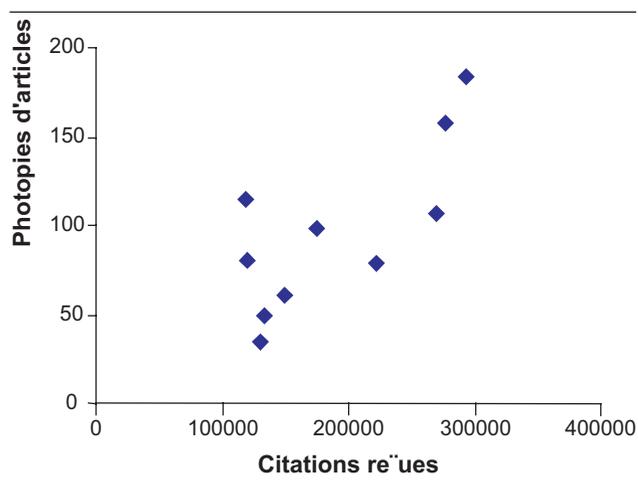
Titres des revues	Citations reçues	Articles photocopiés
Journal of biological chemistry	293024	184
Proceedings of the national academy of science	276659	158
Nature	270077	107
Science	221696	79
Journal of the american chemical society	174540	99
Cell	149477	60
Physical review letter	132797	50
Physical review B -Condensed matter	130789	35
Journal of chemical physics	120281	81
New england journal of medicine	118106	115

Les valeurs des deux variables, X citation et Y photocopie, sont portées sur un graphique de la même façon que l'on porte un point sur une carte (latitude, longitude): l'une est mesurée sur l'axe horizontal X (nombre de citations), l'autre sur l'axe vertical Y (nombre de photocopies).

Le point résultant se trouve à l'intersection des deux perpendiculaires passant par ces points de mesure. On obtient un nuage de points qui nous permet de visualiser la liaison entre ces deux variables (Voir graphique 1).

GRAPHIQUE 1

Relation entre Usage et Citation des revues



Pour déterminer s'il existe une corrélation entre les deux variables, on fait appel au coefficient de corrélation noté  $r$  qui exprime la force de la relation (l'attraction) qui existe entre elles. C'est un nombre calculé à partir des deux variables X et Y:

$$r_p = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

C'est ici le coefficient de corrélation de PEARSON noté  $r_p$ . Égal ici à 0,75, il indique qu'il existe une corrélation linéaire positive importante entre le nombre de demandes de photocopies d'articles d'une revue et le nombre de citations qu'elle reçoit.

#### 4.2 EXEMPLES DE MESURES MATHÉMATIQUES D'UNE REVUE

##### 1) Le cœur de littérature et la Loi de BRADFORD

En 1934, un documentaliste britannique Samuel. C. BRADFORD énonce une loi (qui porte depuis son nom) qui permet au gestionnaire d'un service de documentation de gérer sa collection de périodiques scientifiques. Elle nous permet aussi de fournir à un doctorant les revues les plus importantes qui traitent de son sujet.

Si, dans un domaine scientifique donné, sur une thématique donnée, les périodiques scientifiques sont rangés par ordre décroissant d'articles publiés, il existe un nombre  $q$  (supérieur à 1) et un nombre  $r$  de périodiques tels que si l'on regroupe les revues en considérant les  $r$  premiers, puis les  $rq$  suivants, puis les  $rq^2$  suivants ... (soit une progression géométrique), alors on observe que chaque groupe de revues contient le même nombre d'articles. Le nombre de revues par groupe est une suite géométrique:

$$r, rq, rq^2 \dots rq^i \dots$$

Supposons que la littérature cœur d'un domaine soit composée de 20 revues et que le facteur de BRADFORD  $q$  soit égal à 2. Ces 20 revues contiennent par exemple 100 articles sur le sujet. Si l'on veut consulter trois fois plus d'articles, soit 300 articles, il est alors nécessaire de dépouiller 140 revues au lieu de 20, c'est-à-dire sept fois plus de périodiques c'est-à-dire 140 au lieu de 20:

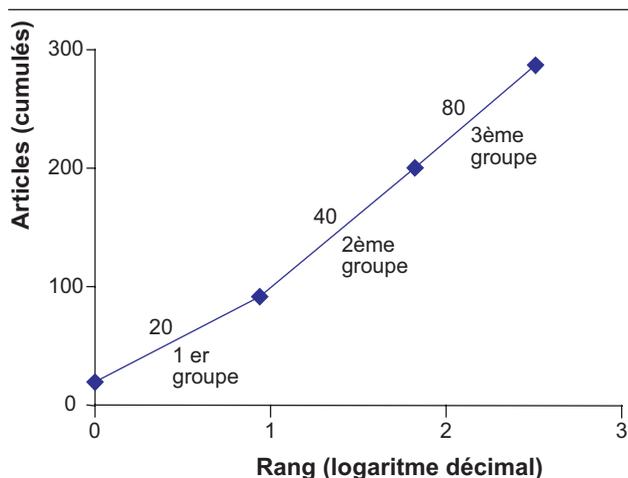
$$20 + 20.2 + 20.2.2 = 140$$

On peut mettre en évidence trois groupes de revues:

- un premier groupe contient les revues qui ont publié plus de 4 articles par an,
- un deuxième groupe contient les revues qui ont publié entre 2 et 4 articles par an,
- et un troisième groupe contient les revues qui ont publié au maximum un article par an.

Si l'on trace la courbe de variation du nombre d'articles en fonction du rang de la revue on constate que c'est une droite sur sa plus grande longueur. On peut sur cette courbe mettre en évidence les 3 groupes de revu

GRAPHIQUE 2  
 Les 3 groupes de BRADFORD



## 2) Obsolescence de l'information

La fonction exponentielle est parfois appelée " fonction de croissance naturelle " car de nombreux processus naturels, comme la croissance d'une forêt ou d'une population, varient de façon exponentielle. La croissance exponentielle de la quantité d'articles est bien connue.

Corollaire à cette croissance rapide du nombre de publications, il existe une obsolescence également rapide du stock d'information disponibles. Ce qui veut dire que si les références à la littérature passée sont distribuées de façon aléatoire, sans rapport avec la date de publication, une majorité d'entre elles renvoie à des travaux récents, puisqu'il y a plus d'articles disponibles pouvant être cités:

$$C(t) = C(0)e^{-at}$$

où a est un nombre positif supérieur à 1.

Le calcul de la demi-vie des littératures scientifiques fournissent des éléments permettant d'éclairer ce type d'interrogation. La demi vie d'une littérature est le temps pendant lequel la moitié de la littérature active a été publiée. Les études d'obsolescence des différentes littératures ont montré des variations importante de cette caractéristique : 4,6 années en physique, 7,2 années en psychologie, 10,5 années en mathématiques. De façon identique, connaissant le nombre total de citations reçues par une revue, la demi-vie de cette revue mesure le nombre d'années pendant lesquelles elle a reçu 50% de ces citations. A titre d'exemple, voici les valeurs de ces demi-vies pour quelques revues de Science de l'information:

TABLEAU 3  
 Demi vie des revues en Science de l'information

Revues	Demi-vie (années)
J AM SOC INFORM SCI	6,8
SOC STUD SCI	9,6
SCIENTOMETRICS	5,1
INFORM PROCESS MANAG	6,8
J INFORM SCI	6,2

(Source JCR) (Année 1999)

## V LA MESURE DES REVUES ÉLECTRONIQUES

Comme on le sait, Internet est devenu une plateforme éditoriale globale et de nombreuses revues électroniques sont maintenant disponibles, les chercheurs les considérant de plus en plus comme des médias légitimes de publication. En informatique aujourd'hui, les articles récents les plus cités sont plus souvent en ligne qu'imprimés. La science de l'information ne fait pas exception comme nous l'avons vu ci-dessus.

On trouvera ci-dessous les résultats d'une des toutes premières études infométriques d'un échantillon de 28 revues électroniques<sup>2</sup>.

### LA PRODUCTION

Au 1er août 2001, 1,120 articles en science de l'information avaient été publiés dans ces 28 revues. Le nombre d'articles publiés était de 26 en 1995 et atteignait 204 en 1999.

La plus grande majorité des auteurs les plus prolifiques provenaient des USA et de Grande-Bretagne. L'auteur le plus prolifique a publié 11 articles pendant ces 6 ans.

<sup>2</sup> Hawkins, Donald T. (2001) "Bibliométrie de revues électroniques en science de l'information" *Information Research*, 7(1) [Available at <http://InformationR.net/ir/7-1/paper120.html>]

Seulement 26 articles avaient comme auteurs des personnes de pays différents montrant ainsi que la technologie électronique n'influencait guère encore la collaboration internationale.

Le lectorat n'a pas été étudié .

#### LE CŒUR DE LITTÉRATURE

Le tracé de la loi de Bradford met en évidence le fait que la littérature cœur est encore mal définie. On peut penser que 6 revues déjà bien établies, fondées en 1996 comme Ariadne, D-Lib Magazine, Journal of Electronic Publishing, Information Research, First Monday, and Issues in Science & Technology Librarianship et publiées de façon régulière mensuellement et trimestriellement en sont l'amorce .

#### MULTI-AUTORAT

Étant donné les facilités de communication qu'offre courriels et collecticiels, on peut espérer voir se développer de nombreuses et importantes collaborations qui vont se traduire par plus d'articles co-auteurs. Pour l'instant, le faible échantillon étudié sur une durée encore trop courte ne permet pas de tirer de telles conclusions.

#### PAYS DES AUTEURS

USA et Grande-Bretagne ont produit à eux deux 81,6% des articles; la France 0,4% (soit 5 articles). On constate avec intérêt qu'un grand nombre de pays sont représentés. La technologie électronique rend plus facile la communication des auteurs avec les éditeurs scientifiques et commerciaux; et ce qu'elle que soit leur localisation.

#### INSTITUTIONS DES AUTEURS

12% (soit 68) des organismes sont responsables de 46% de la production (soit 672 articles). Les 2/3 des auteurs proviennent d'institutions académiques.

#### THÈMES

Les thèmes les plus développés dans ces revues électroniques portaient sur l'information électronique (20,7%), les bibliothèques électroniques (16,9%), l'édition électronique (12,9%). Les études des besoins d'information et des usagers ne comptaient que pour 5,5%.

#### FORMAT DES ARTICLES

Le mimétisme est très fort avec les articles papier. On retrouve des brèves (nouvelles, comptes-rendus de conférences) des textes courts décrivant des recherches mais aussi des services et des produits; mais aussi des articles de fond, rapports de recherche et synthèses.

#### SIGNALEMENT DANS LES BANQUES D'INFORMATIONS D'INFORMATIONS

7 banques signalent et analysent actuellement ces revues électroniques.

#### CONCLUSION

Après avoir évoqué le rôle du nombre et de la mesure lors de l'évaluation des revues scientifiques, nous avons présenté quelques indicateurs de nature statistique et mathématique particulièrement bien adaptés à cette fin. Leur détermination peut servir à améliorer la visibilité et l'impact d'une revue et donc à élargir son public. L'apparition des revues sur support électronique, qui se révèlent être par ailleurs de très viables médias de publication, rend plus facile ces déterminations. D'autres améliorations sont donc à attendre du passage progressif de toutes les revues scientifiques en version électronique.

“Because Electronic is beautiful !”.