Fraud detection and Benford's law: some linked risks

Adrien Bonache and Karen Moris and Jonathan Maurice

MRM-CREGOR-COST, LEG-FARGO, MRM-ERFI-FCCS

1. April 2010

Détection de fraudes et loi de Benford: quelques risques associés

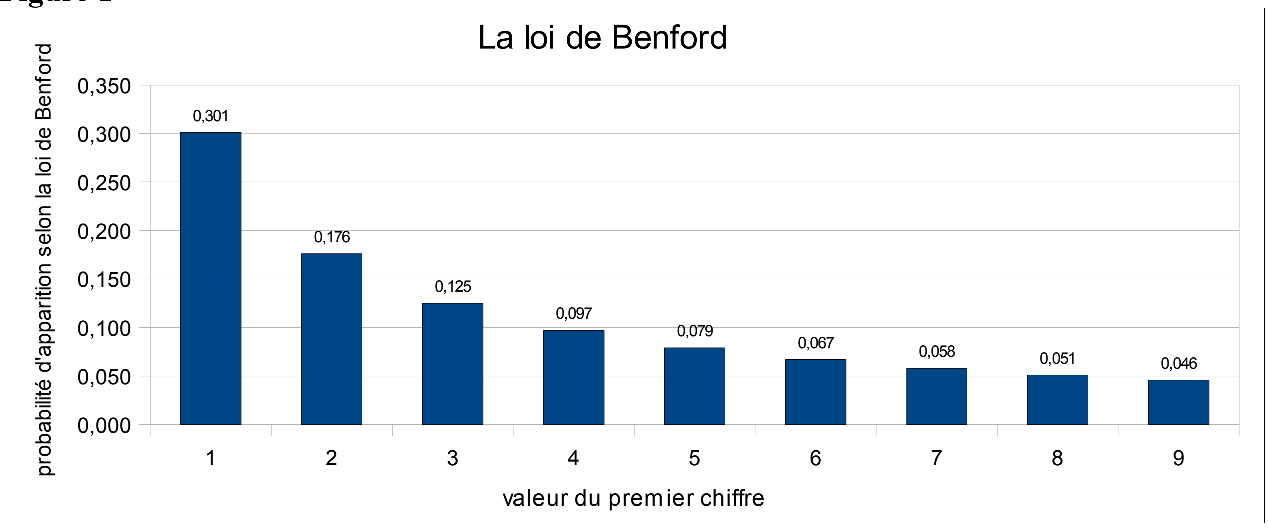
Adrien B. Bonache[*]
Karen Moris
Jonathan Maurice

Résumé
Après avoir brièvement rappelé l'attrait *a priori* de la loi de Benford pour repérer les fraudes, les limites d'utilisation connues sont présentées. De plus, un exemple atteste que cet outil statistique ne permet pas de détecter des fraudes comptables dans les ventes de biens à la mode réalisées par une entreprise.
Abstract
After reviewing the *ex ante* appeal of Benford's law to detect fraud, the known limitations are introduced. Then, an example states that this statistical tool didn't permit to make out the presence of accounting fraud for fashion good sales.

Introduction

Pour discerner rapidement quels sont les postes comptables que les entreprises manipulent et ainsi découvrir les entreprises qui fraudent, certains commissaires aux comptes et auditeurs fiscaux pourraient penser que la loi de Benford serait d'une grande aide : c'est une « loi » statistique (figure1) qui offrirait la possibilité de détecter les fraudes comptables commises par les entreprises. Mais cet outil possède des limites qu'il convient de connaître avant de l'utiliser à cet escient.

Figure 1



À la fin du XIXème siècle, Newcomb constate l'usure inégale des tables de logarithme: Le chiffre 1 est plus utilisé que le chiffre 2, lui-même plus utilisé que le chiffre 3... Il établit de cette observation une loi statistique. Cette loi a été « remise au jour » par Benford à la fin des années 1930 et prend alors son nom. La loi de Benford rend compte de la fréquence d'apparition du premier chiffre d'un nombre dans une série de nombres. Elle est une distribution telle que la probabilité d'apparition d'une valeur « d » en première position d'un nombre est ln(1+1/d)/ln(10). La figure 1 en donne une représentation graphique.

En 1972, Varian a souligné l'utilité potentielle de cette loi pour détecter la présence d'éventuelles fraudes dans des bases de données. L'idée d'utiliser cette loi en comptabilité pour détecter un risque de fraude comptable est apparue à la fin des années 1980 dans un rapport de recherche d'un chercheur néo-zélandais, Charles Carslaw. Par la suite dans de nombreux pays, d'autres chercheurs

[*] Auteur pour correspondance: bonache@rip.ens-cachan.fr; +33 (0)6.28.94.62.40. Ceci est une version preprint de l'article publié dans la Revue Française de Comptabilité d'avril 2010. La version postprint est disponible sur le site : **http://proquest.umi.com/pqdweb?TS=1061376276&RQT=306&lang=fr**

ont confirmé la présence de fraudes fiscales et comptables détectables par l'intermédiaire de cette « loi » statistique (et non détectées par les agents du fisc qui n'utilisaient pas alors cette analyse statistique).
La présence de fraudes serait détectable par une analyse statistique en utilisant la loi de Benford. Il s'agit de comparer la distribution des chiffres dans les postes comptables d'une entreprise avec la loi de Benford. S'il est constaté une différence significative entre les distributions, alors il y a lieu de pousser l'investigation pour rechercher des preuves concrètes de fraudes dans certains comptes de l'entreprise. Autrement dit, lorsqu'il y a inadéquation à la loi de Benford, statistiquement il peut y avoir des fraudes. Mais cela ne suffisant pas pour condamner une entreprise, des investigations complémentaires sont menées pour chercher des preuves concrètes.

Cette loi a déjà fait l'objet d'une étude dans la *Revue Française de Comptabilité*. Labouze et Labouze ont analysé l'utilité de cette loi pour assister un auditeur à la recherche de fraudes comptables. Eu égard à l'objectif de leur article, les limites d'application de cette loi statistique ne sont pas étudiées. L'objet de notre article est de rappeler les limites connues et de les compléter.

Dans le *Journal of Forensic Accounting*, un article publié en 2004 met en avant les clés de l'utilisation efficace de la loi de Benford. Premièrement, il faut connaître les types de postes comptables ou les types de données devant suivre la loi de Benford. Les auteurs proposent la classification suivante donnant des pistes de réponses à cette première condition (table 1).

Table 1
Utilité de la loi de Benford pour détecter des fraudes

Quand la loi de Benford semble utile	**Exemples**
Ensemble de nombres qui résulte de combinaisons mathématiques d'autres nombres- résultat venant notamment de deux distributions de données économiques et/ou comptables	Chiffres d'affaires (ventes*prix) Achats de matières premières et marchandises (achats*prix)
Données sur le volume des transactions	Décaissements, ventes, dépenses
Ensemble de données « larges » -plus on a de données, plus l'analyse est pertinente	Transactions durant un exercice comptable
Comptes ou montants comptables, qui semblent se conformer à la loi de Benford - quand la moyenne de la série est supérieure à la médiane et la skewness (coefficient d'asymétrie) est positive	La plupart des séries de nombres comptables
Quand la loi de Benford semble inutile	**Exemples**
Séries de données constituées de nombres attribués arbitrairement (2)	Numéros de chèques, de factures, codes postaux
Nombres pouvant influencer la « pensée humaine » (2)	Prix fixés à des seuils psychologiques (1,99$), retraits aux guichets automatiques bancaires (20, 40…euros)
Comptes où transitent des gros montants à destination d'un nombre précis d'entreprises (2)	Un compte spécifiquement créé pour enregistrer une opération de refinancement de 100$
Comptes avec une borne minimale et/ou une borne maximale (2)	Ensemble d'actifs qui doivent dépasser un seuil pour être enregistré
Transactions qui ne donnent pas lieu à un enregistrement comptable (1)	Vols, pots-de-vin, « contrats à l'amiable »

Source: Durtschi, Hillison et Pacini (2004, p..24)

Deuxièmement, les auteurs soulignent l'importance du niveau de signification des résultats. En effet, en utilisant la seule statistique de test du chi-deux d'adéquation à la loi de Benford, il est possible que l'on trouve une preuve statistique de fraude dans une entreprise alors qu'il n'y a pas de fraude effective (« faux négatif » ou erreur de première espèce). L'inverse peut aussi se produire: ne

pas déceler de fraude bien qu'une fraude réelle existe (« faux positif » ou erreur de seconde espèce). Pour éviter ces erreurs, il est conseillé de ne pas utiliser uniquement le test du chi-deux d'adéquation mais aussi un test binomial pour chaque chiffre.
Troisièmement, l'article fait apparaître deux types de fraudes indécelables avec le seul outil statistique d'adéquation à la loi de Benford (indexés (1) et (2) dans la table 1):
(1) celles concernant des « transactions qui ne donnent pas lieu à un enregistrement comptable »,
(2) celles touchant des données qui ne suivent pas la loi de Benford pour d'autres raisons que l'existence d'une manipulation des données pouvant traduire l'existence d'une fraude.
Dans ces deux cas, les auteurs conseillent l'usage de méthodes non statistiques: « analyse détaillée de l'actif, inspection directe de l'entreprise et des documents comptables, connaissance détaillée de la culture de l'entreprise et de ses performances relativement aux autres entreprises du secteur, et une analyse approfondie des explications contenues dans les rapports d'activité concernant des écarts éventuels».

Dans cet article, nous supposons qu'il peut exister des séries de données comptables, dans certaines entreprises, qui peuvent ne pas suivre la loi de Benford, sans pour autant signifier qu'il y ait eu une fraude. Cela peut être le cas lorsque ces montants sont issus d'un type de marché ayant des caractéristiques spécifiques.
Nous avons retenu et étudié un cas pouvant correspondre à cette « hypothèse »: les ventes de biens à la mode peuvent ne pas être en adéquation à la loi de Benford, sans pour autant signifier que ces ventes aient fait l'objet de manipulations. Ainsi contrairement à ce qui apparaît sur la table 1, le poste de ventes relatif à ce type de biens peut ne pas suivre la loi de Benford, sans qu'il y ait de fraudes concrètes. Ce résultat pourrait s'expliquer par l'existence d'une dynamique particulière entre consommateurs, comme le montrent certains travaux récents en mathématiques (Granovetter et Soong, 1986; Tolle, Budzien et Laviolette, 2000). D'après ces travaux, ce type de dynamique peut ne pas suivre la loi de Benford.

L'objectif de notre papier est de montrer qu'il existe des postes comptables pouvant ne pas suivre la loi de Benford en l'absence de fraudes: les postes de ventes de biens à la mode dans des entreprises en vendant. Autrement dit, existe-t-il des ventes de produits à la mode ne suivant pas la loi de Benford?

Méthode d'investigation

Nous décrivons dans cette partie la base de données collectées et la procédure qui nous a permis de révéler un type de données comptables pouvant ne pas suivre la loi de Benford en l'absence de manipulations comptables.

Nous avons analysé des ventes de consoles de jeux vidéo, qui peuvent être théoriquement intéressantes, car il s'agit de biens innovants pour lesquels il existe des phénomènes de mode. Suivant l'article de Mark Granovetter et Roland Soong, il pourrait y avoir des dynamiques non-linéaires dans les ventes de ce type de produits. Ces dynamiques peuvent faire naître des comportements chaotiques. Or, des mathématiciens ont dernièrement montré que ces dynamiques « chaotiques » peuvent ne pas être en adéquation avec la loi de Benford[1].
Une base de ventes hebdomadaires de consoles est disponible sur longue période, celle du site internet www.vgchartz.com. Elle a été utilisée dans cette étude. Suivant Varian, l'adéquation de cette base de données à la loi de Benford (statistique du chi-deux$\simeq$1,53, *p-value*$\simeq$0,99) montre une absence de biais ou de fraudes dans celle-ci.

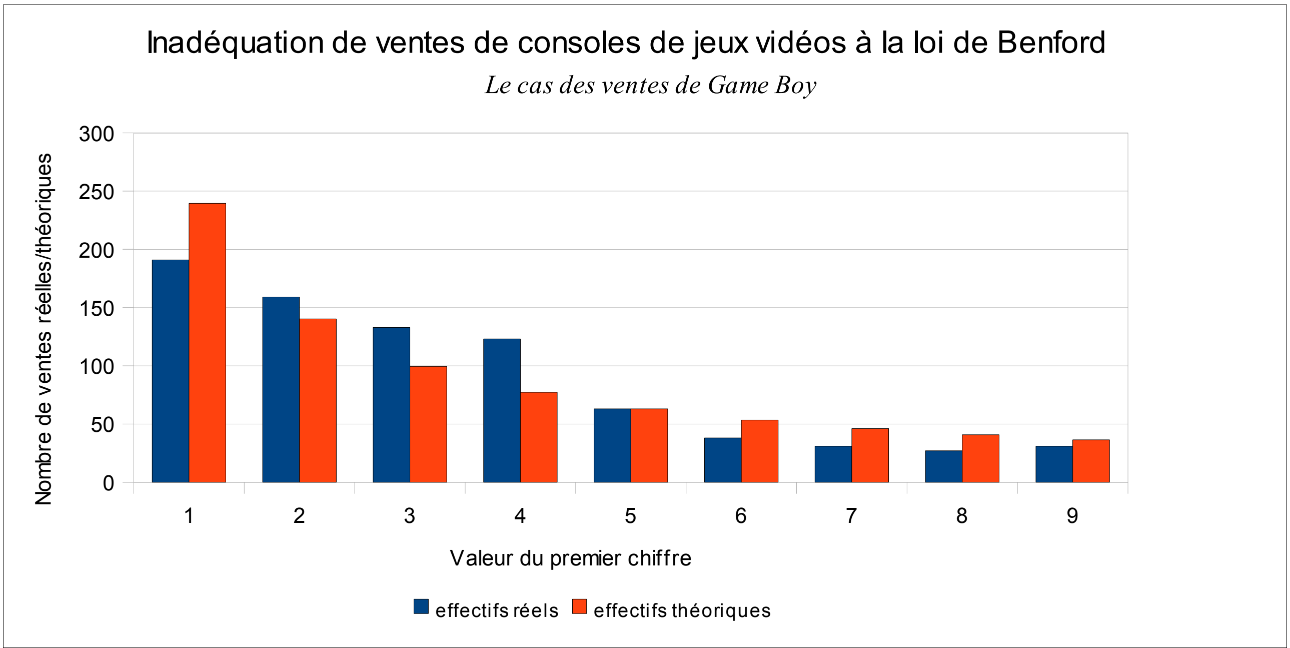
Pour l'analyse, nous avons séparé les chiffres des quantités vendues hebdomadaires avec la

[1] Nous invitons le lecteur intéressé par la théorie du chaos et ce résultat en mathématique à se référer à l'ouvrage de Dominique Guégan et à l'article de recherche de Charles Tolle, Joanne Budzien et Randall Laviolette, dont les références figurent en fin d'article.

procédure suivante. Nous avons remplacé le signe « 1 » par « 1, » et fait de même pour tous les autres chiffres. Par exemple, s'il avait été vendu 195 463 Game Boy, nous avons obtenu, par cette démarche, la ligne suivante : « 1,9,5,4,6,3, ».
Ensuite, la série des ventes pour chaque console de jeu a été copiée et collée sur un tableur, avec la fonction collage spécial, en indiquant de prendre en compte la virgule comme séparateur. Ainsi, les premiers chiffres de toutes les ventes d'un bien, pour chaque semaine, apparaissaient sur la même colonne. Puis, nous avons compté le nombre de « 1 » dans la colonne où étaient inscrits les premiers chiffres; puis le nombre de 2, de 3 jusqu'à 9, toujours dans cette première colonne.
Enfin, suivant la méthode prescrite dans l'article précédemment cité de Xavier et Robert Labouze, des tests d'adéquation à la loi de Benford ont été réalisés.
Par exemple, dans notre étude, les résultats suivant ont été obtenus pour les ventes hebdomadaires de Game Boy du 22 avril 1989 au 18 juillet 2004.

Résultats d'un test d'adéquation à la loi de Benford
La figure 2 ci-dessous présente les différences entre la répartition des premiers chiffres pour les ventes de Game Boy (en bleu) et la distribution des ventes conformément à la loi de Benford (en orange).

Figure 2



Sur la table 2 ci-dessous, ces deux distributions correspondent respectivement à e et e*. e* a été calculé en multipliant le nombre de semaines de vente de la Game Boy (soit la somme des e égale à 796) par la probabilité théorique d'apparition d'un chiffre en première position selon la loi de Benford (« loi théorique »).

Table 2
Détail d'un test d'adéquation à la loi de Benford [a]

Premier chiffre	Effectif e	Loi théorique	Effectif théorique e*	(e-e*)²/e*	Signe du biais	P-value d'un test binomial
1	191	0,3010	239,62	9,87	-	0,000020849 ***
2	159	0,1761	140,17	2,53	+	0,008007619 ***
3	133	0,1249	99,45	11,32	+	0,000100388 ***
4	123	0,0969	77,14	27,26	+	0,000000086 ***
5	63	0,0792	63,03	0	-	0,052307382 *
6	38	0,0669	53,29	4,39	-	0,004934505 ***
7	31	0,0580	46,16	4,98	-	0,003736427 ***
8	27	0,0512	40,72	4,62	-	0,004982558 ***
9	31	0,0458	36,42	0,81	-	0,046725773 **
total	796	1	796	65,77 ***		

[a] « + » («–») signifie que le nombre de ventes réelles dépasse (resp. est inférieur à) l'effectif théorique-selon la loi de Benford- pour une valeur donnée du premier chiffre. « e » et « e* » représentent respectivement le nombre de ventes et l'effectif théorique associé à une certaine valeur pour le premier chiffre. *,**,*** indiquent que le degré de signification du résultat est inférieur à 10%, 5% et 1%.

En calculant une statistique du chi-deux, on trouve une différence significative entre les deux distributions (Table 2). En effet, la somme des ratios (e-e*)²/e* associé à chaque chiffre -statistique du chi-deux- est supérieure à la borne du test pour un risque de se tromper de 1%, 65.77 contre 20.09. La différence entre les deux distributions est donc significative.

La table 2 permet de retrouver des informations représentées sur la figure 2 concernant l'écart d'effectifs par chiffre (signe du biais) et donne une information complémentaire quant à leur degré de signification (*p-value* d'un test binomial). Notamment, la figure 2 fait apparaître un écart faible pour les chiffres 5 et 9. La table confirme cela, car ces écarts ne sont pas significatifs si l'on n'accepte pas un risque de se tromper supérieur à 1% (*p-value*>0,01). Inversement pour le chiffre 4, le graphique semble indiquer un écart très important et la table montre que le risque de se tromper est de 0,000000086 environ en déclarant que les effectifs théorique et empirique diffèrent. Cette différence est donc significative (*p-value*<0,01).

Somme toute, les ventes de consoles de jeux vidéo ne suivent pas la loi de Benford. Cela devrait amener à conclure statistiquement qu'il existe de possibles fraudes. Pourtant un test préalable concernant l'ensemble de la base de données montre une adéquation (*p-value*≃0,99) à la loi de Benford et donc statistiquement l'absence de fraudes dans celle-ci.[2]

Apports et limites de nos résultats pour les praticiens

L'apport majeur de cette étude est de synthétiser en français les limites connues de l'utilisation de la loi de Benford pour déceler des fraudes comptables (table 1). En effet, il semble indispensable de définir en creux les domaines d'application d'un test statistique avant d'en faire usage.
De plus, il semblerait que les ventes suivant des dynamiques chaotiques peuvent ne pas suivre la loi de Benford. Or, un commissaire aux comptes ou un auditeur peut être amené à utiliser la loi de Benford pour vérifier s'il y a d'éventuelles fraudes dans les ventes d'une entreprise. D'après des travaux de recherche récents, les secteurs pour lesquels des ventes peuvent être chaotiques sont les suivants:
- celui des produits innovants,
- celui de la vente d'articles faisant l'objet de phénomènes de mode et
- de la vente de biens pouvant entraîner une addiction des consommateurs (Feichtinger et al., 1995).
D'autres investigations sont à mener pour confirmer notre résultat dans le secteur de la vente de consoles de jeux vidéo et pour voir si des résultats similaires sont observables dans d'autres secteurs.

---

[2] Rappelons que dans la section « méthode d'investigation », il a été dit que la base de données du site internet www.vgchartz.com a été utilisée et qu'un test d'adéquation à la loi de Benford a montré une absence de fraude significative dans l'ensemble de cette base de données de ventes: statistique du chi-deux≃1,53 (*p-value*≃0,99).

Cependant, notre étude n'est pas exempte de critiques. La plus importante est de n'avoir testé la non-adéquation à la loi de Benford que sur des ventes en volume et non sur les ventes en valeur. Il se peut que sur les chiffres d'affaires hebdomadaires par produit cette non-adéquation à la loi de Benford disparaisse. Mais il se peut aussi que les ventes hebdomadaires de consoles de jeux vidéo en valeur ne soient pas significativement en adéquation avec cette loi. D'autres investigations sont à mener pour éclaircir ce point.

Somme toute, dans l'attente d'investigations supplémentaires, s'ils utilisent la loi de Benford, les commissaires aux comptes et les auditeurs doivent le faire avec précaution. Premièrement, il existe des fraudes indécelables avec cette analyse statistique (table 1). Deuxièmement, dans les secteurs de la vente de produits rendant dépendants et particulièrement des biens innovants et/ou à la mode, il se peut que les ventes ne suivent pas la loi de Benford sans être frauduleuses (figure 2 et table 2).

Références

Carslaw C., 1988, Anomalies in income numbers: Evidence of goal-oriented behavior, *The Accounting Review* 63 (2).

Durtschi C., Hillison W. et Pacini C., 2004, The effective use of Benford's law to assist in detecting fraud in accounting data, *Journal of Forensic Accounting* 5 (1).

Feichtinger G., Prskawetz A., Herold W. et Zinner P. (1995), « Habit formation with threshold adjustment: addiction may imply complex dynamics », *Journal of Evolutionary Economics* 5(2).

Granovetter M. et Soong R., 1986, Threshold models of interpersonal effects in consumer demand, *Journal of Economic Behavior and Organization* 7 (1).

Guégan D., 2003, *Le chaos en finance: Approche statistique (Statistique mathématique et probabilité),* Economica.

Hill T.P., 1995, A statistical derivation of the significant-digit law, *Statistical Science* 10.

Labouze X. et Labouze R., 2000, La Loi de Benford: la détection des fraudes comptables, *Revue française de comptabilité* 321.

Nakayama S. et Nakamura Y., 2004, A fashion model with social interaction, *Physica A: Statistical Mechanics and its Applications* 337 (3).

Tolle C., Budzien J. et Laviolette R., 2000, Do dynamical systems follow Benford's law?, *Chaos: An Interdisciplinary Journal of Nonlinear Science* 10 (2).

Varian Hal, 1972, Benford's law, *The american statistician* 26 (3).