



**HAL**  
open science

## Effet long tail ou effet podium : une analyse empirique des ventes de produits culturels en France

Pierre-Jean Benghozi

► **To cite this version:**

Pierre-Jean Benghozi. Effet long tail ou effet podium : une analyse empirique des ventes de produits culturels en France. 2008. hal-00407205

**HAL Id: hal-00407205**

**<https://hal.science/hal-00407205v1>**

Submitted on 24 Jul 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**CNRS**

CENTRE NATIONAL  
DE LA RECHERCHE  
SCIENTIFIQUE

PREG

# ***Effet Long Tail ou effet podium : une analyse empirique des ventes de produits culturels en France***

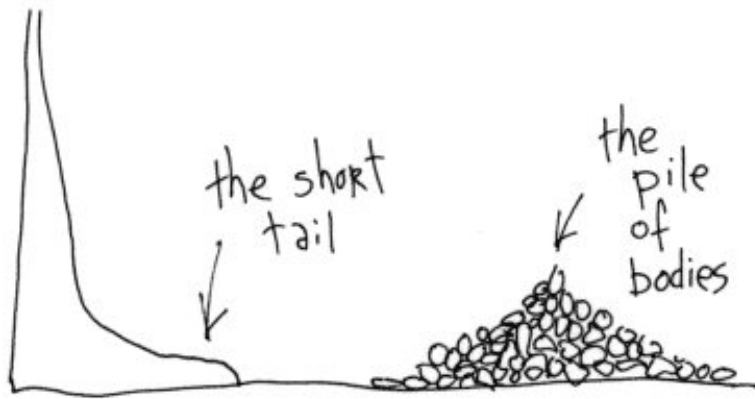
*Pierre-Jean Benghozi*

Mars 2008

• • • • • • • • •

Le document qui suit remet en forme le travail réalisé par DIDION Christian, DIMCA Georges, GRANDPRE Julien, GROTTO Lionel, KAFO Ronald, LUCIANI Vincent, WOLBER Olivier dans le cadre d'un projet scientifique collectif d'élèves de l'Ecole polytechnique. Il s'appuie aussi sur le mémoire effectué par Sami BIASONI, dans le cadre du Master Gestion et Dynamiques des Organisations de l'Université Paris Nanterre.

Ces travaux ont bénéficié du soutien du Département des Etudes, de la Prospective et des Statistiques du Ministère de la Culture et de la Communication



(c)gapingvoid.com

# Table des matières

## **PARTIE 1 – INTRODUCTION**

- 1. Présentation et discussion du phénomène « Long Tail »**
- 2. Etats des lieux de la situation américaine**
  - 2.1 Présentation
  - 2.2 Méthodologie
  - 2.3 Principaux résultats
  - 2.4 Analyse de l'étude et perspectives

## **PARTIE 2 - L'ANALYSE DE LA SITUATION FRANÇAISE**

- 1. Eléments de méthode**
  - 1.1. Présentation des données étudiées
  - 1.2. La construction de séries temporalisées
  - 1.3. Les limites de la base GfK
- 2. Le cas de la musique**
  - 2.1. Le contexte de la musique en France en 2005
  - 2.2. Premiers constats
  - 2.3. Un premier résultat : la saisonnalité
    - 2.3.1 *Résultats en magasin*
    - 2.3.2 *Résultats des ventes par Internet*
    - 2.3.3 *Conclusion*
  - 2.4. Caractérisation statistique de la Tail
  - 2.5. L'atténuation de l'effet superstar sur la période 2002 – 2005
  - 2.6. Synthèse
- 3. Le cas de la vidéo**
  - 3.1. Le contexte de la vidéo en France en 2005
  - 3.2. Saisonnalité du marché de la vidéo
  - 3.3. Long Tail et réduction de l'effet superstar
    - 3.3.1 *L'évolution des ventes de vidéo sur Internet et en magasin*
    - 3.3.2 *Concentration et déconcentration du marché*

## **PARTIE 3 - CONCLUSION**

- 1. Rappel des principaux résultats**
- 2. Une illusion de Long Tail ?**
- 3. Les limites méthodologiques**
- 4. Modèles d'affaires, étalement des prix et business models**
- 5. Recommandations pour une prolongation au téléchargement**
  - 5.1. Fiche technique

## **REFERENCES BIBLIOGRAPHIQUES**

### **ANNEXES**

**A.1. Distribution par décile du marché de la musique au cours de l'année 2005**

**A.2. Caractérisation statistique comparée des ventes de musique : chiffre d'affaires et volumes, sur et hors internet, semaines 9, 13 et 51**

**A.3. Représentation de la tête de distribution : Top 10 des ventes de CD sur et hors Internet**

**A.4. Représentation de la queue de distribution : ventes de CD sur et hors Internet**

**A.5. Données GfK – Le logiciel de consultation fourni en ligne**

**A.6. Rappels statistiques et techniques**

## **PARTIE 1 - INTRODUCTION**

L'apparition de nouveaux réseaux de distribution, en particulier la vente en ligne suscite une modification structurelle des marchés de la distribution de biens culturels, en particulier de musique et de vidéo. Portée par la réduction des coûts de recherche (pour la demande) et la réduction des coûts de catalogage et transaction (pour l'offre), le développement de la vente en ligne a facilité la distribution de gammes de produits de plus en plus diversifiés. Les conditions et les effets de cet élargissement de l'offre ont donné lieu à de nombreux travaux.

Cependant, les différents acteurs impliqués dans ces secteurs s'accordent mal sur les évolutions concrètes que cela implique. En effet, des désaccords existent sur les tendances actuelles de ces marchés : tendent-ils à se concentrer sur les titres les plus vendeurs – effet « podium » ou « superstar » – ou au contraire à donner une place toujours plus grande aux marchés de niche ?

L'une des hypothèses les plus débattues est connue sous le long d'effet long Tail et a été spécifiquement élaborée à partir de l'exemple des produits culturels. Elle postule que la variété des produits offerts sur les nouveaux canaux de distribution nourrit et élargit la coupure entre best sellers (mise en avant et concentration des succès sur quelques produits stars) et marchés de niches (multiplication de segments aux ventes spécialisées et plus restreintes).

On constate ainsi un double mouvement - concentration croissante des recettes / multiplication de l'offre – rejoignant ce que les économistes ont qualifié de star system (Rosen 1981; Benhamou 2002). La massification de la diffusion provoque le découplage entre le succès de quelques produits-stars de moins en moins nombreux et la multiplication des échecs ou des petites ventes pour les autres. Ce phénomène, très général, s'est vu renforcé par une dynamique de e-commerce au point de convergence des stratégies de producteurs, des distributeurs et des intermédiaires (Benghozi et Paris 2003).

Ce que l'on entend Plus précisément sous le nom d'effet « Long Tail » a été formalisé dans un article séminale de Chris Anderson (2004) qui postule que l'importance et la variété des marchés de niche sur internet fait que l'agrégation de ces petits marchés représente désormais un niveau de recettes comparable si ce n'est supérieur à celui des best sellers. Il appuie son hypothèse sur le cas du marché des biens culturels, livre et musique essentiellement.

Cette notion de Long Tail a connu un succès exceptionnel dans la littérature et parmi les praticiens. Si elle semble a priori très commentée, les résultats empiriques restent malgré tout très rares et peu convaincants : résultats peu démonstratifs et observations contradictoires notamment. La notion de « Long Tail » est pourtant, au départ, un concept analytique très précis identifiant, dans une distribution statistique, l'importance du poids des événements peu fréquents et de faible amplitude.

Le rapport qui suit s'appuie sur un travail de recherche original, effectuée par des élèves dans un cadre pédagogique, sur la base de données inédites, pour tester, identifier et caractériser la nature des effets « Long Tail » sur les marchés français de la musique et de la vidéo. La recherche s'est appuyée sur le traitement statistique des données complètes de vente de supports physiques de musique et vidéo enregistrés (CD et DVD), sur la période 2002-2005. L'analyse a principalement été orientée dans trois directions :

1. l'évolution chronologique annuelle de la structure de distribution des ventes
2. les oppositions canal Internet vs. canal traditionnel
3. La différenciation de la structure des ventes selon les périodes de creux ou de pics de ventes

Les résultats obtenus traduisent plusieurs effets notables. Les premiers montrent que la saisonnalité pèse sur la structure de distribution des ventes et pas uniquement sur leur volume. Les deuxièmes concernent la déformation et l'accentuation, dans le temps, du phénomène de Long Tail au détriment des ventes « moyennes ». Les derniers suggèrent un étalement des prix sur l'ensemble des ventes, plus forts sur les marchés traditionnels que sur Internet.

Au delà de l'intérêt de la recherche pour la validation ou réfutation d'une hypothèse théorique qui a connu une forte audience, la recherche permet aussi de répondre aux fortes attentes des acteurs privés et publics partenaires de la recherche. En effet, l'existence, ou non, de la long Tail a des incidences importantes sur la structuration des pratiques des différents opérateurs : efforts et investissements éventuels de numérisation pour l'élargissement des répertoires et l'autonomisation de plate-formes thématiques, stratégies d'exclusivité dans l'acquisition des droits conditions de maintien d'une diversité culturelle de l'offre.

## 1. Présentation et discussion du phénomène « Long Tail »

L'effet « Long Tail », qui peut être traduit en français par effet « longue traîne » ou « longue queue », est un outil de la théorie des statistiques permettant de mettre en valeur l'importance du poids des événements peu fréquents ou de faible amplitude par rapport à celui des événements à grande fréquence ou amplitude dans le cadre de l'étude d'une distribution.

Dans l'économie de la culture, le terme a été popularisé la première fois sous la plume du rédacteur en chef de Wired (Anderson 2003<sup>1</sup>) mais est resté peu documenté à l'exception, notamment, des travaux menés au MIT (Brynjolson et al., 2003 et 2006) ou à Harvard (Elberse et al., 2007). De façon générale, il est particulièrement important de souligner que dans ses acceptions les plus courantes, tant du côté académique que du côté des praticiens, la notion revêt une acception à la fois descriptive et normative. Descriptive car appliquée aux chiffres de ventes des industries culturelles traditionnelles et en ligne, le phénomène permet de rendre compte de la diversification de la consommation en une multiplication de micro-niches renvoyant à autant de marchés et communautés d'intérêts. Normatif car la théorie de la Long Tail postule également que l'agrégation des produits faisant l'objet d'une faible demande représente un marché de taille comparable à celui des best-sellers et peut donc constituer le support de stratégies et de modèles d'affaires alternatifs de la part des éditeurs et distributeurs de contenus. Ce que l'on qualifie aujourd'hui couramment de « l'économie de la Long Tail » repose ainsi sur des stratégies de vente de contenus de niche (titres anciens, segments spécifiques, versionings particuliers), autrefois largement inexploités en raison de l'insuffisance de leurs niveaux de diffusion effective<sup>2</sup>.

---

<sup>1</sup> Même si C. Anderson revendique la paternité du concept de Long Tail, il convient de noter que ce dernier était déjà mentionné dès 2003 par C. Shirky.

<sup>2</sup> Ainsi, en 2005, les dirigeants de Google n'ont-ils pas hésité à titrer l'une de leur présentation d'entreprise : *Serving the Long Tail*, Google Analyst Day, 9 février 2005.



Le concept de Long Tail est donc générique mais le poids qu'il a pris dans nombre d'analyses de l'économie numérique tient à ce que la notion semble contribuer à différencier fortement et spécifiquement les marchés physiques et les marchés numériques. En effet, dans le premier cas, les contraintes de la matérialité se traduisent pour les distributeurs de contenus par des arbitrages en coûts d'opportunités favorisant les best sellers : ces derniers y sont favorisés car ils permettent de générer des économies d'échelles en matière de logistique et distribution physique. C'est d'une certaine façon la base de l'économie du star system (Rosen, 1981) : elle postule que la réduction des coûts de distribution et de transaction conduit à une homogénéisation et une concentration des formes de consommation autour des produits les plus porteurs. Au contraire, les produits de niche, présentant de faibles taux de rotation, alourdissent à la fois les coûts de distribution et les coûts de stockage, d'où leur moindre présence dans les rayons des distributeurs.

Il est donc intéressant de souligner que la répartition des ventes entre best sellers et marchés de niche donne déjà lieu, dans les circuits de distribution physique, à une différenciation des stratégies entre les trois catégories de distributeurs physiques: les détaillants spécialisés (disquaires, vendeurs de vidéos), la grande et moyenne distribution généraliste et la grande distribution spécialisée (FNAC, Virgin, etc.). Elle se manifeste très clairement dans le nombre de références disponibles dans chacun des canaux de distribution.

Canal de distribution	Ordre de grandeur du nombre de références (2005)
Grande distribution généraliste (Hypermarchés et moyennes surfaces)	5 000
Réseaux spécialisés (FNAC, Virgin)	100 000
Disquaires	25 000
Vente en ligne	900 000

Face à des contraintes logistiques incitant à la mise en avant des titres les plus vendeurs, l'effet « longue traîne » semblerait en particulier opérer pour nombre de détaillants spécialisés choisissant de concentrer leur activités sur des contenus de niche, c'est-à-dire sur la partie basse de la traîne. Généralement situés au niveau de zones géographiques accessibles (grandes villes, etc.), ces distributeurs misent sur leur complémentarité avec le reste des distributeurs physiques ; la faible rotation des contenus qu'ils proposent est contrebalancée par l'exclusivité de ces derniers. Si leur cible adressable est restreinte, leur part de marché peut s'y avérer très importante, du fait d'une concurrence souvent réduite, voire inexistante pour un périmètre géographique donné.



Cette situation de stratégie de niche est somme toute classique dans l'économie de l'information<sup>3</sup>. Les canaux de distribution habituels favorisent les artistes de renom au détriment de la distribution d'artistes plus marginaux, le phénomène étant encore renforcé par la possibilité qu'ont les produits stars d'ajuster leurs dépenses de promotion pour distordre la demande en leur faveur. Dans ce contexte, le P2P contribue au contraire à fournir les contenus à un coût et un niveau de prix beaucoup plus faible, il permet de ce fait à des artistes marginaux, évacués de la compétition sur les canaux traditionnels de distribution, d'entrer malgré tout sur le marché. Comme certains auteurs le montrent<sup>4</sup>, le prolongement en ligne des modes traditionnels de la distribution physique apparaissent dès lors socialement inefficaces dès lors qu'existent des technologies Peer to Peer.

L'analyse des stratégies de niche permet ainsi de souligner plusieurs ambiguïtés importantes dans la manière dont l'exploitation de la Longue Traîne est souvent perçue et décrite.

D'abord, elle est envisagée, selon les cas, comme spécifique ou non de l'économie numérique. En particulier, plusieurs auteurs<sup>5</sup> relèvent que la conjonction entre une multiplication de l'offre (livres, disques, films notamment) et une concentration de la demande est observable pour l'ensemble des biens culturels et s'avère relativement indépendante (pour le livre notamment) du développement du téléchargement et des ventes en ligne. Le phénomène appelle, en particulier des réflexions spécifiques sur l'économie de la traîne. On peut ainsi s'interroger pour savoir si les low sellers relèvent de modèles d'affaires et d'un équilibre économique spécifique (de type low cost ou réservé à des structures de production indépendantes), ou s'ils ne constituent pas, au contraire, des modèles non rentables, déséquilibre structurel, dont la pérennité n'est assurée que par l'arrivée régulière sur le marché de nouveaux intervenants et de nouveaux investissements<sup>6</sup>.

Ensuite, l'économie de la longue traîne peut correspondre par conséquent, soit à une stratégie cumulative d'exploitations individualisées de niche (dans le cadre d'un marché de marchés ou chaque niche et chaque communauté associée est gérée individuellement, dans une logique de portefeuille), soit à une perspective d'agrégation de l'ensemble de la traîne (dans le cadre d'un marché global se différenciant uniquement par les pratiques et les parcours des utilisateurs, dans une logique de *bundling*). Alors que dans les marchés physiques, seul le premier point est réellement à l'œuvre (dans la spécialisation de boutiques par exemple), ce n'est pas le cas dans les marchés en ligne où selon la configuration des sites les deux stratégies peuvent être envisagées. Il faut souligner que l'importance de l'enjeu ne concerne pas seulement la distribution des œuvres mais aussi, plus largement, les investissements que les éditeurs de contenus doivent éventuellement engager pour la valorisation de leur patrimoine. La question qui se pose à eux est en effet celle de l'éventualité de la numérisation complète de leurs bibliothèques de titres<sup>7</sup>.

La Longue Traîne prend donc, dans le cadre des marchés en ligne supportés par l'internet, une importance renouvelée sous l'effet de plusieurs facteurs. La mondialisation des réseaux permet, d'une part, la constitution d'un marché agrégé à l'échelle mondiale, rendant plus facile l'agrégation de préférences individuelles spécifiques en communautés d'échange structurés représentant autant de niches et de micro marchés. C'est la même dynamique qui est à la base de succès de sites d'échanges P2P, de transactions tels que E-Bay, ou de sites communautaires à l'image de FaceBook. La numérisation abaisse, d'autre part, les coûts de diffusion et permet d'assurer à faible coût, sans contraintes de stockage, la mise en ligne et la distribution d'un nombre considérable de titres, sans

---

<sup>3</sup> Elle a en particulier été bien documentée et étudiée, notamment par Varian et Shapiro (1999)

<sup>4</sup> Cf. Zhang 2002, par exemple

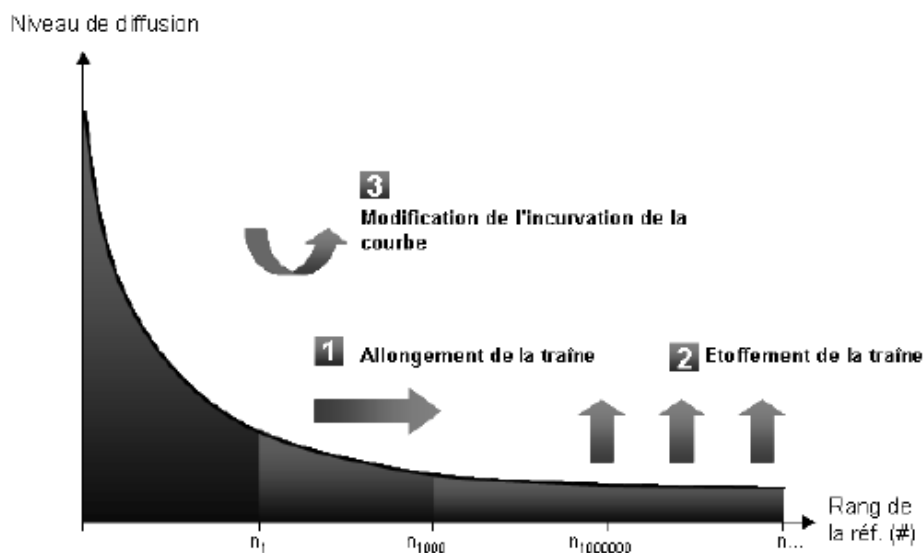
<sup>5</sup> En France, citons notamment Benhamou, Benghozi, Chantepie...

<sup>6</sup> Sur un schéma très proche de celui décrit par Faulkner et Anderson (1987) pour le marché de l'emploi dans le cinéma

<sup>7</sup> Nous verrons plus loin les questions que pose l'appréhension de cette déformation de la structure de la traîne au fil du temps.

commune mesure avec les offres disponibles dans les marchés physiques (magasins ou vente par correspondance). Ces contraintes s'expriment à la fois de manière endogène et exogène : elles pèsent sur les capacités propres des distributeurs (volumes de stockage, superficies de rayonnage, étendue du catalogue disponible), elles conditionnent également l'environnement accessible et adressable (gamme de fournisseur, clientèle avoisinante, déterminisme géographique de segmentation). Du côté des consommateurs, la numérisation et les applications associées (moteurs de recherche notamment, mais aussi listes de recommandations contextuelles, *push* de promotions) facilitent l'exploration et permettent de réduire les coûts de recherche rendant de ce fait plus facile l'accès à des titres de moindre notoriété. Du côté des offreurs et des diffuseurs, la numérisation des transactions ouvre la voie à des possibilités radicalement nouvelles de tarification permettant des stratégies adaptées pour promouvoir les œuvres et assurer une meilleure discrimination des consommateurs (paiement forfaitaire ou variabilité des prix unitaires par exemple).

La conjonction de ces facteurs se traduit non seulement par une augmentation considérable de la consommation globale de musique, mais aussi par une plus grande variété dans les structures d'offre. Le dernier rapport IFPI sur la musique digitale fait ainsi état de 500 services (plateformes légales) de musique numérique dans le monde (contre 50 en 2003), proposant plus de 6 millions de titres (contre 1 millions en 2003).



On peut dès lors expliquer la constitution de la longue traîne par une dynamique en plusieurs étapes :

- 1- allongement de la traîne sous l'effet d'une démocratisation des outils de production (homestudio, blogs, sites personnels) et de la démultiplication correspondante de l'offre
- 2- épaissement de la traîne lié au renforcement des formes de diffusion en ligne (plateforme, agrégateur, réseaux P2P)
- 3 – incurvation et basculement de la distribution par l'amélioration de l'ajustement offre/ demande (communautés d'intérêts) et sous l'effet de la réduction des search costs (moteurs de recherche, recommandations, cookies) : cf. exemple ci-dessous sur le site Amazon.fr.



## 2. Etats des lieux de la situation américaine

### 2.1 Présentation

Le premier travail empirique systématique mené sur la long Tail est sans doute le travail réalisé par Elberse et Oberholzer et publié en 2006. Leur étude porte sur la vente de DVD et de VHS aux Etats-Unis de 2000 à 2005. L'objectif de cette recherche était de tester l'hypothèse de l'émergence d'un phénomène de Long Tail sur ces marchés. Il s'agissait, plus précisément de mesurer si, sur cette période, la proportion des ventes des « petits » artistes – dont les ventes sont inférieures en volume à une quantité fixée – augmente proportionnellement par rapport à celle des « gros artistes », définis de manière identique comme ceux dont les ventes excèdent en volume une certaine quantité.

L'étude utilise un échantillon de 5455 programmes vidéo, dont les caractéristiques de vente sur la période 2000-2005 sont bien connues, à la fois pour les supports DVD et les supports VHS (artiste, genre, sous-genre, langue originale, distributeur, type de vente – location, « sell-through », catalogue...). Sont aussi connues les ventes pour chacun des supports par semaine – donc aussi par mois et année – et par canal de distribution.

### 2.2. Méthodologie

À partir de ces données, le premier travail des chercheurs a été de répartir les titres étudiés par quantiles pour chacun des supports DVD et VHS, suivant le chiffre de vente annuel du titre rapporté au chiffre de vente total pour le support considéré. Les chercheurs disposent donc a priori d'une description complète de la distribution des ventes – donnée par les chiffres de vente pour chaque titre et pour chaque support. Cette répartition en quantiles constitue la base de leur modélisation et de l'exploitation des données dont ils disposent.

Le premier test réalisé sur les données est celui de Kolmogorov-Smirnov. Le but est ici simplement de montrer que la distribution des ventes s'est modifiée sur la période considérée. Il s'agit d'un test très

simple qui consiste fondamentalement à vérifier que la masse de la distribution des ventes se transforme d'année en année – en constatant notamment le déplacement de sa médiane, ou du rapport des masses contenues dans deux parties de la distribution. Cette étude leur permet déjà de conclure à une cohérence des chiffres de vente observés avec l'hypothèse de l'apparition de l'effet Long Tail sur ces marchés, même si l'augmentation du nombre de titres qui ne vendent pas perturbent assez largement leurs données et limite l'intérêt du test<sup>8</sup>.

Deux autres tests sont réalisés, respectivement de régression linéaire et de régression binomiale négative sur les quantiles. Cette méthode permet d'étudier de manière plus spécifique l'évolution sur la période de la distribution des ventes en fonction de leurs caractéristiques – à savoir les filtres vus plus haut, tels que genre, distributeur, mode de distribution, etc. L'idée de ce test est globalement de se demander à quel niveau la mise en évidence de la Long Tail se fait de manière la plus significative, dans un but d'en rechercher la cause. Le défaut de l'étude à ce niveau est d'être relativement peu exploitable : en dépit de l'intérêt central de l'objectif visé, les tableaux obtenus sont en revanche extrêmement difficiles à interpréter et ils ne sont pas tous en lien direct avec la question de départ. Ainsi, un grand nombre de chiffres dans les tableaux et de courbes qui en découlent ne renvoie que très indirectement à la question de la longue traîne : par exemple la répartition en quantiles des ventes sur l'ensemble de la période 2000-2005 en fonction de caractéristiques telles que les ventes de films pour enfant ou de documentaires. En effet, ces courbes ne montrent pas du tout l'évolution sur la période puisqu'elles s'appuient sur des ventes cumulées sur la période en question. De plus, elles ne permettent pas non plus à elles seules de conclure à un effet Long Tail puisqu'elles représentent des coefficients de corrélation entre des chiffres de ventes par quantile et les caractéristiques de ventes, ce qui n'est pas directement exploitable. Seules certaines mesures semblent accréditer l'hypothèse de la Long Tail.

### 2.3. Principaux résultats

Les conclusions de l'étude montrent d'abord que le développement de la diffusion en ligne a profondément modifié le volume et la structure des ventes pendant la période étudiée en favorisant simultanément le développement des stars et l'étoffement de la « traîne ». Chacun de ces deux phénomènes s'opère cependant de façon spécifique en portant des dynamiques contradictoires : renforcement individuel des superstars s'accompagnant d'une plus faible part relative, multiplication de la traîne mais renforcement simultané des méventes.

La vidéo manifeste notamment très nettement un glissement de ces ventes vers la queue de distribution, plus prononcée pour les années récentes et davantage pour le DVD que pour le VHS.

L'étude et l'analyse des données a permis à l'équipe d'Harvard de tester deux explications alternatives des changements qui s'opèrent dans la distribution vidéo. La première repose sur l'existence de changements dans la nature de la population des consommateurs de DVD, correspondant notamment à des *early adopters* de CSP+, aux goûts plus éclectiques et à même de favoriser les produits présents dans la queue de distribution. Sur ce point, les résultats semblent cependant indiquer le contraire car le phénomène s'opère également, même si c'est dans une moindre mesure, sur les ventes de VHS en ligne.

La seconde hypothèse testée est que l'évolution de la consommation de la longue traîne résulte plus d'une augmentation mécanique de l'offre (nouveaux titres et rééditions) que d'un changement dans la demande. Concernant la structure des ventes, les auteurs montrent qu'alors que les revenus générés

---

<sup>8</sup> Notons que la suppression de ces titres dans l'exploitation des données aurait pu être utile ici, en facilitant le travail d'analyse

par les hits tendent globalement à être moins importants, la part individuelle de chacun des best sellers tend elle à croître au fil du temps. Ils le rattachent très directement à une explication en termes de développement d'un marché « *winner-take-all* ».

En queue de distribution, l'étude montre également le caractère particulièrement plat de la structure des ventes : beaucoup de titres étant peu ou pas du tout vendus. Il est à noter que ce phénomène ne semble pas avoir été particulièrement affecté par la croissance du nombre de titres disponibles.

Du point de vue normatif, l'étude s'avère confirmer la difficulté de transformer l'évolution de la structure de distribution en une stratégie de marché profitable orientée sur une valorisation systématique de la traîne. En effet, les opportunités de commercialisation ouvertes par la réduction des coûts de transaction et de diffusion en ligne sont aussi compensées par une démultiplication considérable de l'offre rendant plus difficile la captation de l'attention des consommateurs.

#### **2.4. Analyse de l'étude et perspectives**

Plusieurs éléments intéressants nous paraissent devoir être soulignés et retenus dans l'étude proposée. Au premier chef, même s'il s'agit d'une idée relativement simple, la répartition des ventes en quantiles comme point de départ à l'analyse de l'évolution de la distribution des ventes au cours du temps.

Cependant, un des défauts importants de la modélisation proposée dans l'étude américaine est l'utilisation insuffisante des filtres, notamment dans les régressions linéaires, où l'on s'intéresse pourtant spécifiquement aux facteurs susceptibles de contribuer à l'effet Long Tail. Il semble également intéressant de reproduire et comparer les courbes américaines<sup>9</sup>, par exemple en traçant le nombre de ventes par centile et en lissant l'histogramme obtenu – par année par exemple, sur la période 2002-2005. Pour comparer ces courbes, il faut les déflater individuellement afin de les rendre d'intégrale unité et ne pas voir l'observation de l'effet Long Tail perturbée par une modification du volume des ventes d'année en année – il semble d'ailleurs que l'étude de Harvard n'ait pas vu cette difficulté qui est pourtant tout à fait fondamentale. On pourra alors observer l'évolution de la distribution des ventes sur la période donnée de manière efficace. De plus, en incluant des filtres bien choisis – distributeurs, genre, etc.–, on sera en mesure de fournir une étude qualitative fine sur l'apparition éventuelle d'une Long Tail sur différents marchés bien spécifiques. Les modèles utilisés dans l'étude ne semblent pas en mesure démontrer ou d'infirmer l'hypothèse de l'apparition d'une Long Tail sur de tels marchés de genre par exemple.

## **PARTIE 2 - L'ANALYSE DE LA SITUATION FRANÇAISE**

### **2. Eléments de méthode**

#### **1.1. Présentation des données étudiées**

La recherche réalisée s'est appuyée sur les bases de données privées fournies par la société GfK, et portant sur le marché des ventes de supports physiques de musique et de vidéo – deux bases de

---

<sup>9</sup> Notamment celles fournies p. 42 de l'article original

données séparées – sur la période 2002-2005. Ces bases de données sont extrêmement complètes – a priori beaucoup plus que celles dont disposaient les chercheurs de Harvard.

Les données ont été traitées avec le logiciel WEbIntelligence<sup>10</sup> qui permet de faire des requêtes directement sur une des deux bases de données, en produisant des analyses d’objets variées en fonction de différents filtres.

La base de données GfK permet donc la manipulation de nombreux indicateurs, à la fois d’ordre qualitatif - circuit de distribution, zone géographique (pays, région), référence des produits (noms, code EAN, interprète), genre, format, distributeur, éditeur - et d’ordre quantitatif - intervalle temporel (année, semestre, mois, semaine), prix moyen, chiffre d'affaires, quantité, vente moyenne en volume, rang, nombre de références. La plupart de ces indicateurs sont purement descriptifs (ventes, CA...), d'autres sont des construits (rang, prix moyen, meilleur classement...).

De façon plus détaillée, les objets ou filtres utilisables sont notamment :

- le circuit et le sous-circuit de distribution – disquaires, grandes surfaces spécialisées, grands magasins, Internet,... ;
- la région en France ;
- l’année, le semestre, le trimestre, le mois, la semaine ;
- le code EAN du produit ;
- le nom du produit – des noms de titres pour le marché de la musique par exemple ;
- le distributeur – en tout plusieurs milliers, dont Fnac, Universal, Warner...– ;
- l’éditeur ;
- le format – DVD, VHS... ;
- le nom de l’interprète ;
- le prix moyen ;
- le genre et le sous-genre ;
- le chiffre d'affaires ;
- la quantité ;
- la vente moyenne en volume ;
- le rang.

Ci-dessous, nous présentons, par exemple le chiffre d’affaire des différents circuits de distribution au cours du mois de mai 2005 en utilisant un tableau croisé affichant les produits par genre x circuit de distribution, et en appliquant un filtre temporel (mai 2005).

	AMBIANCE	AUTRE SMINDEFIN BOF	CLASSIQUE	ENFANT	FRANCAISE	INTERNATIONAL	JAZZBLUES	MUSIQUE DU MONDE	
Disquaires	54772	2757	161400	402679	48506	1013230	1215419	220674	352272
Grands Magasins	4114	1847	19459	40118	4651	142960	136496	21197	31014
Grands & Multi-Spécialistes (GMS)	15105	3263	51463	19474	4935	407049	692924	31629	83884
Hypermarchés + (>6500m²)	342486	471037	800048	456798	202421	13261040	11387610	454552	1019336
Hypermarchés - (<=6500m²)	143195	91135	381109	182024	45713	7297787	5526188	160091	482512
Internet	41928	6414	201962	341567	45971	915122	1185330	194043	190175
Librairies niveau 1	13943	1218	49527	91294	14888	314516	376889	52634	73142
Spécialistes Culture	480963	121541	2522617	3826245	360750	11265080	16941795	2576063	3858347
Supermarchés > 1200 m²	9142	157960	134516	2329	1218	2753599	1504855	15562	89986
Vente par correspondance	97284	174	64787	310248	67012	1758174	568567	138033	112904

<sup>10</sup> développé par Business Objects

## 1.2. La construction de séries temporalisées

L'historicisation de ces analyses permet d'établir des courbes spécifiques de variation par période. En revanche, l'accès aux quantiles n'est pas possible directement : il suppose l'exportation des données pour un traitement ultérieur off line. Pour ce faire, la distribution par décile a été effectuée de la manière suivante, synthétisée dans les tableaux de l'Annexe 1 et présentant les distributions de ventes d'articles musicaux par le réseau de distribution traditionnel (magasins spécialisés, supermarchés, disquaires...) et par Internet.

L'analyse est effectuée sur les données hebdomadaires (cf . infra). Chaque semaine est repérée par un code 200552xy où 2005 donne l'année, 52 donne le nombre de semaines dans l'année, et xy donne la semaine considérée. Pour chacune des semaines, un traitement des données identique a été opéré. Tout d'abord, les quantités de ventes sont classées dans l'ordre décroissant. On considère ensuite les quantités cumulées, puis on tire de celles-ci le nombre d'articles différents à considérer pour atteindre le n<sup>ième</sup> décile d'articles vendus.

Quantité	Quantité vendue	Quantité cumulée	Nombre d'articles à considérer pour atteindre le n <sup>e</sup> décile
12	37	37	3 <sup>e</sup> décile : 1
6	33	70	5 <sup>e</sup> décile : 2
25	25	95	7 <sup>e</sup> décile : 3
37	12	107	8 <sup>e</sup> décile : 4
2	7	114	9 <sup>e</sup> décile : 5
7	6	120	10 <sup>e</sup> décile : 8
6	6	126	
33	2	128	

Ainsi, pour chacune des semaines,

1. une première ligne présente les quantités vendues
2. une deuxième ligne donne le pourcentage d'articles différents à considérer pour atteindre les n premiers déciles considérés.
3. une troisième ligne donne le pourcentage d'articles à considérer pour vendre le n<sup>ième</sup> décile seulement.

Pour l'exemple donné, on aurait dans la septième colonne le pourcentage d'articles à considérer pour vendre le 9<sup>e</sup> décile, soit :  $(4-3) \times (100/8) = 12,5 \%$ .

## 1.3. Les limites de la base GfK

A posteriori, malgré le caractère extensif des données disponibles, la base GfK permet de bien saisir les évolutions des ventes physiques mais s'avère inappropriée dans le cadre d'une analyse du phénomène Long Tail étendu aux échanges immatériels. Compte tenu de la nature de la base (biens physiques, canaux traditionnels, intervalle temporel réduit), il est difficile de prétendre caractériser beaucoup plus que des distributions de Pareto déformées. À condition de restreindre ces dernières à leurs parties antérieures, les données sont cependant beaucoup plus intéressantes pour analyser la short head, y compris sur les marchés en ligne : en effet, l'analyse montre qu'il existe une corrélation forte entre les

hits et mégahits des distributeurs traditionnels, et ceux des distributeurs online. Ces résultats sont confortés par les travaux ayant plus directement portés sur le téléchargement.

Les bases de données GfK soulèvent également une autre difficulté, d'ordre purement technique. Elle résulte des conditions d'exploitation ou des modalités d'export proposées.

Business Object, l'outil intégré à la base GfK, apparaît de prime abord puissant. Moyennant un choix raisonnable d'objets et de filtres, le logiciel permet le traitement et la présentation des résultats en tableaux, puis la visualisation de ces derniers en courbes, sous forme graphique. En revanche, l'accès aux quantiles n'est pas possible directement, et la fonction de traçage de courbes reste de portée limitée. Nous fournissons en annexe quelques exemples des résultats que nous avons pu obtenir à l'aide des bases, une fois la prise en main du logiciel effectuée et les principales fonctionnalités maîtrisées.

Le logiciel intégré ne peut donc pas être utilisé pour des traitements sur l'ensemble de la distribution, en raison de lenteurs excessives à l'utilisation. Pour ce faire, il faut le plus souvent exporter les données sous une autre application pour tracer alors les courbes souhaitées.

L'outil Excel, privilégié dans un premier temps pour dégager rapidement des grandes tendances a vite montré ses limites en matière de capacité de traitement. Il rend difficile ou empêche toute manipulation dès que les séries dépassent 256 lignes (2 octets). La mise en évidence de structures saisonnières a permis de se limiter à des analyses hebdomadaires mais il aurait été également nécessaire de travailler plus systématiquement avec des données agrégées sur données semestrielles et annuelles).

## **2. Le cas de la musique**

### **2.1. Le contexte de la musique en France en 2005**

Pour essayer de caractériser et mesurer le phénomène de Longue traîne, nous avons étudié la distribution des ventes de musique et de vidéo en France dans la période comprise entre 2003 et 2005.

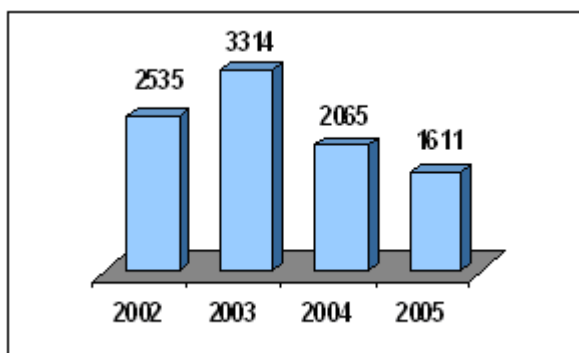
En 2002, au début de la période de référence retenue dans la recherche, il est important de noter que la situation du marché de la musique en France paraît, aux yeux des professionnels, parfaitement saine. L'année 2001 a été exceptionnelle avec un rythme remarquable de croissance du marché : 10,8 %. Si cette croissance est moitié moindre en 2002, les éditeurs phonographiques affichent leur optimisme et se réjouissent cependant d'un chiffre d'affaires éditeurs de 1,3 milliard d'euros (H.T.), représentant malgré tout une progression des ventes de disques, cette année là, de 4,4 %. Le site du SNEP précise même, à cette époque : « La crise internationale qui touche le marché du disque depuis deux années consécutives a, en 2002, une nouvelle fois épargné la France. (...) Les résultats des principaux marchés constatés par l'IFPI à fin septembre 2002 permettent de prévoir, en toute hypothèse, que la France sera cette année le seul marché affichant une croissance des ventes. »



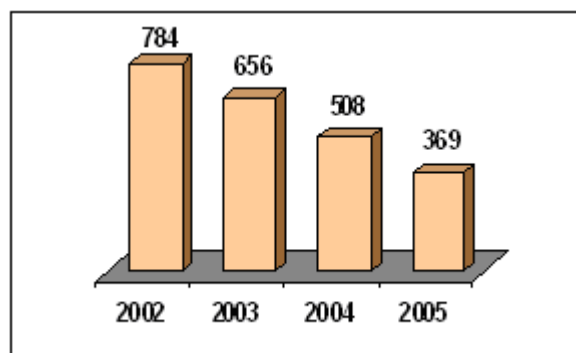
	2001	2002	Evolution
CHIFFRE D'AFFAIRES (ventes gros H.T. en millions d'euros)	1246	1302	+ 4,4%
dont Singles	128,7	133,6	+ 3,8%
dont Albums	1058,7	1095,7	+ 3,5%
dont Vidéo	44,1	53,3	+ 20,7%
UNITES VENDUES (en millions)	166	171	+ 3,3%
dont Singles	38	39,3	+ 3,6%
dont Albums	122,8	125,7	+ 2,3%
dont Vidéo	3	3,8	+ 28,3%

La structure des ventes de musique à cette période permet de mieux comprendre les dynamiques à l'œuvre. La part des titres francophones classés en singles représente 48,5 % en 2002 (contre 46 % en 2001, 53 % en 2000 et 40,1 % en 1999). Le chiffre est à peu près comparable pour les albums (49,6%) même si les ventes de ces derniers s'avèrent un plus stables (48,5 % en 2001, 48,2 % en 2000 et 41,8 % en 1999). De façon synthétique, la part des nouveaux talents dans le classement représente 32 % en 2002 (contre 28 % en 2001 et 26 % en 2000). Il faut cependant noter que les chiffres semblent indiquer un rétrécissement de l'offre disponible entre 2002 et 2005

Nombre d'albums commercialisés



Nombre de singles



A titre de mémoire, les singles qui sont restés le plus longtemps classés en 2002 étaient : 37 semaines en variété nationale pour *Bratisla Boys* (Stach Stach) et 31 semaines en variété internationale pour *Mad House* (Like a prayer). En matière d'albums, Sur les 20 meilleures ventes d'albums, 15 étaient francophones ; 7 albums<sup>11</sup> francophones étaient classés toute l'année (52 semaines consécutives), 2 albums internationaux classés quasiment toute l'année (plus de 50 semaines consécutives).

Le taux de renouvellement dans les ventes est également important à situer.

A la fin de la période considérée, en 2005, la présence moyenne d'un titre dans le top était de 11,4 semaines et 6,8 % des titres ne sont restés qu'une semaine dans ce top. Cela représente, en moyenne, un renouvellement qui s'établit sur la base de 6,9 titres nouveaux par semaine (soit 6,8 % par semaine).

<sup>11</sup> Dany Brillant " Dolce Vita ", Jean-Jacques Goldman " Chansons pour les pieds ", Jean-Louis Aubert " Comme un accord ", Lara Fabian " Nue ", Laurent Voulzy " Avril ", Marc Lavoine " Marc Lavoine ", Noir Désir " Des visages Des figures "

La durée maximale de présence dans le classement d'un titre francophone a été de 41 semaines<sup>12</sup>, les suivants étant à 35<sup>13</sup> et 28 semaines seulement<sup>14</sup>.

La structure de ces chiffres pour la catégorie des albums apparaît sensiblement différente car elle manifeste une plus forte concentration de la consommation sur les best sellers. La présence moyenne dans le top, de 9,8 semaines, est plus faible ; la part des albums n'y restant qu'une semaine est plus importante (15,4 %) ; le niveau de renouvellement est plus accentué (16,8 titres nouveaux par semaine, soit 8,4 % par semaine).

La situation en toute tête de liste est par contre plus contrastée. Au cours de l'année, huit singles différents seulement ont été classés en numéro 1, dont 4 titres francophones<sup>15</sup>. Le délai moyen après sortie pour devenir numéro 1 était de 3 semaines. En matière d'albums, on a compté par contre 22 numéro 1 : Le délai moyen après sortie reste, dans les deux cas, le même pour obtenir cette place, mais ¼ des titres entrent par contre directement à la première place.

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Titres classés	481	480	511	489	511	491	494	505	541	516	502	458
Présence moyenne (semaines)	11	11	10,2	10,6	10,2	10,6	10,4	10,3	9,6	10,1	10,1	11,4
Renouvellement (titres/semaine)	9,5	7,8	8,1	7,6	8,1	8	8	8	8,6	8,2	7,7	6,9
Numéro 1	10	11	12	10	5	10	9	12	14	13	14	8
Délais pour arriver numéro 1 (semaines)	9	6	7	4	8	6	6	4	5	5	4	3
Titres en première semaine (%)	13	10	9	6,9	6,7	6,1	5,3	6,9	7,6	5,8	5	6,8

Au niveau de la situation du commerce en ligne, c'est à partir de l'introduction de l'ADSL en France par France Telecom, en 1999, que les ventes Internet ont commencé à augmenter de manière significative. L'évolution est sensible sur les 3 années plus particulièrement étudiées, qui correspondent au décolllement du commerce en ligne. A titre d'exemple, le site de vente en ligne Fnac.com a doublé, sur la seule année 2004, son nombre de visiteurs uniques. D'après le baromètre ACSEL (Association pour le commerce et les services en ligne), le chiffre d'affaires du commerce en ligne général a augmenté de 60% en 2003, 65% en 2004, 50% en 2005.

<sup>12</sup> Pour Le casse de Brice (Jean Dujardin) et Un monde parfait (Ilona Mitrecey)

<sup>13</sup> Tout le bonheur du monde (Sinsemilia)

<sup>14</sup> Caravane (Raphaël) et C'est les vacances (Ilona Mitrecey)

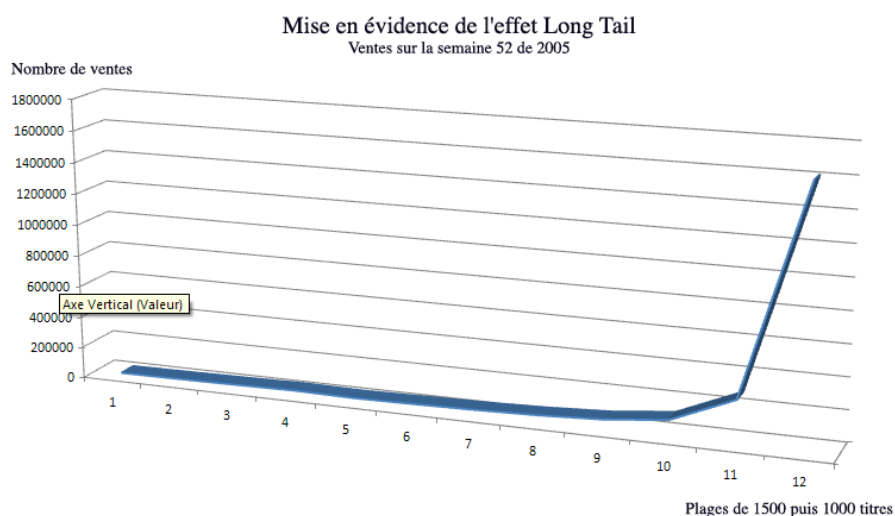
<sup>15</sup> « Ma philosophie » de Amel Bent, « Un monde parfait » de Ilona Mitrecey, « Santiano » de Star Academy 5 et « Mon plus beau Noël » de Johnny Hallyday

France : Part de l'e-commerce dans la VAD (incluant pure-players, VAdistes historiques et enseignes-magasins)		
Année	Chiffre d'affaires de la VAD (en milliard d'Euros)	Part du e-commerce
2003	10,5	35%
2004	11,9	48%
2005	14	62%

## 2.1 Premiers constats

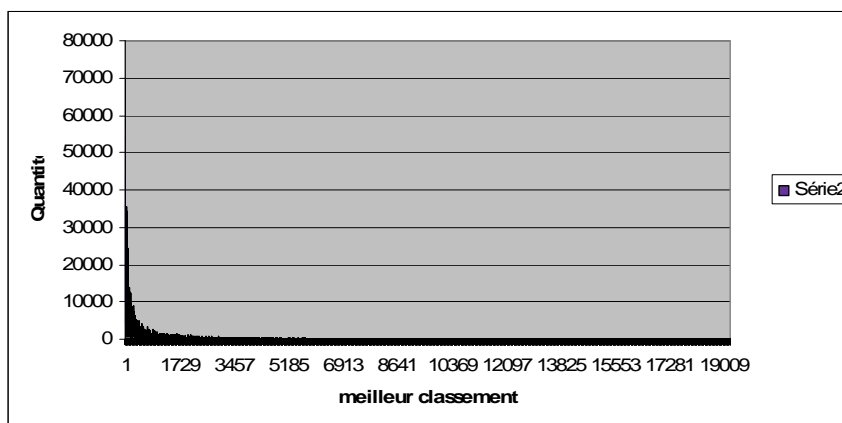
Les premiers tests effectués sur la base de données ont d'abord visé à tester la nature de la distribution des ventes, en mettant simplement à plat les données cumulées par titre, en extrayant de manière aléatoire des semaines de référence.

Le premier des tests a été conduit sur les ventes de CD de la semaine 52/2005 (titres rangés par ordre croissant de ventes). Sur le graphe, les titres en abscisse sont regroupés par tranches de 1500 puis 1000 titres, ce qui correspond à 12 groupes de volumes de ventes :



Au-delà de la forme attendue de la courbe, ce premier test permet déjà un premier constat d'ordre méthodologique. Le découpage par tranches égales ne tient pas compte des résidus aléatoires. En effet, environ 20 000 titres ne sont vendus qu'à un seul exemplaire. Si, d'un point de vue statistique, il serait possible de considérer que ces titres ne représentent que du bruit aléatoire, ce n'est pas le cas du point de vue analytique : en effet, le nombre de ces titres modifie considérablement l'allure du graphe mais il reste difficile d'envisager la vente d'un unique titre comme un véritable marché de niche.

Le second test effectué permet de confirmer ce premier constat. Il porte sur les ventes par artiste pour la semaine 46/ 2004.



### 2.3. Un premier résultat : la saisonnalité

Les travaux menés jusqu'à présent sur la Long Tail, notamment américains, portent en général sur des données agrégées par année. Or tous les chiffres montrent que la répartition des ventes au long de l'année connaît de fortes disparités, particulièrement marquées pour les périodes de fêtes. Il nous a donc paru tout spécialement important de tester la saisonnalité de la Long Tail en cherchant si les variations des volumes de ventes correspondaient ou non à des variations observables de la structure des ventes. Afin de vérifier cette hypothèse, nous avons d'abord retenu une année particulière – 2005 – en mesurant la traîne sur des semaines précises, à différentes périodes de l'année. Notre objectif était d'abord de vérifier l'hypothèse de l'existence de types différents de structure selon les périodes, ensuite de pouvoir identifier des semaines représentatives pour ces types.

L'analyse a été effectuée à partir d'une semaine tous les deux mois, en différenciant les ventes réalisées en magasin et celles réalisées sur Internet.

#### 2.3.1 Résultats en magasin

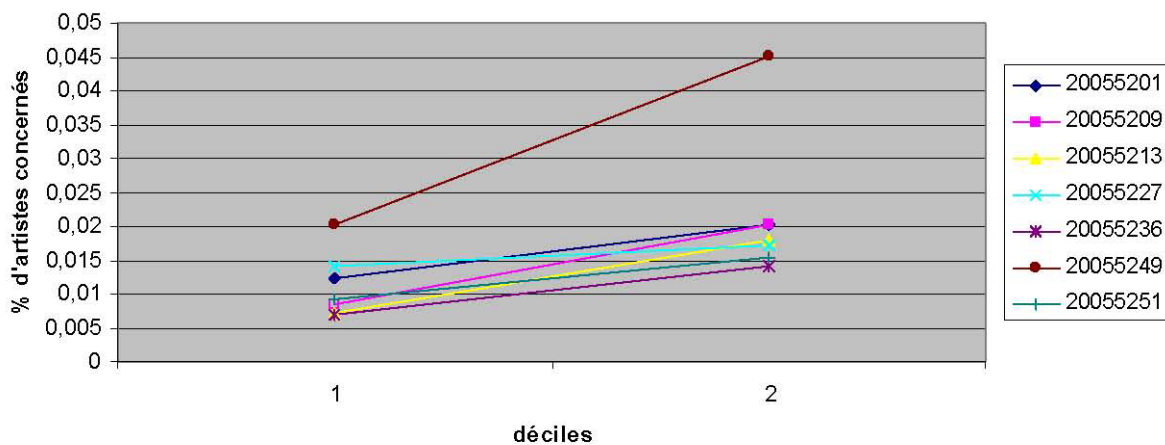
Quand on considère d'abord le simple chiffre d'affaires en valeur absolue, le plus gros volume de ventes se situe, comme on pouvait s'y attendre, en décembre (51<sup>ème</sup> semaine) avec 460 000 ventes contre environ 200 000 au cours des autres semaines de l'année ; un second pic de ventes est également à relever en mars (300 000 ventes pour la 9<sup>ème</sup> semaine).

Quand on envisage ensuite la répartition des ventes par titres, on retrouve également des variations intéressantes à souligner, qui confirment notre hypothèse de départ. On constate bien que le poids des « produits star » varie selon les semaines de l'année.

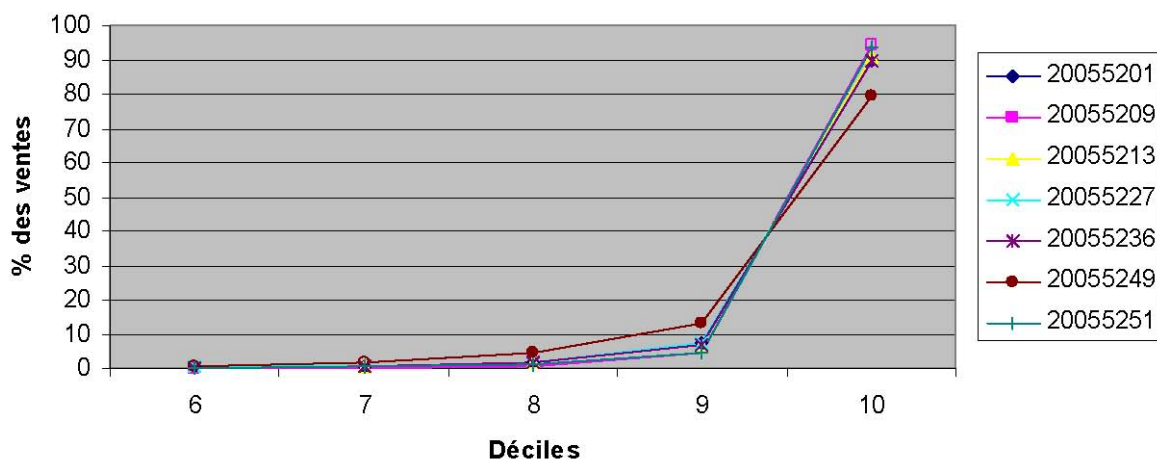
Les premiers déciles apparaissent particulièrement faibles pour la période de Noël et au printemps (mars à mai : semaines 9 et 13) : 4 ou 5 produits réalisent 10 % des ventes, contre environ 10 à 15 pour les autres périodes de l'année. Parallèlement, l'effet Long Tail apparaît tout aussi nettement hors des périodes de Noël et du printemps. En effet, les 70 premiers % des ventes sont effectués par 0,4 à 0,6 % des articles les mieux vendus aux périodes de Noël et au printemps, contre 1 à 1,5 % des articles le reste du temps. On constate en outre, pour la période précédant Noël (semaine 49) que ces mêmes 70 % des ventes ont été effectués par 3 % des articles. Cela signifie donc que les 30 derniers % des ventes sont

effectués, respectivement par environ 99,5 %, 98,5 % et 97 % des articles. L'effet Long Tail est donc bien plus particulièrement marqué hors des périodes de Noël et de printemps.

**Ventes de CD en 2005 hors Internet -visualisation de la tête**



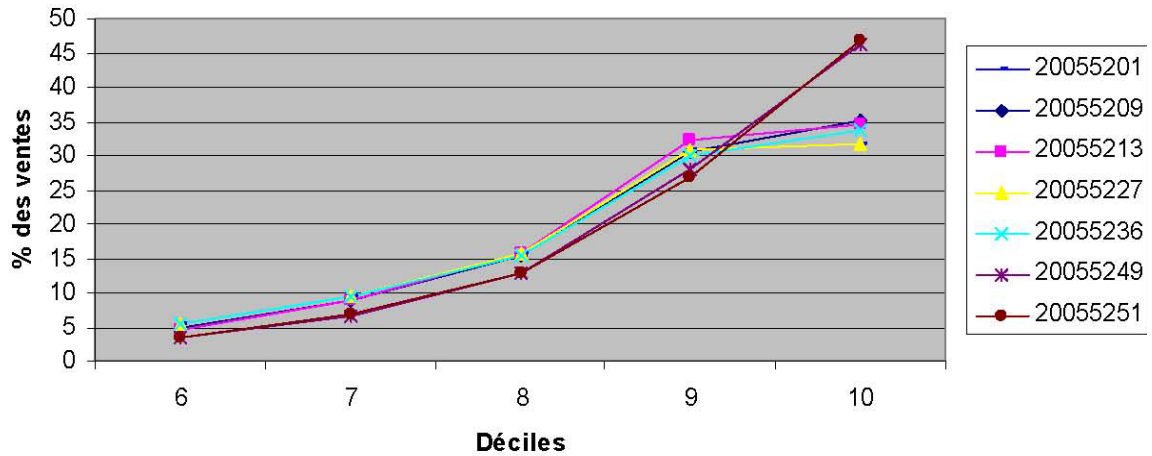
**Ventes de CD en 2005 hors Internet - visualisation de la tail**



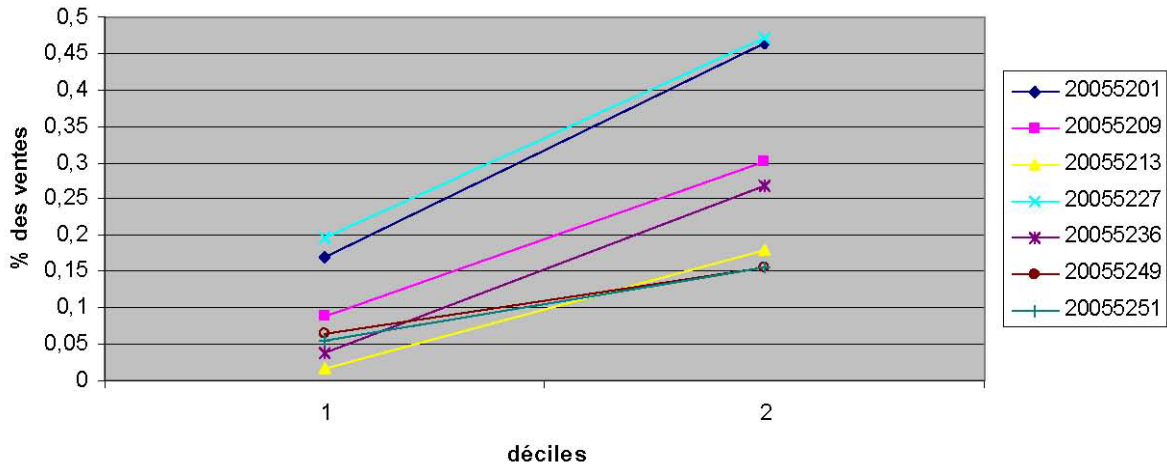
### 2.3.2 Résultats des ventes par Internet

En valeur absolue, les volumes sont bien moindres sur internet avec environ 50 000 ventes pour une semaine, avec des pics à 100 000 durant la période de Noël. L'effet « produit star » varie par contre beaucoup plus fortement selon les semaines puisque le premier décile des ventes correspond selon les cas à 2 articles (semaine 13), une dizaine (semaines 9, 49 et 51) et atteint près de 25 (semaines 1 et 27). Ces résultats sont certes à moduler car ils peuvent aussi s'expliquer par les sorties (ou non) de produits fortement médiatisés. Cette explication est toutefois limitée car l'écart est nettement observable entre les résultats des ventes en ligne et en magasin.

### Ventes de CD en 2005 sur Internet - visualisation de la tail



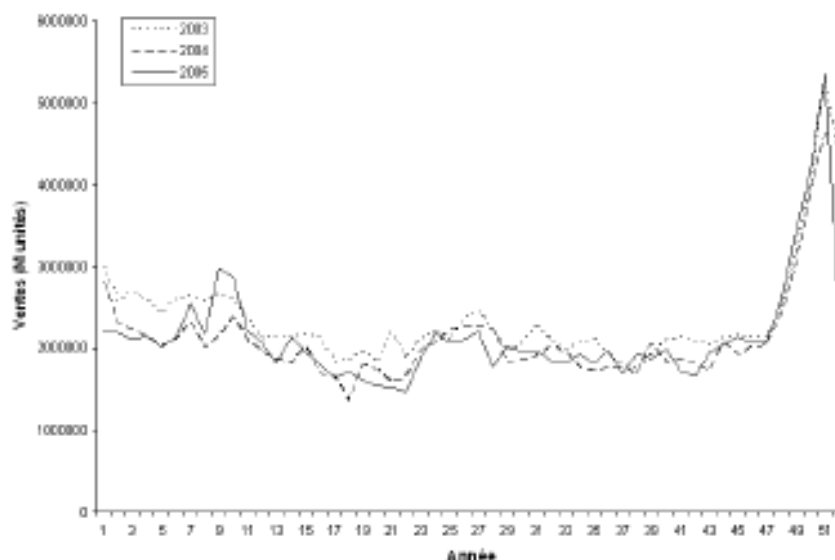
### Ventes de CD en 2005 sur Internet - visualisation de la tête



### 2.3.3 Conclusion

Globalement, on peut donc conclure qu'une année (ici, l'année 2005) connaît deux types de périodes : les périodes calmes, qui peuvent servir d'échantillon de ventes types, et les périodes de ventes plus dynamiques que sont Noël et le printemps pour les ventes en magasin, et pour les ventes en ligne, uniquement Noël. Les périodes calmes sont moins sensibles à l'effet « produit star » que les périodes connaissant des ventes plus importantes. En revanche, l'effet Long Tail s'observe plus en dehors des périodes de ventes dynamiques.

Pour tester ce résultat, nous nous sommes intéressés aux années antérieures, que nous avons mises en relation avec l'année 2005. Les résultats sont synthétisés dans la figure ci-dessous.



Évolution hebdomadaire des ventes cumulées (France, 2003-2004).

Il apparaît que si le pic de Noël est flagrant, Celui du printemps est moins manifeste, dès lors que l'on tient compte d'années autres que l'année 2005. Son importance apparaît plus conjoncturelle.

Nous relevons des corrélations supérieures à 80 % entre les canaux Internet et hors Internet sur chacune des années testées. Ce résultat empêche, à ce stade, de mettre en évidence un étoffement de la traîne en dehors des périodes de pic. En revanche, la comparaison interannuelle permet de confirmer une accentuation de l'effet superstar lors de la période de Noël : ainsi en 2005, le premier percentile représente 74 % des ventes totales en semaine 51, contre environ 65 % en période normale (semaine 20). Ce résultat est observable dans des proportions similaires en 2003 et 2004.

En conclusion, l'analyse nous permet de confirmer un effet inédit de saisonnalité annuelle non relevé jusque là dans la littérature. Deux types de périodes peuvent être mis en évidence, marquées par des manifestations propres du phénomène Long Tail. Des semaines, dites de pic, se situent pendant les fêtes de Noël et de fin d'année. Elles se caractérisent par une hausse significative et bien connue des ventes en volume mais au cours de cette période particulière, on constate également un renforcement de l'effet « produit star » et un creusement de la traîne, aussi bien sur Internet qu'en magasin. Ce résultat s'explique par le caractère d'étreintes des ventes de cette période. Or dans un contexte de cadeaux, les consommateurs n'achètent pas pour eux mais pour d'autres. Il existe, dans ce cas, une forte incertitude quant aux goûts et à la satisfaction du consommateur final destinataire du cadeau : l'acheteur privilégié de ce fait des choix consensuels, en privilégier les hits et les titres les plus achetés par d'autres, justement ceux qui sont le plus surexposés lors des périodes de pic (publicité, promotion télévisée, positionnement en têtes de gondoles, etc.)

## 2.4. Caractérisation statistique de la Tail

Afin de vérifier l'effet superstar ainsi que l'effet Long Tail, deux indicateurs statistiques seront utilisés : le coefficient d'asymétrie et le coefficient d'aplatissement<sup>16</sup>. Le calcul du coefficient d'asymétrie permet de

<sup>16</sup> La présentation détaillée de ces indicateurs statistiques est disponible dans l'Annexe 5

détecter si la distribution est, par rapport à la moyenne, plus étalée à gauche ou à droite ou si, au contraire, les observations sont également réparties de part et d'autre de la moyenne.

L'analyse est effectuée sur la base d'une caractérisation statistique de la situation dans chacune des semaines retenues. A titre d'exemple, nous fournissons ci-dessous les détails correspondant aux ventes de la 9<sup>ème</sup> semaine de l'année 2005 sur internet (à gauche) et hors internet (à droite). L'ensemble des données utilisées dans l'analyse qui suit est développée dans l'Annexe 3.

Nombre	14580	Nombre	59613
Moyenne	56,39	Moyenne	500,70
Médiane	24	Médiane	37
Écart moyen	56,39	Écart moyen	797,43
Écart type	314,11	Écart type	12607,22
Variance	98666,75	Variance	158941953,96
Coefficient de variation	557%	Coefficient de variation	2518%
Asymétrie	50,34	Asymétrie	168,78
Aplatissement	3664,00	Aplatissement	35045,11
Minimum	0	Minimum	0
Maximum	26318	Maximum	2692124
Étendue	26318	Étendue	2692124
Premier quartile	15	Premier quartile	20
Troisième quartile	41	Troisième quartile	106
Int. semi-interquartile	13	Int. semi-interquartile	43
Premier décile	10	Premier décile	13
Neuvième décile	82	Neuvième décile	356

En comparant les données sur les Chiffres d'Affaires (CA), on peut remarquer que les médianes sont du même ordre de grandeur alors que les moyennes sont sensiblement plus élevées hors Internet. Ce constat laisse supposer que les hits rapportent plus hors Internet que sur Internet ; c'est un indice d'une consommation plus large en ligne, le consommateur s'orientant vers des titres plus divers que les nouveautés présentes de manière forte dans les magasins.

De même, une grande variabilité est observable quand on considère l'écart moyen<sup>17</sup>. Ce dernier est beaucoup plus important hors Internet que sur Internet. Ce constat s'explique a priori par le plus grand nombre de produits disponibles en ligne, qui génère une dispersion plus grande du chiffre d'affaires. En revanche, le coefficient de variation<sup>18</sup> est sensiblement le même sur les deux marchés : le résultat peut indiquer que l'on retrouve finalement une variabilité du même ordre sur Internet que hors Internet. Afin de préciser ce résultat contradictoire avec les constats précédents, nous nous intéresserons par la suite au volume de ventes et non plus au chiffre d'affaires.

L'étude plus particulière de l'asymétrie et de l'aplatissement tend à confirmer les premières observations effectuées sur l'étude graphique : l'effet longue traîne apparaît plus prononcé sur Internet (sans doute car les consommateurs y ont plus de facilité pour trouver un article spécifique rare), alors que l'effet superstar est plus prononcé hors Internet (car les ventes de grands hits en magasins touchent une frange de consommateurs différentes et qu'en outre elles peuvent résulter plus facilement d'achats

<sup>17</sup> Mesure de dispersion indiquant la moyenne des valeurs absolues des écarts de chaque valeur par rapport à la moyenne

<sup>18</sup> C'est-à-dire l'écart-type sur la moyenne



d'impulsion<sup>19</sup>).

Si l'on considère désormais les volumes de ventes, on s'aperçoit que la médiane – à la fois sur Internet et hors Internet – est bien inférieure à la moyenne sur les marchés traditionnels. Le coefficient de variabilité cette fois-ci est de 2 à 3 fois supérieur sur les marchés traditionnels que lorsqu'on considérait le chiffre d'affaires. Cet effet était masqué du fait de la pondération par le prix. De cette analyse de la médiane, on peut déduire qu'un CD appartenant à la longue traîne sera positionné sensiblement dans la même partie de la distribution des ventes sur Internet et hors Internet. En revanche, concernant la partie comprise entre la médiane et le début de l'effet superstar, on ne peut a priori pas dire grand-chose hormis le fait que la variation est plus importante sur les marchés traditionnels.

L'asymétrie et l'aplatissement nous donnent davantage d'informations : l'asymétrie est du même ordre de grandeur en volume de ventes alors qu'en général le coefficient d'aplatissement est au minimum 3 fois supérieur hors Internet. Dans ce cas, l'effet superstar en volume se manifeste de manière identique sur les deux marchés alors que l'effet Long Tail est plus important sur Internet.

## 2.5. L'atténuation de l'effet superstar sur la période 2002 – 2005

Le tableau et les graphes qui suivent présentent l'évolution des ventes de CD pour le mois de janvier de 2002 à 2005. Ils permettent de conforter les premières observations sur le poids de la queue et l'étoffement de la queue de distribution. Ils ne permettent par contre pas, en première analyse, de mettre en évidence, au fil du temps, une augmentation ou une diminution spécifique de l'effet superstar.

Ventes de CD (Internet et marchés traditionnels confondus) Janvier - 2002 à 2005					
Plus grosses ventes	10%	25%	50%	90%	99%
Année					
<b>2002</b>	0,003% (2)	0,018% (12)	0,114% (76)	3,642% (2387)	40,505% (26545)
<b>2003</b>	0,009% (6)	0,035% (23)	0,184% (121)	5,967% (3911)	45,713% (29958)
<b>2004</b>	0,006% (4)	0,034% (22)	0,220% (144)	8,543% (5599)	54,828% (35931)
<b>2005</b>	0,008% (5)	0,035% (23)	0,227% (149)	8,680% (5688)	54,034% (35409)

% des titres (et nombre de titres) qui réalisent le % des plus grosses ventes indiqué

Pour préciser cette évolution sur plusieurs années, nous avons choisi deux semaines "étalons", en nous appuyant sur les résultats obtenus dans la partie précédente. Nous avons plus particulièrement retenu l'observation que les périodes de Noël et du Printemps sont dynamiques, mais que Noël profite aux ventes en ligne et en magasin alors que le Printemps profite surtout aux ventes en magasin. Le choix des

<sup>19</sup> D'après une étude Fevad, Médiamétrie-Nielsen Netratings d'octobre 2006, 76% des consommateurs téléchargeant de la musique en ligne ont entre 25 et 49 ans. A l'opposé, les ventes en magasin touchent des populations a priori moins à même d'opérer des téléchargements en ligne : les très jeunes ou les plus de 50 ans.

semaines 09 (printemps) et 51 (Noël) devrait donc permettre de comparer et spécifier les ventes en ligne et en magasin.

Les graphiques présentés dans l'annexe 5 permettent déjà d'opérer une première comparaison visuelle de l'évolution année après année sur les semaines 9 et 51, des ventes en magasin et sur Internet.

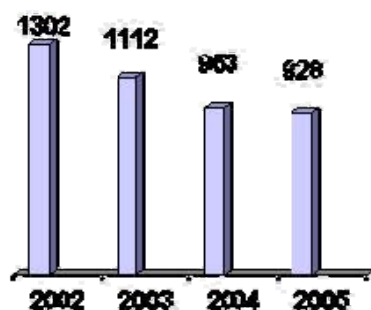
#### Semaine 9, ventes sur Internet

	Quantité Cumulée	Pourcentage cumulé déciles 1 à 9	Pourcentage cumulé déciles 1 à 6	Pourcentage cumulé déciles 1 à 3	Pourcentage 1 <sup>er</sup> décile
<b>2003</b>	55 572,000	80	20	2	0,07
<b>2004</b>	47 788,500	84	12,5	3,5	0,22
<b>2005</b>	55 467,300	85	13,5	1,5	0,13

La comparaison entre années pour la semaine 9 fait apparaître plusieurs dynamiques contrastées intéressantes à relever et qui traduisent une tendance au creusement de la traîne.

- L'évolution du premier décile (le nombre d'artistes qui font les 10% des plus grosses ventes) marque une augmentation au fil du temps et semblerait indiquer une atténuation de l'effet superstar.
- Dans la partie centrale de la distribution, 60% des ventes sont réalisées par un nombre diminuant d'artistes (20 en 2003, 13,5 en 2004), marquant un creusement de la courbe
- Dans la queue de distribution, 90% des ventes sont réalisées par un nombre croissant de titres (de 80 à 85% du total des titres), ce qui irait dans le sens d'un étoffement de la traîne.

Il faut noter que ces dynamiques s'inscrivent dans un contexte global de réduction de la quantité cumulée des ventes correspondant à la croissance du téléchargement en ligne de 2003 à 2005 et à la régression du marché global de la musique (-28,7% de 2002 à 2005). On peut donc parler d'augmentation de la quantité cumulée relative.



Évolution du marché de la musique en millions d'euros

L'observation de la semaine 51 permet d'abord de vérifier l'augmentation importante des quantités vendues pendant les périodes de fêtes : quasiment doublées par rapport à la semaine 9. L'analyse plus fine des pourcentages par déciles confirme les constats effectués à propos de la saisonnalité. Contrairement à la semaine 9, les évolutions des premières ventes ne manifestent pas de tendance notable en matière d'effet superstar. Dans la même temps, la queue de distribution tend plutôt à se renforcer, en augmentant de 4,5 points environ entre 2003 et 2005 – que l'on considère 60 ou 90% des ventes réalisées.

#### Semaine 51 – Ventes sur Internet

	Quantité Cumulée	Pourcentage cumulé déciles 1 à 9	Pourcentage cumulé déciles 1 à 6	Pourcentage cumulé déciles 1 à 3	Pourcentage 1 <sup>e</sup> décile
<b>2003</b>	107 472,000	83	15	2	0,16
<b>2004</b>	110 911,200	91	8	1	0,07
<b>2005</b>	91 860,080	87,5	11,5	1	0,12

Comme attendu, l'analyse des ventes sur les marchés traditionnels montre que l'effet superstar est beaucoup plus présent que pour les ventes sur Internet et que la queue de distribution est beaucoup plus restreinte. L'évolution reste très stable, avec un léger renforcement de l'effet superstar (60% des ventes - réalisées par 4% des titres en 2004 - ne le sont plus que par 2,5% des ventes en 2005).

De leur côté, les quantités cumulées ne présentent pas de tendance remarquable. La faiblesse tout particulière des ventes de la semaine 9 au cours de l'année 2004 peuvent s'expliquer par une offre musicale faiblement attractive durant cette période.

#### Semaine 9 – Ventes hors Internet

	Quantité Cumulée	Pourcentage cumulé déciles 1 à 9	Pourcentage cumulé déciles 1 à 6	Pourcentage cumulé déciles 1 à 3	Pourcentage 1 <sup>e</sup> décile
<b>2003</b>	2 589 104,000	95	4	1	0,12
<b>2004</b>	2 079 341,170	96	3	1	0,09
<b>2005</b>	2 906 543,200	96,5	2,5	1	0,09

Enfin, l'analyse de la semaine 51 sur les canaux traditionnels montre qu'en magasin comme sur Internet, les quantités doublent également pour Noël. La réduction du nombre d'artistes dans le premier décile indique un léger renforcement de l'effet superstar. Ces variations restent cependant limitées, ce qui

semble indiquer que Noël est une période où les consommateurs ont des comportements plus stables que pendant le reste de l'année.

#### Semaine 51 – Ventes hors Internet

	Quantité Cumulée	Pourcentage cumulé déciles 1 à 9	Pourcentage cumulé déciles 1 à 6	Pourcentage cumulé déciles 1 à 3	Pourcentage 1 <sup>e</sup> décile
<b>2003</b>	5 164 508,575	93,5	5,5	1	0,19
<b>2004</b>	5 135 874,864	96	3	1	0,12
<b>2005</b>	5 246 636,610	95,5	3,5	1	0,13

## 2.6. Synthèse

Comme nous l'avons indiqué dans l'introduction, la période retenue dans l'étude a été tout particulièrement marquée par le basculement des pratiques d'achat. Les consommateurs se sont d'une part mis à acheter de plus en plus souvent sur internet ; ils ont, d'autre part, commencé à passer de l'achat de supports physiques de musique enregistrée à des pratiques – payantes ou illégales - de téléchargement. Compte tenu de cette évolution rapide, repérable notamment dans la croissance du marché de la vente de CD à distance sur Internet, on aurait pu s'attendre à des évolutions importantes, même sur une courte période de 3 années. Néanmoins, force est de constater que les variations structurelles restent faibles.

Celles-ci montrent toutefois un étoffement de la longue traîne sur les ventes de CD en ligne par rapport aux ventes en magasin. A l'inverse, les résultats n'indiquent aucune tendance nette quant à l'évolution de l'effet superstar au fil des ans. Ces résultats, qui semblent confirmer les thèses d'Anderson, rendent donc d'autant plus important un élargissement du travail engagé au téléchargement en ligne.

Comme pour les constats de saisonnalité, le caractère limité des évolutions rendent les résultats délicats à interpréter car particulièrement sensibles aux variations résultant des modifications de l'offre artistique d'une année sur l'autre. Comme l'analyse a été faite sur des volumes de ventes normalisés, un ou deux énormes succès peuvent, par exemple, sensiblement affecter l'allure de la traîne en renforçant son aplatissement.

## 3. Le cas de la vidéo

Parallèlement au cas de la musique, le secteur de la vidéo est apparu particulièrement intéressant à étudier pour plusieurs raisons. Il doit d'abord faire face, avec un léger retard - dû notamment aux limites techniques de débit des réseaux - aux évolutions qui ont d'abord touché la musique enregistrée : poids grandissant des téléchargements et développement des échanges P2P notamment. Ces nouveaux

enjeux apparaissent ensuite à un moment où la structure du marché a connu une forte dynamique liée à la substitution entre supports DVD et VHS et au foisonnement des chaînes de télévision et des programmes audiovisuels. Au tournant des années 2000, on a observé ainsi une croissance très importante du marché de la vidéo enregistrée, que ce soit en volume de ventes et en nombre de titres (vendus et disponibles). Enfin, le dernier intérêt d'analyser le secteur vidéo est plus formel ; il offre une bonne base de comparaison pour analyser les situations respectives des marchés français et américains, sur la base de données empiriques très voisines.

### 3.1. Le contexte de la vidéo en France en 2005

Il est inutile de revenir dans le détail sur le mécanisme qui a vu un support de meilleure qualité comme le DVD se substituer rapidement au VHS. Ce phénomène irréversible est désormais bien connu et se trouve lui-même désormais dépassé par les enjeux qui se nouent autour du renouvellement du DVD (avènement du Blue Ray) et sur les interrogations touchant, tout comme pour le CD musical, à sa pérennité même (basculement vers le téléchargement et la VoD).

Les achats de supports vidéo totalisaient, en 2005, 143,41 millions d'unités vendues pour un chiffre d'affaires de 1 784,18 millions d'Euros. Au-delà de la baisse de la part relative du VHS dans le marché de la vidéo, au profit du DVD (98,5% en volume pour le DVD en 2005), on observe également, pendant cette période, des fluctuations intéressantes du volume global des ventes. On constate d'abord une augmentation des ventes en volume (en hausse de 16,9 %). Cette hausse se retrouve également quand on considère le nombre de titres : Selon le SNEP, le nombre de titres délivrés en France par les majors aux plateformes (productions propres, licenciés inclus mais hors distribués) était ainsi, au 31 janvier 2005, de 436 172 titres : cela représentait, en l'occurrence, en augmentation de 44,1% sur le trimestre précédent<sup>20</sup>. A l'inverse, la valeur du marché du DVD et de la cassette VHS connaît, dans le même temps, une tendance significative à la baisse : d'abord légère, elle atteint 8,8 % en 2005. Ce constat permet de souligner la troisième considération importante de l'évolution du marché pendant la période. Elle touche l'évolution des prix des DVD. En 2005, un DVD a été en moyenne vendu 18,9 % de moins qu'en 2004, passant de 15,25 euros à 12,36 euros. Cette diminution unitaire des prix de vente contribue bien sûr, toutes choses égales par ailleurs, à la baisse du chiffre d'affaires global.

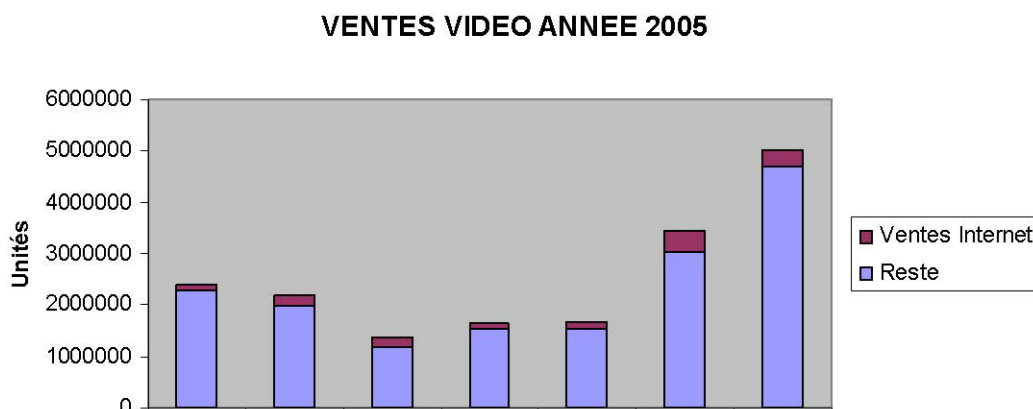
Du point de vue structurel, il est intéressant de comparer la présence des titres dans les tops de vente avec les données analogues présentées pour la musique.

	2005
<b>Titres classés</b>	333
<b>Présence moyenne (nombre de semaines)</b>	6,2
<b>Renouvellement (pourcentage de titres/semaine)</b>	5,6
<b>Nombre de numéros 1</b>	19
<b>Délais pour arriver numéro 1 (semaines)</b>	1,5
<b>Titres présents 1 semaine (%)</b>	22,2

<sup>20</sup> Chiffre obtenu sur la base Médiadisque à partir du code ISRC.

### 3.2. Saisonnalité du marché de la vidéo

Le graphe ci-dessous montre que les ventes de supports vidéo fluctuent le long de l'année. On retrouve, en première observation, le même type de périodicité que dans la musique, marquée par une forte prépondérance des ventes à l'occasion des fêtes de fin d'année et un creux pendant l'été. La fourchette de variation apparaît simplement de moindre importance.

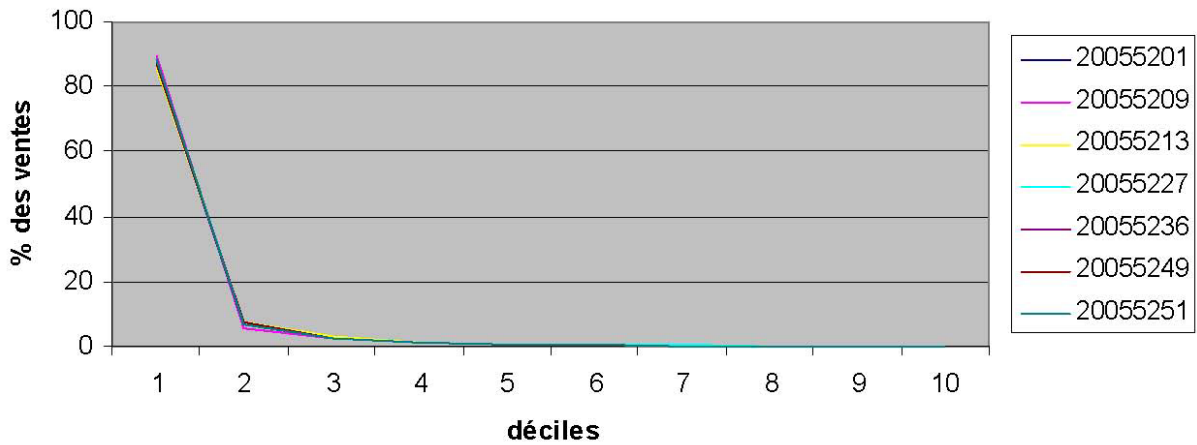


D'avantage encore sans doute que pour la musique, les pics de fin d'année s'expliquent en partie par le rôle important qu'a joué le DVD - et son caractère récent de nouveauté - comme support de cadeau. Les enquêtes établies par le CNC indiquent par exemple que 65,7% des Français déclarent offrir des DVD de temps en temps et, parallèlement, que 19% de l'approvisionnement foyers se fait sous forme de cadeaux.

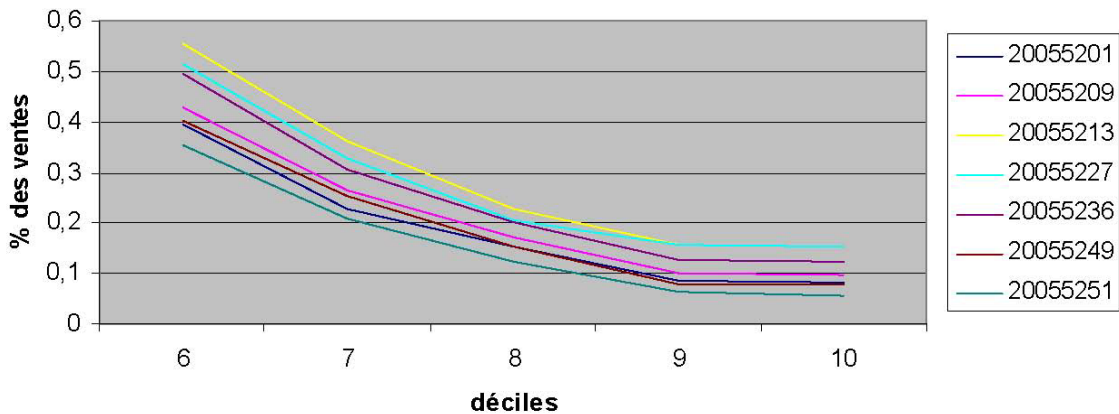
La courbe des ventes se retrouve de façon très voisine quand on distingue l'évolution des chiffres de ventes selon les réseaux habituels (hypermarchés, vidéothèques...). Pendant les années considérées, la proportion des ventes de DVD par Internet restait faible. Il faut toutefois souligner que la part qu'elle représentait sur les ventes total connaît des disparités au long de l'année, traduisant des premiers éléments de différences avec le secteur de la musique. D'une part, le niveau le plus faible - 4,14% - est obtenu, de façon attendue, pour la période de creux maximum (semaine 1), soit juste après le pic des fêtes, au tout début du mois de janvier.

D'autre part, au-delà de cette première observation sur la périodicité. L'analyse des ventes de vidéo confirme également l'importance de l'effet superstar, très clairement observable. Environ 90% des ventes est réalisé par seulement 10 % des productions, quelle que soit la période de l'année. La structure des ventes, similaire sur toute l'année, témoigne, de ce point de vue de l'hégémonie des productions les mieux vendues sur le marché. L'effet Long Tail est par contre presque inexistant, avec 50% des productions réalisant seulement 1,12 % des ventes physiques de vidéo sur toute l'année.

### Ventes en 2005 de vidéos hors Internet



### Ventes en 2005 de vidéos hors Internet- visualisation de la "Tail"

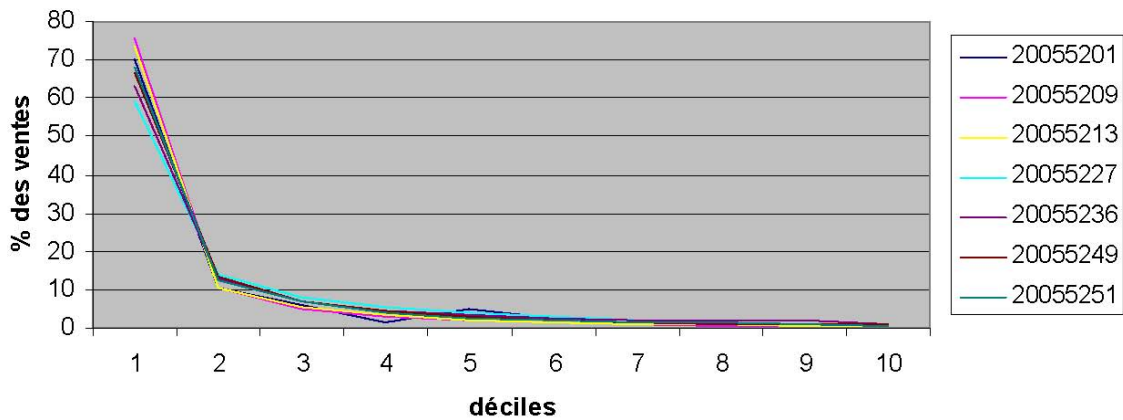


La comparaison des ventes en ligne et des ventes physiques indiquent par contre des facteurs intéressants de différenciation. On constate d'abord, de façon globale, que la proportion maximale de vente en ligne par rapport aux ventes traditionnelles - 11,97% - est obtenue au cours de la première semaine du mois de mars (semaine 13) et non pas sur les périodes de pic ou de creux de consommation par exemple. Une étude plus précise des achats de vidéo sur Internet confirme cette première observation. Elle montre que les fluctuations des achats ne respectent a priori pas les mêmes rythmes que les achats physiques. Au-delà des pics observés durant les dernières semaines de l'année, la semaine 3 présente notamment, un sommet local, qui indiquent que les flux de consommation peuvent connaître des variations de périodicité plus complexes durant l'année pour la vente en ligne que pour les réseaux traditionnels.

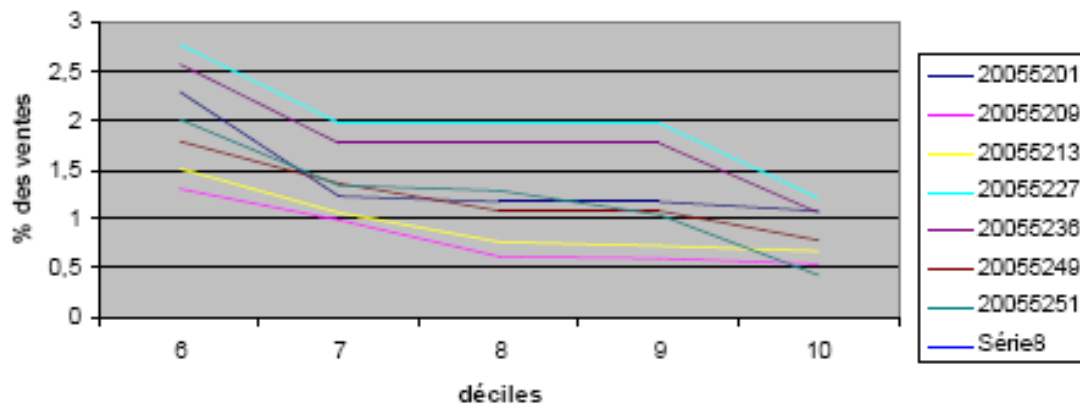
Ces données semblent indiquer que les habitudes de consommation sont sensiblement modifiées dans les achats en ligne. Dans la distribution des ventes, la part du premier décile perd en moyenne près d'une vingtaine de points pour s'établir à 67,9% en moyenne. Le deuxième décile ainsi que le reste des productions les moins vendeuses gagnent en importance. En résumé, les ventes de vidéo sur internet présentent donc :

- l'aspect d'un graphe à la pente moins raide,
- dont la traîne est plus étoffée même si son importance reste limitée
- connaissant une distribution variée le long de l'année. Les cinq derniers déciles accumulent 6,67% des ventes en ligne contre 1,12% dans le cas des ventes physiques, un chiffre 6 fois plus élevé.

### Ventes de vidéo sur Internet

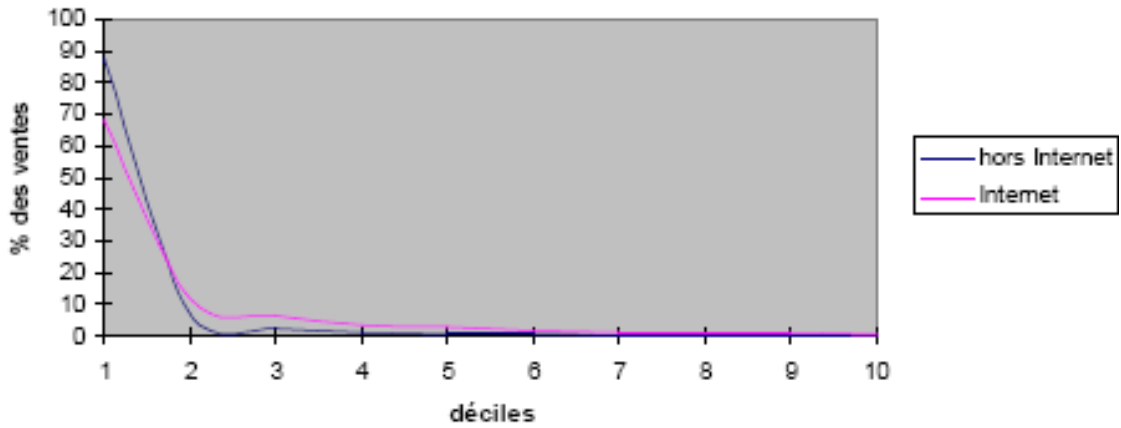


### ventes en 2005 de vidéos sur Internet-visualisation de la "Tail"

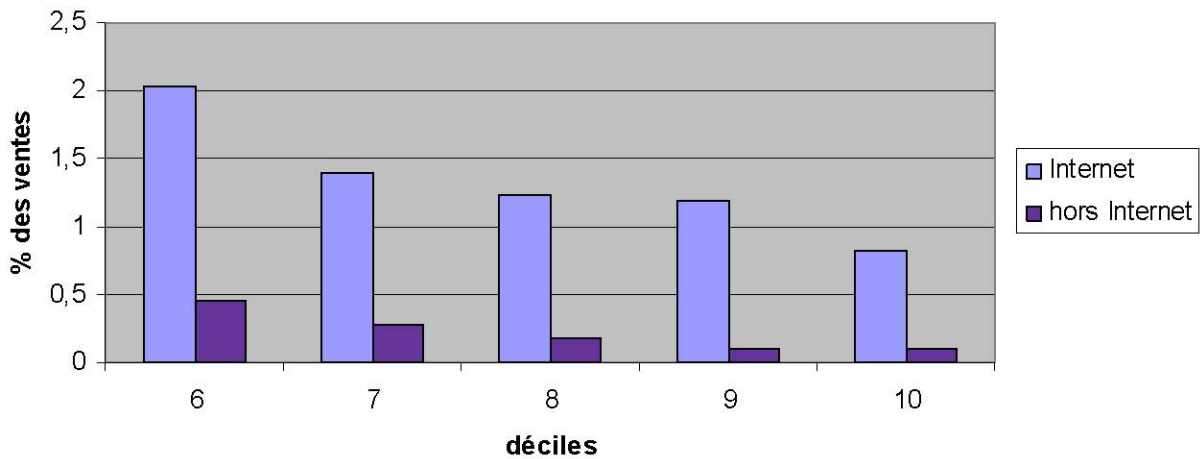




### Mise en évidence de l'Effet Long Tail sur le marché de la vidéo, année 2005



### Effet Long Tail sur vidéos en 2005 : moitié la moins vendeuse des productions



S'ils se confirmaient dans le téléchargement et dans la VoD, ces données pourraient laisser entrevoir une plus grande variété pour les achats en ligne de vidéo que dans la musique, en ouvrant donc des perspectives plus favorables pour les petites productions.

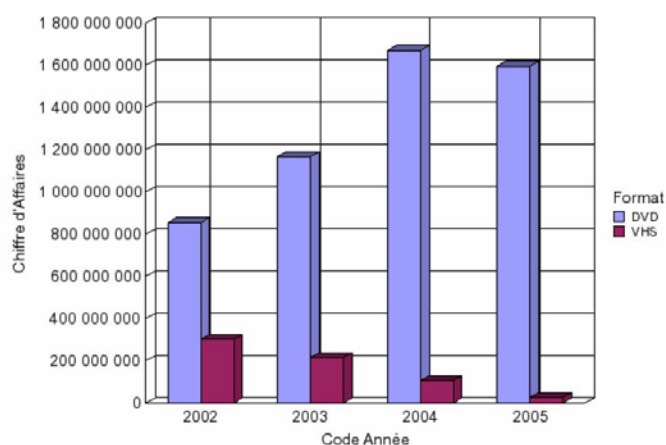
## 3.3. Long Tail et réduction de l'effet superstar

### 3.3.1 L'évolution des ventes de vidéo sur Internet et en magasin

Parallèlement au travail effectué plus haut sur la musique, nous étudions, ci dessous, les évolutions annuelles du marché de la vidéo de 2002 à 2005. Nous essaierons, en particulier de caractériser l'impact du développement de la vente vidéo en ligne sur la vente « traditionnelle ».

Pour mener à bien ce travail, nous avons choisi deux semaines particulières dans l'année : la 25ème (semaine « normale ») et la 51ème (semaine de Noël). Les données GfK nous donnaient accès à 3 années

seulement : 2003, 2004, 2005. Ces années sont néanmoins cruciales car elles correspondent aux années charnières pour la vidéo : développement d'Internet et extinction du VHS par rapport au DVD.



Evolution du Chiffre d'Affaires annuel VHS / DVD<sup>21</sup>

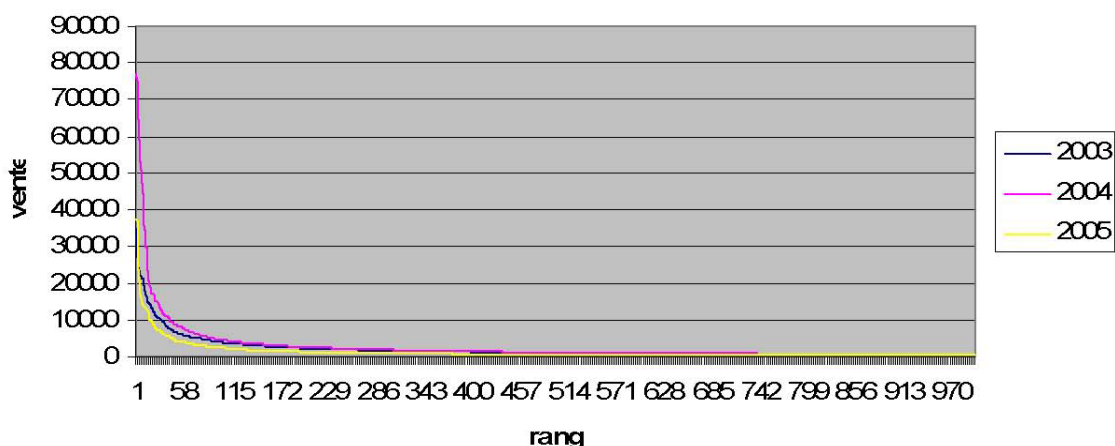
Afin de mesurer l'évolution de la structure des ventes au fil des ans, nous nous sommes d'abord focalisés sur les réseaux de distribution classiques. Le tableau suivant présente, pour les trois années 2003, 2004 et 2005, la synthèse des données correspondant au nombre de titres vendus hors Internet sur les semaines 25 (été) et 51 (hiver) ainsi que la proportion de vente correspondant aux meilleurs et moins bons 10 pourcents. Cette comparaison permet de situer l'évolution simultanée des best sellers et des low sellers, pendant les périodes de fêtes (où prédominaient, dans la musique, les marchés des hits) ou sur les périodes de creux ou « banalisées ».

	<i>S51 - Hiver</i>			<i>S21 - Printemps</i>		
	<i>Premier décile</i>	<i>Dernier décile</i>	<i>Effectif</i>	<i>Premier décile</i>	<i>Dernier décile</i>	<i>Effectif</i>
2003	86,09	0,07	2 986 944	82,75	0,20	724 148
2004	88,11	0,06	3 820 184	88,39	0,12	1 460 409
2005	83,8	0,12	2 079 583	91,25	0,13	1 727 918

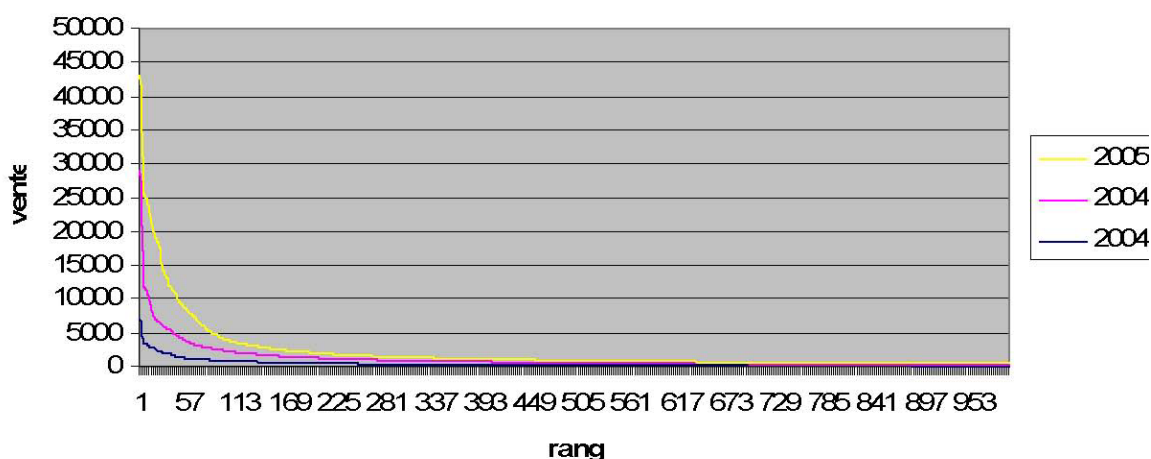
Les données du tableau sont reprises dans les graphes ci-dessous. Ils comparent les courbes de vente de titres en fonction du classement au printemps comme en hiver.

<sup>21</sup> Le tableau ci-dessus montre très clairement des données désormais classiques sur la diminution, au profit du DVD, de la part relative du VHS dans le marché de la vidéo. Ils indiquent pourtant également une légère diminution du volume des ventes de DVD, qui correspondent aussi au développement des ventes en ligne.

**Vente Hiver non Internet, en fonction du rang**



**Vente Eté non Internet, en fonction du rang**



La comparaison des différentes situations permet de pointer plusieurs phénomènes.

On note, comme attendu, l'importance des chiffres de vente sur la semaine de Noël, significativement plus importants que pour l'été. Par contre, il est intéressant de souligner qu'une tendance globale semble se dessiner d'une année sur l'autre : évolutions à la baisse pour les semaines importantes et à la hausse pour les semaines creuses. Les résultats semblent donc indiquer une tendance différente de celle observée pour la musique.

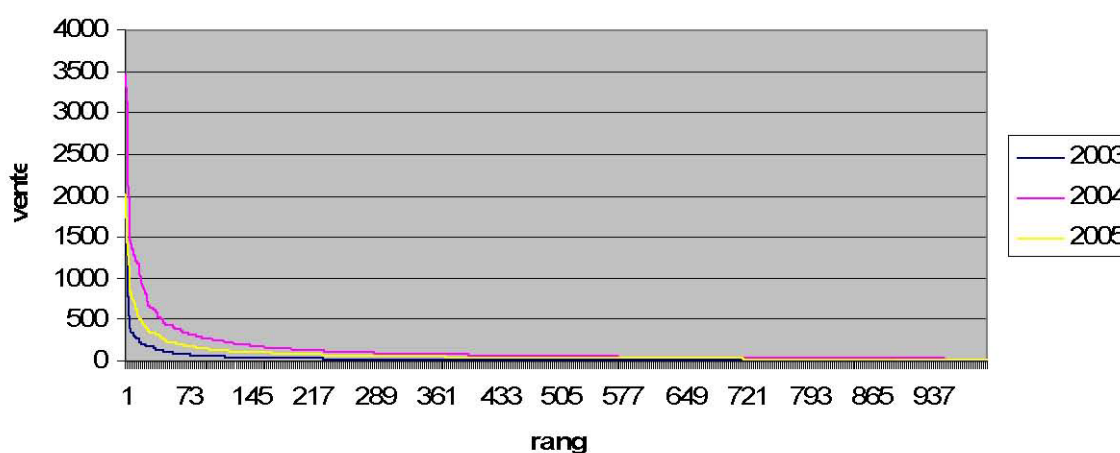
De façon plus générale, les dix premiers pourcents du marché de la vidéo hors Internet vendent plus de quatre-vingt pourcents des titres. En 2003, cette proportion était plus importante à Noël mais on observe que ce poids de l'effet superstar semble s'atténuer voire s'inverser au cours de l'année 2005 avec une forte augmentation en été et une diminution à Noël. Parallèlement, les dix derniers pourcents hors Internet vendent de l'ordre de 0,1 pourcent des titres. Cette proportion augmente à Noël sur les trois années et diminue pendant l'été, marquant une accentuation de l'effet Long Tail lors des périodes de plus fortes ventes et une diminution pendant les périodes creuses.

Après les circuits de distribution physique, nous avons considéré la distribution de DVD en ligne. De même que précédemment, nous avons regroupé, dans le tableau ci-dessous, les données synthétiques présentant les chiffres de ventes ainsi que les pourcentages de ventes du premier et du dernier décile,

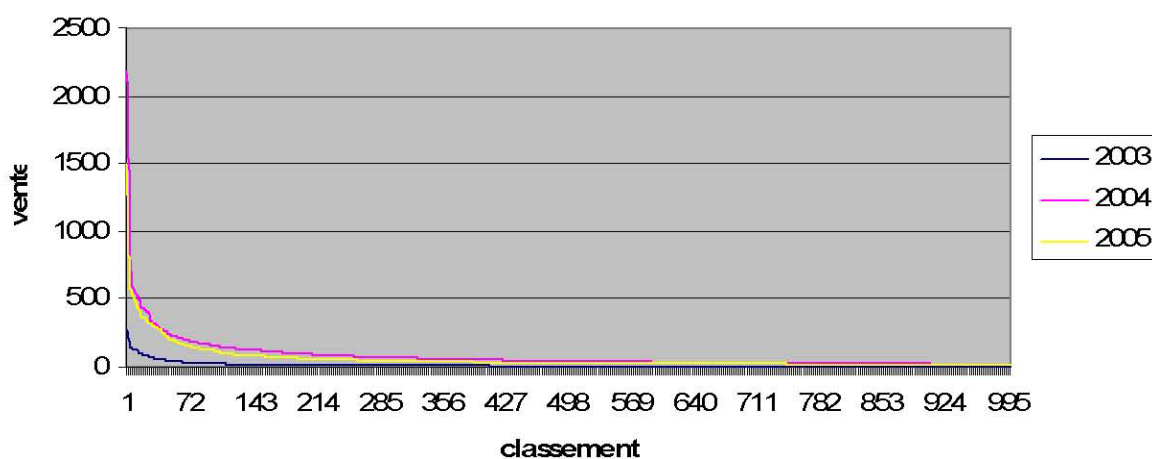
en hiver et en été , ainsi que les courbes de distribution des ventes en fonction du rang.

	<i>S51 - Hiver</i>			<i>S21 - Printemps</i>		
	<i>Premier décile</i>	<i>Dernier décile</i>	<i>Effectif</i>	<i>Premier décile</i>	<i>Dernier décile</i>	<i>Effectif</i>
2003	62,69	1,27	54 227	56,03	2,23	17 592
2004	72,18	0,66	200 560	74,60	0,63	126 010
2005	58 ,20	1,42	145 291	62,83	1,28	89 789

**Vente internet en Hiver en fonction du rang**



**Vente internet en Eté en fonction du rang**



Là encore, la comparaison permet de poser un certain nombre de constats.

On retrouve d'abord, bien entendu, un important écart entre le niveau des ventes effectuées via Internet et celui relevant des circuits de diffusion classique. Le rapport varie d'un facteur se situant selon les périodes entre dix et quarante.

En termes de saisonnalité, sur Internet comme hors internet, les ventes restent très importantes en période de fêtes : en l'occurrence plus du double entre la semaine de Noël et le milieu d'année.

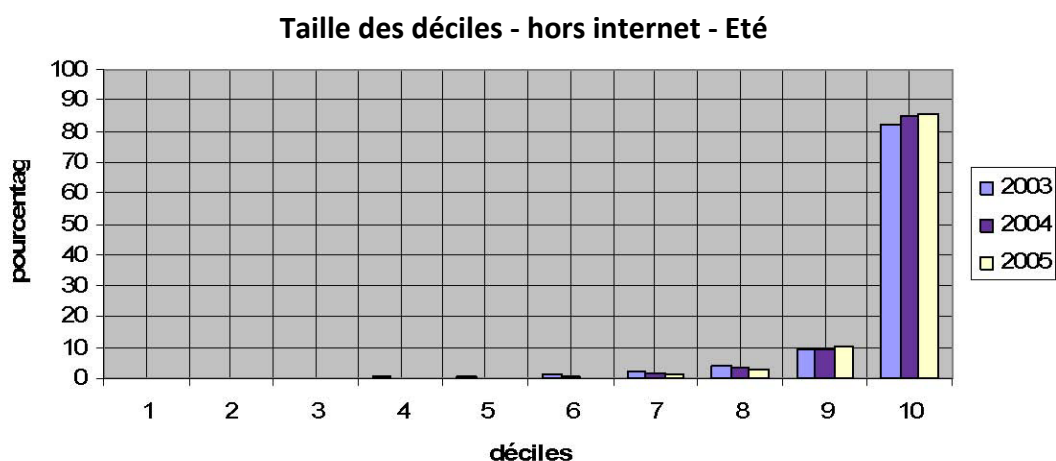
L'étude sur seulement trois années ne permet pas de dégager de tendances nettes sur l'évolution éventuelles des marchés de hits et de niches, aussi bien d'ailleurs pour l'hiver que pour l'été.

On peut néanmoins souligner des effets de séquence. Ainsi, il existe une nette corrélation entre l'augmentation des superstars et la baisse des dernières ventes. C'est tout particulièrement observable en 2004 : cette année est marquée par une forte présence des best sellers et par un effondrement simultanée des marchés de niches. Parallèlement, on note également, mais moins nettement, une tendance à la diminution de la part de marché des superstars en hiver, associée à une augmentation de l'effet Long Tail ; le mouvement exactement inverse s'opérant en été. Si les résultats sont intéressants à souligner, il convient toutefois de les prendre avec précaution car ils restent difficiles à interpréter ; les données utilisées ne permettent pas, en particulier, d'identifier et autonomiser les sorties spécifiques de titres très porteurs pouvant peser de manière significative, même ponctuelle, sur la structure des ventes.

### 3.3.2 Concentration et déconcentration du marché

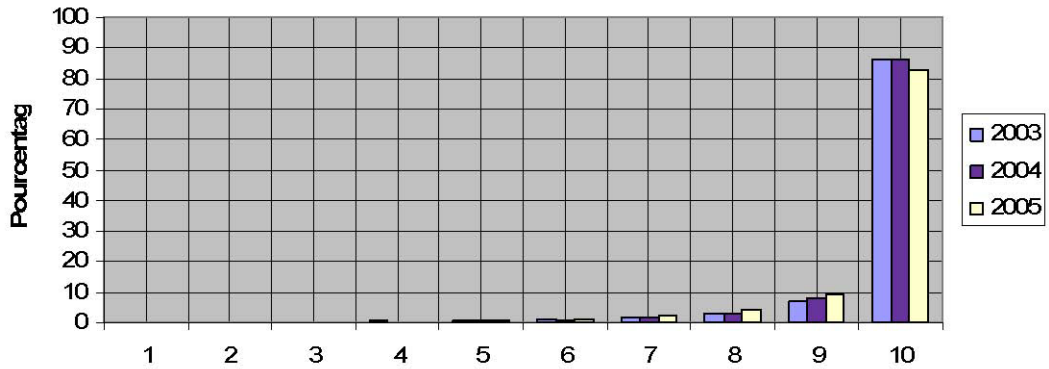
L'analyse présentée ci-dessous complète la précédente en étudiant une autre dimension de la distorsion temporelle de la distribution et du poids respectif des effets Long Tail et Superstar. Après avoir segmenté le marché en fonction du nombre de vidéos vendues, nous allons, dans les lignes qui suivent, nous concentrer plutôt sur l'évolution du nombre de vidéos correspondant au premier et au dernier déciles, en observant donc de manière spécifique le comportement des extrêmes (part respective des premiers et derniers déciles),

Nous présentons, dans les quatre diagrammes et tableaux ci-dessous, la taille de chaque décile, ainsi que les valeurs des deux premiers déciles et deux derniers déciles, hors internet et sur internet, pour les semaines 23 (été) et 51 (hiver).



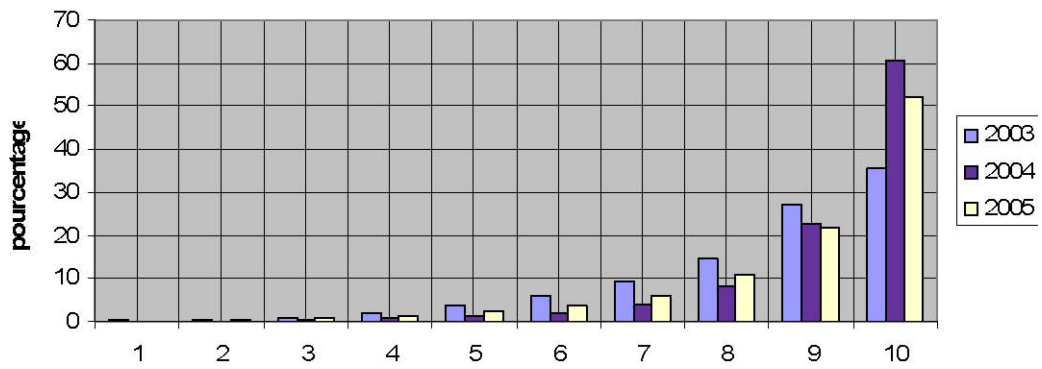
Année	2003	2004	2005
1 <sup>e</sup> décile	0,0139821	0	0
2 <sup>e</sup> décile	0,06291946	0,00590702	0,009238729
9 <sup>e</sup> décile	9,08836689	9,2444917	10,12102735
10 <sup>e</sup> décile	82,1308725	84,895741	85,76311899

### Taille des déciles - hors internet - Hiver



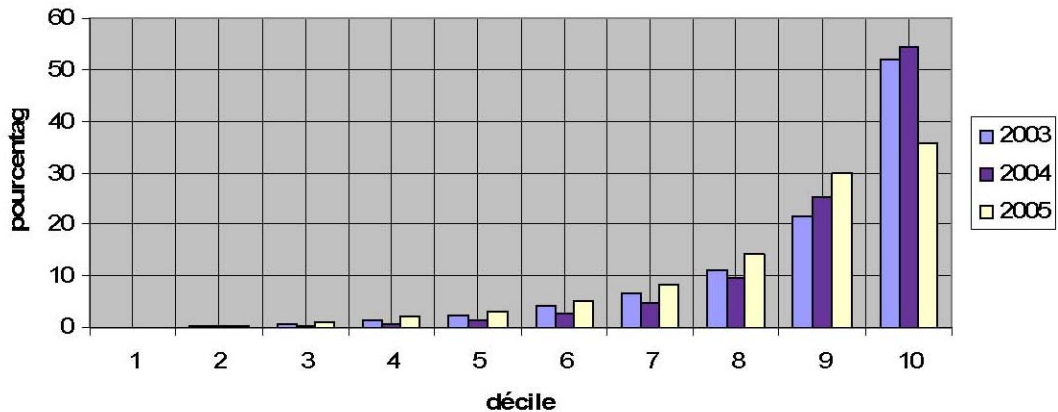
	décile		
Année	2003	2004	2005
1 <sup>e</sup> décile	0,03052969	0,0181373	0,016625795
2 <sup>e</sup> décile	0,09667735	0,03174027	0,058190282
9 <sup>e</sup> décile	6,71144355	8,04389226	9,044432437
10 <sup>e</sup> décile	86,2819926	85,9118527	82,78814581

### Taille des déciles - internet - Été







	décile		
Année	2003	2004	2005
1 <sup>e</sup> décile	0,25601639	0	0,157499394
2 <sup>e</sup> décile	0,51203277	0,07671653	0,339229465
9 <sup>e</sup> décile	27,2145417	22,5930188	21,81972377
10 <sup>e</sup> décile	35,6630824	60,6188467	52,21710686

### Taille des déciles - internet - Hiver



Année	2003	2004	2005
1 <sup>e</sup> décile	0,08704483	0,08589053	0,161244428
2 <sup>e</sup> décile	0,31916437	0,18739752	0,46476335
9 <sup>e</sup> décile	21,4420427	25,2205825	29,79227924
10 <sup>e</sup> décile	51,9077325	54,3687046	35,61604856

La comparaison, aux différentes périodes, de la taille des déciles et du volume correspondant des ventes permet d'avancer plusieurs résultats remarquables.

-  Lorsque la part du premier décile augmente, le marché tend à s'élargir vers les produits à vente plus faible : l'effet superstar diminue, ce qui correspond à une baisse de vente des meilleures vidéos.
-  Lorsque la part des derniers déciles diminue, le marché de niche se réduit et les petites ventes perdent de leur importance : globalement, les ventes sont plus importantes.
-  Dans les circuits de diffusion traditionnels, les ventes de vidéo sont très concentrées sur les Superstars (les deux premiers déciles concernent moins de un pourcent de toutes les vidéos en vente). Cet effet Superstar s'accroît encore pendant l'été ; cette période est en effet marquée par une concentration des ventes sur les bestsellers et un affaiblissement de la vente de « produits de niche » (donc une diminution de l'effet Long Tail). Pendant les périodes de fêtes correspondant aux pointes de vente, on relève par contre une légère diminution de l'effet Superstar, associée cette fois à un renforcement du marché des produits de niche, donc à une augmentation de l'effet Long Tail.
-  Dans les ventes de DVD sur Internet, le marché apparaît, globalement, moins concentré que dans les circuits traditionnels : le dernier décile ne représente plus en moyenne que 50% des vidéos possibles contre 80% dans le cas des marchés physiques. Globalement, l'effet Long Tail y est davantage manifeste, notamment dans les périodes de fêtes, où le pourcentage du dernier décile passe de 54% à 35%.

## **PARTIE 3 - CONCLUSION**

Au terme de ce travail, nous pouvons résumer ici les éléments de réponses obtenus aux hypothèses et questions posées au début de cette étude : plus spécifiquement quelle est l'évolution dans le temps du poids que représente les best sellers (superstar) et les queues de distribution (Long Tail) dans les marchés de contenus culturels que constituent la musique et la vidéo ?

### **1. Rappel des principaux résultats**

De manière générale, on a relevé une accentuation de la queue de distribution plus marquée sur Internet que sur les autres circuits de distribution et un effet Superstar plus important pour les ventes physiques. En un mot : les hits rapportent plus hors Internet que sur Internet.

De façon plus spécifique, le marché de la musique atteste d'une forte saisonnalité des ventes, non seulement en termes de volume des ventes mais aussi dans leur répartition et leur distribution. Les périodes connaissant les ventes les plus dynamiques (Noël et le printemps pour le marché physique et Noël uniquement pour le marché en ligne) sont caractérisées par un effet Superstar prononcé. Leur structure s'oppose à celles que l'on observe dans les périodes de ventes plus calmes (le reste de l'année) qui sont caractérisées par une part moins forte des produits best sellers et une plus forte répartition des ventes sur les produits de niche les moins vendus individuellement (ce que l'on tend souvent à qualifier d'allongement de la Tail).

La comparaison de la situation sur plusieurs années et à ces deux périodes spécifiques permet de raffiner le constat. Il semble se dégager notamment une tendance à l'augmentation de la Tail dans le temps aussi bien sur Internet ou sur les circuits de distribution classique. En revanche, on ne peut identifier aucune tendance nette quant à l'évolution de l'effet Superstar sur la période étudiée.

Sur le marché de la vidéo, l'étude de la saisonnalité met en évidence une *Tail* en général plus importante en hiver et moins importante en été pour les marchés traditionnels. Sur Internet en revanche, la long Tail apparaît davantage au printemps, et est minimale en été (effet intermédiaire à Noël). L'étude sur ces deux périodes entre 2003 et 2005 nous a permis de constater une augmentation de l'effet Long Tail à Noël et une augmentation de l'effet superstar en été.

### **2. Pareto déformée ou illusion de Long Tail ?**

Il nous semble nécessaire à ce stade de préciser un point méthodologique essentiel. Comme les résultats statistiques le montrent les distributions des ventes de CD<sup>22</sup> répondent à des schémas de Pareto, caractérisés par une prépondérance de la tête de distribution :

- périodes de Noël et été 2005 : premier décile constitué de 4 à 5 produits
- périodes hors Noël et été 2005 : premier décile constitué de 10 à 15 produits.

La part des différents quantiles dans les grandeurs annuelles cumulées nous apportent quelques informations supplémentaires :

---

<sup>22</sup> Le même constat est également applicable pour la vidéo



	10 %	25 %	50 %	90 %	95 %
2002	0,00 %	0,02 %	0,11 %	3,64 %	40,51 %
2003	0,01 %	0,04 %	0,12 %	5,97 %	45,71 %
2004	0,01 %	0,03 %	0,22 %	8,54 %	54,83 %
2005	0,01 %	0,04 %	0,23 %	8,68 %	54,03 %

On constate que quelle que soit l'année considérée, moins de 10 % des produits représentent plus de 90 % des ventes. On peut dès lors s'interroger pour savoir si le terme même de Long Tail (au sens normatif et non plus statistique) a encore sa justification alors que la contribution de la traîne en termes de ventes est à ce point négligeable ? Dans les meilleurs des cas, les tendances à l'étoffement identifiées plus haut restent proportionnellement faibles et semblent peu à même de constituer la base d'un véritable marché.

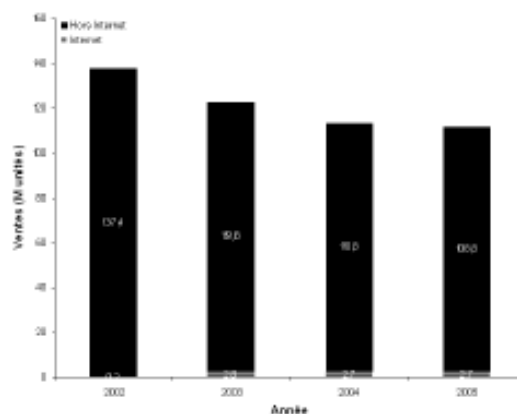
Le fait que les données étudiées concernent pour l'essentiel des biens physiques classiques distribués via les circuits traditionnels n'est sans doute pas la seule explication. Si structurellement, les traînes obtenues sont effectivement longues, elles ne sont pas suffisamment étoffées pour caractériser un phénomène de Long Tail : elles mériteraient plutôt d'être qualifiées de Pareto déformée. On peut en particulier avancer l'hypothèse que le phénomène tient essentiellement au fait que l'on agrège des chiffres concernant des marchés accessibles disjoints. Agréger ces marchés conduit à obtenir une offre composite majoritairement inaccessible pour le client considéré. C'est de ce décalage entre une offre en apparence large et une demande réellement étendue naît l'illusion de Long Tail.

Bien que moins prononcée, cette illusion apparaît également lorsque l'on isole les données internet, dû fait notamment de la faible maturité de la consommation des clients sur les plateformes de ventes en ligne de biens numériques.

### 3. Les limites méthodologiques : Exemple de l'accentuation dans le temps du phénomène de Long Tail

La Figure suivante présente l'évolution des ventes de biens musicaux en France pour les années entières étudiées, de 2002 à 2005.

Evolution des ventes de disques en France de 2002 à 2005 (en millions d'unités)



### Références par canal de distribution en France, de 2002 à 2005 (en millions d'unités)

	2002	2003	2004	2005	TCAM
Internet	38,3	89,1	116,2	120,2	+45,6 %
Hors Internet	420,7	425,1	535,2	516,8	+7,1 %

Ces données montrent une diminution claire des ventes de CD (albums et singles) en France, marquée par un recul des distributeurs traditionnels face aux distributeurs Internet. Étonnamment, la part de ces derniers reste très faible, alors que leurs catalogues s'enrichissent rapidement en nombre de titres.

Sur les semaines 9, 13 et 51 de l'année 2005, de nombreux indicateurs statistiques ont été calculés pour les canaux Internet et hors-Internet. On constate des valeurs de médianes similaires, alors que les moyennes diffèrent selon le canal et la période considérée. Nous avons, au début du travail, avancé une première explication : si les moyennes sont sensiblement plus élevées hors Internet, c'est probablement que sur Internet, le consommateur va tenter de trouver autre chose que la nouveauté qui est présente de manière forte dans les marchés physiques. Cette explication doit toutefois être relativisée.

D'une part, dans la base considérée, le canal Internet présente en 2005 près de cinq fois moins de références ; toutefois, même si l'ensemble des références disponible en ligne est moins nombreuse, elle reste supérieure à celle que peut matériellement proposer tout magasin physique.

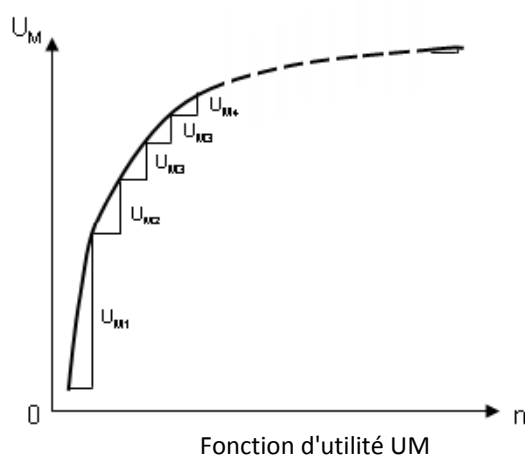
D'autre part la comparaison des données se heurte à l'écart important entre les volumes en jeu, qui rend difficile une stricte validation statistique : le rapport est ainsi d'un ordre 50 entre les canaux Internet et Hors Internet pour la semaine 51 de l'année 2005. Ces faibles volumes en jeu pour le canal Internet peuvent fausser l'interprétation du phénomène en contribuant à sous-valoriser l'impact des hits<sup>23</sup>. Cette distorsion tient au caractère très incomplet de la base de données pour le canal Internet. L'utilisation de la *skewness* (asymétrie) et de la *kurtosis* (aplatissement) fournissent cependant des bons indicateurs pour caractériser des distributions de type Long Tail. Leur application aux données traitées permet de confirmer l'aplatissement de la distribution significativement plus marqué hors des périodes de pic, marquant une queue de distribution plus épaisse.

Les observations précédentes sont intéressantes car elles contribuent à discuter les hypothèses de base sur laquelle repose la théorie de la longue traîne : un plus grand choix dans l'offre disponible implique plus grande variabilité de la consommation côté demande, et l'utilité moyenne tend à croître dans un contexte de marché marqué par une grande variété d'items disponibles.

Cette théorie s'inscrit dans des constats déjà anciens concernant la singularité de la fonction d'utilité observée dans l'économie des biens culturels. Elle tient dans la tension entre deux mouvements relevés dans la littérature. Le premier correspond à la décroissance, constatée empiriquement<sup>24</sup>, de la fonction d'utilité marginale standard, au fur et à mesure que le nombre de biens achetés croît pour une même période.

<sup>23</sup> Cf. comparaison des écarts-type

<sup>24</sup> Cf. J. Heilbrunn et al. (2001)



A l'inverse, les observations sur le caractère de biens d'expérience des produits culturels et de l'addiction rationnelle<sup>25</sup> montrent que sur un intervalle de temps suffisamment long, les appétits musicaux d'un consommateur peuvent évoluer sous l'effet conjugué d'évènements exogènes (influence de la publicité, découverte d'un artiste inconnu, recommandations, etc.) et endogènes (apprentissage, renforcement ou affaiblissement de goûts spécifiques). La question et le constat même de la Longue traîne remettent notamment en cause à la fois l'hypothèse d'une pente initiale forte et une convergence asymptotique rapide ; elle postule en outre une staticité qui correspond mal aux dynamiques relevées<sup>26</sup>.

#### 4. Modèles d'affaires de la Tail, étalement des prix et business models

Au-delà des données empiriques relevées dans les pages précédentes, il est essentiel de garder à l'esprit un facteur essentiel lié à la numérisation des contenus et au développement du téléchargement. Au cours du temps, le nombre de références disponibles et la longueur de la distribution, tendent, toutes choses égales par ailleurs, à s'étendre mécaniquement, et ce quels que soient les volumes annuels de création ou les stratégies des éditeurs.

En effet, du fait de coûts de stockage et de référencement quasi nuls, tout titre numérisé a désormais vocation à rester éternellement en catalogue. Il s'agit donc là d'une situation radicalement différente de celle observable dans les marchés physiques. En effet, dans ce dernier cas, sauf cas exceptionnels les titres ont une durée de vie limitée. Pour les titres qui se vendent, l'épuisement du tirage de départ n'appelle une réimpression que si le niveau de vente reste, en fin de vie, à un rythme suffisamment haut ; pour les titres ne rencontrant pas rapidement leur public et n'ayant que des faibles taux de rotation, les retours et coût de stockage conduisent à une mise au pilon et à un retrait rapide des bacs : ils sont irrémédiablement supprimés. Dans le monde physique, la conjonction de ces deux phénomènes conduisait donc à une auto-régulation du volume global des références disponibles : ce chiffre ne croissait que beaucoup modérément malgré un rythme annuel très soutenu des nouvelles créations.

En revanche, pour les marchés numériques, la quasi nullité des coûts de stockage et de maintien au catalogue fait que cette « mortalité naturelle » des titres n'existe plus. Les conséquences de cette

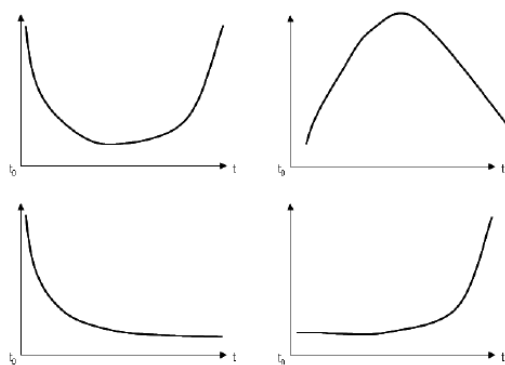
<sup>25</sup> G. Becker et K. Murphy (1998)

<sup>26</sup> C'est notamment ce que montrent G. Becker et G. Stigler (1977) lorsqu'ils mettent en évidence l'accroissement de la fonction d'utilité marginale standard dans le temps.

situation sont de plusieurs ordres. : elles touchent d'une part la structure et la configuration d'ensemble des ventes, elles pèsent d'autre part sur les stratégies et les modèles d'affaires susceptibles d'être développés dans la filière.

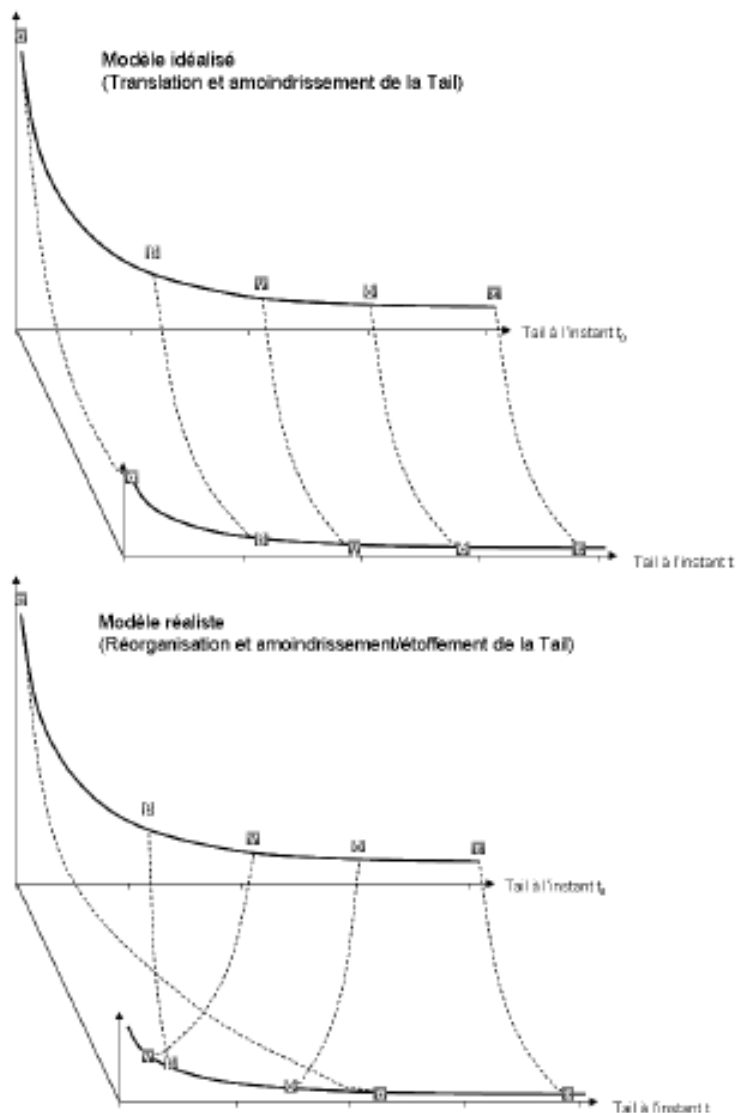
Au niveau de l'évolution de la configuration, l'absence de « mort naturelle » renforce, dans l'économie numérique, la capacité pour chaque titre de disposer d'un cycle de vie propre, que ce soit au sein de la traîne, où parmi les hits. Plusieurs configurations principales sont couramment évoquées par les responsables de marketing :

- les cycles [ sortie → oubli → redécouverte] , survenant notamment suite au décès d'un artiste (parfois appelés *smile*)
- les séquences [ croissance forte → oubli rapide] caractéristique de titres faisant l'objet d'hypermédiatisation et d'effets de mode (parfois appelés *smirks*);
- les titres connaissant un déclin naturel progressif (parfois appelés *bear trend*)
- la montée en puissance progressive, appuyés notamment par les effets de réputation, buzz ou bouche-à-oreille (parfois appelés *bull trend*);
- les ventes stabilisées que connaissent les œuvres majeures, anciennes, « de répertoire » ou, à l'inverse, les œuvres de queue de distribution (parfois appelés *flat*).



Ces évolutions spécifiques de chaque titre peuvent avoir des conséquences différentes sur la déformation d'ensemble de la distribution. Le schéma suivant présente un cas stylisé de cette situation, à partir du cas d'un ensemble ordonné [  $\alpha, \beta, \gamma, \delta, \varepsilon$  ] de titres pris à un instant  $t_0$  le long d'une courbe de distribution de type Long Tail.

Si l'on s'intéresse aux titres pris individuellement et que l'on en étudie la vie dans le temps, on note un déplacement de chacun d'entre eux. Alors que certains déclinent naturellement (e.g.  $\varepsilon$ ), d'autres au contraire voient leurs taux de rotation évoluer favorablement (e.g.  $\gamma$ ). Il s'ensuit un remaniement de la courbe dans sa composition : certains titres de niche deviennent ainsi des hits relatifs, alors que des hits vont se perdre dans la traîne. Les évolutions de distribution observables ne résultent donc pas seulement d'évolutions uniformes des taux de rotation mais tiennent aussi de la conjonction de mouvements individuels (et notamment la position relative des *best sellers*, *long sellers* et *low sellers*) ; cette conjonction contribue, dans l'exemple, à définir, à l'instant suivant  $t$ , une distribution globalement plus « tassée », marquée par une pente faible dans la short head et une traîne plus fine.



Du point de vue des modèles d'affaires et de la concurrence, la transformation de la structure des coûts associés à la commercialisation en ligne a aussi des effets notables. Les entreprises spécialisées dans la distribution traditionnelle de biens culturels (CD musicaux, vidéo, etc.) doivent faire face à une base de coûts fixes incompressibles ainsi qu'à des coûts variables (relation linéaire en première approximation). De façon traditionnelle, les entreprises en compétition essaient de construire et proposer des modèles plus efficaces en allégeant les coûts fixes associés à leur structure (moins de personnel, réduction des circuits et espaces de distribution, etc.) ou bien en élaborant des stratégies logistiques permettant de réduire les coûts variables (optimisation et économies d'envergure par réduction de la diversité, focalisation sur les best sellers, mutualisation et concentration et par exemple).

La situation est en revanche toute autre pour les marchés numériques en ligne. La constitution de positions concurrentielles est, dans ce cas, à la fois plus simple par rapport aux marchés physiques (du fait de la faible part des coûts variables liés à la distribution), et plus complexes puisque la structure de coûts ne permet pas une différenciation aisée. Ainsi, le coût marginal de diffusion est quasi nul pour l'ensemble des concurrents : pour un titre numérisé, ceux-ci ont en outre des structures de coûts unitaires très voisines pour la réplique, le stockage sur disque logique et la diffusion via les

plateformes online dédiée<sup>27</sup>. On observe simplement des structures de coûts en escalier : les paliers successifs sont associés à des effets de seuils liés à des réévaluations discrètes de coûts fixes (élargissement de la bande passante, achats d'espaces de stockage supplémentaires, etc.). Cette structure de coûts singulière permet – en théorie - aux distributeurs Long Tail de vendre des contenus à faible rotation, sans concession quant à leur rentabilité ; elle rend par contre difficile pour eux une réelle différenciation par les prix<sup>28</sup>.

Il faut malgré tout noter que si les coûts marginaux de la queue de distribution sont faibles, ils ne sont pas nuls et posent donc la question de leur amortissement (volume minimal de vente). Ces coûts relèvent de la constitution du catalogue, du stockage<sup>29</sup>, mais aussi de l'établissement et la fiabilisation de la base de données : numérisation, saisie et constitution des métadatas, normalisation des données associées aux fichiers<sup>30</sup>. La question reste encore secondaire, ou tout au moins acceptable, pour la vente de musique en ligne car les coûts de numérisation et de transferts y sont très faibles. Elle risque par contre de prendre toute son importance pour le téléchargement de vidéo car les coûts associés risquent d'y apparaître plus importants même s'ils restent faibles : encodé traditionnellement, un long-métrage est en effet 10 à 15 fois plus volumineux qu'un album musical, et 200 à 300 fois plus qu'un titre unique.

Certains diffuseurs ont ainsi arrêté la numérisation de leur catalogue faute d'un réel « effet longue traîne », malgré, dans certains cas, des efforts significatifs effectués sur leurs fonds de catalogue (remasterisation, édition en VO...). L'une des principales plateformes notait par exemple que sur les 2 millions de titres qu'elle met en ligne, au moins la moitié n'a jamais été téléchargée.

Les options ouvertes aujourd'hui sur la queue de distribution semble donc reposer sur le transfert d'une partie de ces coûts sur les auteurs ou les éditeurs (dans un modèle de type web 2.0<sup>31</sup>), soit sur une valorisation conçue globalement substituant à la rentabilité par titres - en vigueur dans le monde physique ou pour les hits - une appréhension globalisée du modèle d'affaires : valorisations publicitaires et média reposant sur l'agrégation des marchés permettant l'attractivité et d'une offre élargie, ou bien rémunération forfaitaire sur la base d'abonnements ou de la fourniture de services associés.

## **5. Recommandations pour une prolongation de l'étude sur des données de vente en ligne et téléchargement**

La partie qui suit a pour vocation de présenter la démarche d'analyse quantitative du phénomène Long Tail que nous suggérons d'adopter sur la base de données concernant les téléchargements musicaux sur les plateformes numériques des deux des principaux distributeurs online sur le marché français<sup>32</sup>.

---

<sup>27</sup> Même s'il ne l'explique pas complètement, ce constat est sans doute à mettre en relation avec le fait que les prix des produits culturels sont plus étalés sur les marchés traditionnels que sur Internet

<sup>28</sup> Elle empêche notamment les stratégies de soldeur « low cost » associés à des effets de déstockage tels que l'on peut les trouver chez des acteurs tels que CDiscount par exemple.

<sup>29</sup> 50 terabits de stockage, par exemple, pour une des plateformes numériques leaders en France

<sup>30</sup> Les coûts sont d'autant plus importants que la constitution des bases de données suppose des contributions humaines à plusieurs étapes du processus : d'une part, une proportion significative des indications transmises par les maisons de disques sont fausses ou marquent des incohérences entre support physique et numérique, d'autre part, la constitution de plateformes de vente demande un travail d'accompagnement « à la main », pour intégrer, par exemple les erreurs d'identification ou variations de titres ou d'orthographe effectuées par les acheteurs potentiels à la recherche d'un titre (les différentes manières d'écrire Tchaïkovski par exemple).

<sup>31</sup> Très proche de celui en vigueur dans le monde physique pour ou l'édition « à compte d'auteurs » par exemple

<sup>32</sup> Les contacts pris dans le cours de l'étude confirment la possibilité de disposer de telles données

Compte tenu des contraintes liées aux sources qui pourraient être mises à disposition par les distributeurs en ligne (pas d'accès direct, confidentialité des données notamment), deux options sont envisageables.

La première consiste à récupérer, pour chaque point d'analyse, les extraits de base utiles, si possible non anonymisés. La seconde requiert un pré-traitement des données en extrayant, à chaque fois que cela est requis, les informations souhaitées (généralement chiffres de ventes) correspondant aux bornes des 10<sup>e</sup> (voire 100<sup>e</sup>) de centiles logarithmiques.

Si elle présente l'inconvénient d'être approximative, la deuxième méthode nous paraît la plus réalisable d'un point de vue opérationnel, dans la mesure où elle nous permettrait d'obtenir des bases d'analyses moins lourdes en termes de traitement. Pour autant, elle nécessite une phase de calibrage préalable, nécessaire à la détermination du meilleur arbitrage entre précision et praticité.

Idéalement, il faudrait que l'analyse puisse être menée :

- indépendamment sur les bases « titres » et « albums »<sup>33</sup> et ce pour chacune des années,
- sur les données cumulées de toutes les années disponibles
- sur les différents canaux exploitables (contenus dématérialisés online, contenus physiques online et contenus physiques off line)
- dans le cadre de subdivisions temporelles le plus fines possible (si possible quotidiennes).

Les difficultés de traitement de l'information de traîne rencontrées sur la base GfK ont mis en lumière un paradoxe opératoire intéressant : par définition, la Long Tail est constituée par un grand nombre d'items uniques, et plus ce nombre est important, plus la Long Tail se révèle. Concrètement, cela signifie que les tailles des bases de données à manipuler doivent être élevées pour que le phénomène puisse être étudié de façon suffisamment précise, complexifiant d'autant l'élaboration statistique.

Une étape de pré-traitement permettrait de remédier à ce problème de la taille des bases d'étude. Elle permettrait d'abord d'établir une partition en quantiles logarithmiques afin d'obtenir une analyse plus précise pour les premiers points de la Tail + interpolation spline à partir de l'échantillon de point obtenu, afin de reconstituer la courbe originelle.

Les modèles de régression par quantile permettront de caractériser et de distinguer les effets de différents paramètres (influence du genre, de l'ancienneté, du format, etc.) sur les principaux déterminants de la Long Tail (taille de la traîne, ratio surface de hit/miss, indicateurs de diversité etc.). Les tests de statistiques descriptives trouvent toutefois vite leurs limites dans la mesure où ils ne permettent pas d'expliquer les variations qu'ils mettent en évidence. Ainsi, il n'est pas spontanément

---

<sup>33</sup> La distinction des données qui ont trait aux albums, de celles qui concernent des titres uniques apparaît particulièrement importante, notamment pour appréhender l'évolution des pratiques de consommation et les mettre en regard avec les stratégies d'éditeurs. En effet, même si les albums physiques sont pour certains déclinés en singles, ils ne le sont jamais que très partiellement. Aujourd'hui toutes les plateformes de téléchargement musicales offrent aux consommateurs le choix entre des albums et leurs titres pris séparément. Ainsi dépecé, l'album apparaît de plus en plus comme un véhicule promotionnel. En effet, sous forme dématérialisée, l'objet « album » n'existe plus vraiment en tant que tel, mais simplement en tant que bundle de titres, dont la principale valeur ajoutée n'est autre que le prix. Il s'ensuit un glissement terminologique intéressant : alors que dans le modèle physique traditionnel, on considère qu'un titre est extrait d'un album, le rapport se renverse dans les modèles de biens dématérialisés, où l'on dira que ce sont les titres qui constituent l'album. L'accentuation de l'effet « produit star » - plus marqué pour les titres uniques que pour les « albums » - conduit à s'interroger sur l'accentuation éventuelle de la concentration sur les titres et une plus grande volatilité pour les albums.

possible de savoir si l'évolution dans le temps des indicateurs de diversité est due à un effet Long Tail plus prononcé, ou à une évolution de la composition des items constitutifs de la distribution étudiée.

Les phases d'analyses statistiques descriptives devront donc pouvoir être complétées par des traitements économétriques plus précis sur des extraits de bases complets<sup>34</sup>.

## 5.1. Fiche technique

### *Présentation de la base et des grandes masses – indicateurs clés*

- Taille de la base : nombre de titres/albums uniques disponibles
- Evolution annuelle des nombres de références et des volumes cumulés de ventes
- Chiffres de ventes aux points-bornes des fractions de centiles logarithmiques pour chacune des années, et pour l'ensemble des années cumulées
- Contribution des différents déciles (centiles) : quels pourcentages des ventes sont respectivement réalisés par les 10 (100) premiers pourcents de titres/albums les plus vendus, les 10 (100) suivants, etc.
- Test du *98 percent rule* : quel pourcentage du catalogue se vend au moins une fois par an, trimestre, mois, semaine
- Ancienneté moyenne et écart-type de chaque centile

### *Option pour l'interpolation*

- Chiffres de ventes aux bornes des 100<sup>e</sup> ou 10<sup>e</sup> de centiles logarithmiques (soient 10 001 ou 1 001 points) par semaine pour une année donnée, soient 52 x nombre d'années disponibles
  - o Idem pour les ventes annuelles par genre
- Ancienneté moyenne et écart-type des anciennetés (par rapport à la date de sortie) sur les différents centièmes de centiles logarithmiques des courbes annuelles
- Pour chaque album situé aux bornes de chaque 10<sup>e</sup> de centile logarithmique, chiffres de ventes de l'album et de chacun de ses titres uniques

### *Influence du facteur genre*

- Nombre de titres disponibles par genre et année de production
- Pour chaque genre, chiffres de ventes aux points-bornes des fractions de centiles logarithmiques pour chacune des années, et pour l'ensemble des années cumulées
- Pour chaque année (+ années cumulées), genre le plus/moins représenté par centile (à défaut, décile) ou mieux : contribution de chaque genre à chaque décile
- Focalisation et analyse complète des données d'un genre déterminé (caractérisation des phénomènes de *nichebusters*)
  - o Analyse spécifique de la niche « Bossa Nova » par exemple

### *Saisonnalité de la Tail et structure de la distribution*

- Nombre de ventes par semaine pour chaque année, + années cumulées
- Queue de distribution

---

<sup>34</sup> On peut par exemple souligner l'importance de pouvoir prendre en compte, dans les données extraites, les comportements d'achats de consommateurs : fréquence d'achat, variété de choix. Le traitement de ces données permettrait de mesurer dans quelle mesure les produits de niche sont-ils achetés par tous les consommateurs ou principalement par quelques spécialistes, ainsi, à l'inverse, que la spécificité des comportements d'achats des spécialistes : limités aux produits de niche ou comportant une part représentative de hits p



- Pour chaque année (+cumul), caractérisation du dernier décile (albums/ titres) : moyenne des ventes, écart-type des ventes, nombre de titres par genre
- Pour chaque année, nombre d'invendus
- Déformation de la short head
  - Pour chaque année, nombre de ventes et évolutions annuelles de titres/albums sélectionnés
    - pour l'année 1, relever la date de sortie et le nombre de ventes de titres/albums situés aux bornes des intervalles [0-10 %], [10-20 %] ... [90-100 %]
      - ex. : si 50 titres différents ont été vendus, relever les chiffres de vente du premier, 5e,... 45e, dernier
    - Relever les chiffres de ventes de ces titres pour chacune des années suivantes, ainsi que les positions correspondantes.
    - Répéter l'opération pour l'année 2 comme année de référence
  - Pour une semaine donnée de la première année disponible (par ex. : S5 1<sup>ere</sup> semaine de février), nombre de ventes de chacun des titres/albums disponibles, puis chiffres de ventes et positions de ces titres pour la même semaine de chacune des années suivantes/

#### *Rapport albums/titres*

- Pour chaque album situé aux bornes de chaque centile logarithmique, fournir les chiffres de ventes de l'album, ainsi que de chacun de ses titres uniques
- Nombre moyen de titres uniques téléchargés par centile d'album (prendre la Tail d'une année et donner le nombre moyen de titres correspondants téléchargés, ainsi que l'écart-type)

## Références bibliographiques

- Anderson C. (2004). The Long Tail. *Wired Magazine*, 12.10, [www.wired.com/wired/archive/12.10/Tail.html](http://www.wired.com/wired/archive/12.10/Tail.html)
- Anderson C. (2006). *The Long Tail: Why The Future of Business Is Selling Less of More*. New York, NY: Hyperion.
- Bakos, Y. et Brynjolfsson, E. (2000). Bundling and Competition on the Internet. *Marketing Science* 19 (1), 63-82.
- Becker G. et Murphy K., "A Theory of Rational Addiction", *The Journal of Political Economy*, 1988, 96, p. 675 – 700
- Becker G. et Stigler G.J. (1977). "De Gustibus Non Est Disputandum", *The American Economic Review*, 67: 76-90.
- Benghozi P.-J., Paris T., \_ L'industrie de la Musique à l'Âge Internet, *Gestion 2000*, 2001, 2, pp. 41-60.
- Benghozi, P.-J. et T. Paris (2005b). The Economics and Business Models of Prescription in the Internet. *Internet Economics*. E. Brousseau et N. Curien, Cambridge University Press.
- Benghozi P.-J. et Paris T., (2005a). Analysing the distribution to understanding the markets of cultural goods, *8th International Conference on Arts & Cultural Management*, HEC Montreal.
- Benhamou, F. (2002). *L'économie du star system*. Paris, Odile Jacob.
- Benhamou F. (2004). *L'économie de la culture*, Paris : Éd. de la Découverte.
- Brynjolfsson E., Yu J. H., et M.D. Smith (2006). From Niches to Riches: Anatomy of the Long Tail. *MIT Sloan Management Review* 47(4), 67-71.
- Brynjolfsson E., Yu J. H., et M.D. Smith (2003), Consumer Surplus in the Digital Economy : Estimating the Value of Increased Product Variety at Online Booksellers, *Management Science*, 49, 1580-1596.
- Chantepie P. (2005). *Révolution numérique et industries culturelles*, Paris, Éd. de la Découverte.
- Chkravarti I.M., Laha R.G. et Roy J. (1967), *Handbook of Methods of Applied Statistics*, John Wiley and Sons, Volume I, pp. 392-394.
- Crasneanu M. (2007), To peer or not to peer : the economic revolution of cultural products, *Elenbi Strategic Review* .
- Dang Nguyen G. et Mevel O. (2007), Nouvelle et ancienne économie : vers une intégration réussie, *Revue Française de Gestion*, N° 173/4, 113-130
- Elberse A. et Oberholzer-Gee F. (2007), Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales, *Harvard Business School Working Paper Series*, No. 07-015
- Faulkner R, & Anderson A. 1987. Short-Term Projects and Emergent Careers: Evidence from Hollywood. *American Journal of Sociology*, 92 (4): 879-909
- Heilbrunn J. et Graay C.M. (2001), *The Economics of Art and Culture*, Cambridge, Cambridge University Press.
- Mason T. (2004) Quantifying Quality and Other problems, in *Economics of Art and Culture*, Victor A Ginsburgh (ed.), Amsterdam : Elsevier.
- Peitz M. et Waelbroeck P. (2003) Piracy of Digital Products: A Critical Review of the Economic Literature, *Cesifo Working Paper* no. 1071 Category 9: Industrial Organisation, November.
- Roland Berger Strategy Consulting, (2005), *La Long Tail*, Document de travail.
- Rosen, S. (1981). "The Economics of Superstars." *American Economic Review* 71: 845-58.
- Shirky C., (2003), Power Laws, *Weblogs and Inequality*, Document de travail.
- Throsby D. (2001), *Economics and Culture*, New York : Cambridge University Press.
- Shapiro, C. and H. R. Varian (1999). *Economie de l'information*. Paris, De Boeck Université.
- Vogel, H. (1998). *Entertainment Industry Economics*. Cambridge, Cambridge University Press.
- Wolf, M. J. (1999). *The Entertainment Economy: How Mega-media Forces are Transforming our Lives*. New York, Times Books.
- Zhang, M.X., 2002, "Stardom, Peer-to-Peer and the Socially Optimal Distribution of Music", *mimeo*, Sloan School of Management, MIT.

## Quelques blogs et liens portant sur la longue traîne

- X *MySpace et le long Tail* : <http://blogs.zdnet.fr/index.php/2006/09/07/myspace-et-le-mirage-du-Long-Tail/>
- X *Le marché des noms de domaine et la longue traîne* : <http://adscriptum.blogspot.com/2006/07/les-noms-de-domaine-et-la-longue-04>
- X *Economie du e-commerce* : <http://www.elanceur.org/Traductions/LaLongueQueue.html>
- X *Blog sur la longue traîne* : <http://louvre-boite.viabloga.com/news/la-longue-queue>
- X *Le filtrage du long Tail* : <http://blog-fr.criteo.com/archive/2006/09/18/le-Long-Tail-a-besoin-de-filtrage.html>
- X *The long Tail* : <http://longTail.typepad.com/>

# Annexes

*A.1. Distribution par décile du marché de la musique au cours de l'année 2005*

*A.2. Caractérisation statistique comparée des ventes de musique : chiffre d'affaires et volumes, sur et hors internet, semaines 9, 13 et 51*

*A.3. Représentation de la tête de distribution : Top 10 des ventes de CD sur et hors Internet*

*A.4. Représentation de la queue de distribution : ventes de CD sur et hors Internet*

*A.5. Données GfK – Le logiciel de consultation fourni en ligne*

*A.6. Rappels statistiques et techniques*



<b>Déciles</b>										
<b>Date</b>	1	2	3	4	5	6	7	8	9	10
20055201	24 (5083,9)	90 (10167,8)	244 (15254,7)	515 (20335,6)	957 (25419,5)	1726 (30503,4)	3062 (35587,3)	5298 (40671,2)	9703 (45755,1)	14224 (50839)
20055201	0,168728909	0,632739408	1,715410574	3,62064117	6,728065242	12,1344207	21,52659863	37,24690664	68,21569179	100
20055201	0,168728909	0,464004489	1,082677165	1,905230596	3,107424072	5,406355456	9,392575928	15,71991001	30,96878515	31,79430621
20055209	13 (6546,7)	57 (11093,4)	149 (16640,1)	343 (22186,8)	730 (27733,5)	1439 (33280,2)	2738 (38825,9)	4888 (44373,6)	9450 (49920,3)	14580 (55467)
20055209	0,089163237	0,30946502	1,021947874	2,352537723	5,008858711	9,865684499	18,77914952	34,21124629	61,81481481	100
20055209	0,089163237	0,301783265	0,631001372	1,330589849	2,654320988	4,882825789	8,909465021	15,43209877	30,60356653	35,18518519
20055213	2 (5291)	24 (9127)	86 (13728)	222 (18236)	502 (22809)	1058 (27371)	2144 (31928)	4064 (38488)	8017 (41049)	12250 (45610)
20055213	0,016326531	0,195918367	0,702040816	1,812244898	4,097959184	8,636734694	17,50204082	33,1755102	65,44489796	100
20055213	0,016326531	0,179591837	0,506122449	1,110204082	2,285714286	4,53877551	8,865306122	15,67346939	32,26938776	34,55510204
20055227	28 (5083,9)	95 (10167,8)	269 (15254,7)	625 (20335,6)	957 (25419,5)	1726 (30503,4)	3062 (35587,3)	5298 (40671,2)	9703 (45755,1)	14224 (50839)
20055227	0,196650394	0,667885284	1,891189854	4,393892002	6,728065242	12,1344207	21,526989663	37,24690664	68,21569179	100
20055236	0,196650394	0,471034871	1,223284589	2,502812148	2,33408324	5,406355456	9,392575928	15,71991001	30,96878515	31,79430621
20055236	0,039348391	0,306917447	1,062406548	2,722300653	5,902258698	11,35594654	20,72086252	36,18478004	66,26268986	100
20055236	0,039348391	0,267569056	0,7554891	1,660502085	3,179349965	5,453686844	9,364916975	15,46391753	30,07790981	33,73731014
20055249	12 (9571,4)	41 (19142,8)	104 (28714,2)	240 (38285,6)	532 (47857)	1145 (5428,4)	2385 (66998,8)	4769 (76571)	9992 (86142)	18637 (95714)
20055249	0,064389045	0,219992488	0,558029726	1,287760906	2,854536674	6,143692654	12,797124	25,58888233	55,61377904	100
20055249	0,064389045	0,155604443	0,338037238	0,72973118	1,566775769	3,28915598	6,653431346	12,79175833	28,02489671	46,38622086
20055251	10 (8650)	38 (18351)	104 (27524)	251 (36729)	552 (45938)	1166 (55123)	2405 (64313)	4726 (73499)	9597 (82687)	18094 (91874)
20055251	0,055269399	0,210014369	0,574776169	1,387200177	3,05073505	6,444125124	13,29160981	26,11915552	53,03968166	100
20055251	0,055269398	0,15474743	0,364761799	0,812424008	1,663634873	3,393390074	6,847573781	12,82745662	26,92052614	46,96031834

## ***A2 Caractérisation statistique comparée des ventes de musique : chiffre d'affaires et volumes, sur et hors internet, semaines 9, 13 et 51***

Semaine 20055209

CA Internet

CA Hors Internet

Nombre	14580
Moyenne	56,39
Médiane	24
Écart moyen	56,39
Écart type	314,11
Variance	98666,75
Coefficient de variation	557%
Asymétrie	50,34
Aplatissement	3664,00
Minimum	0
Maximum	26318
Étendue	26318
Premier quartile	15
Troisième quartile	41
Int. semi-interquartile	13
Premier décile	10
Neuvième décile	82

Nombre	59613
Moyenne	500,70
Médiane	37
Écart moyen	797,43
Écart type	12607,22
Variance	158941953,96
Coefficient de variation	2518%
Asymétrie	168,78
Aplatissement	35045,11
Minimum	0
Maximum	2692124
Étendue	2692124
Premier quartile	20
Troisième quartile	106
Int. semi-interquartile	43
Premier décile	13
Neuvième décile	356

Semaine 20055213

CA Internet

CA Hors Internet

Nombre	12250
Moyenne	58,32
Médiane	22
Écart moyen	62,72
Écart type	916,13
Variance	839291,51
Coefficient de variation	1571%
Asymétrie	101,16
Aplatissement	10786,15
Minimum	0
Maximum	98252
Étendue	98252
Premier quartile	15
Troisième quartile	36
Int. semi-interquartile	10,5
Premier décile	9
Neuvième décile	75

Nombre	55630
Moyenne	371,33
Médiane	34
Écart moyen	571,53
Écart type	5158,16
Variance	26606580,65
Coefficient de variation	1389%
Asymétrie	72,89
Aplatissement	8431,65
Minimum	0
Maximum	733242
Étendue	733242
Premier quartile	19
Troisième quartile	95
Int. semi-interquartile	38
Premier décile	13
Neuvième décile	312

Semaine 20055251

CA Internet

CA Hors Internet

Nombre	18093
Moyenne	95,85
Médiane	25
Écart moyen	115,70
Écart type	1115,03
Variance	1243289,73
Coefficient de variation	1163%
Asymétrie	101,02
Aplatissement	12067,57
Minimum	0
Maximum	135556
Étendue	135556
Premier quartile	16
Troisième quartile	51
Int. semi-interquartile	17,5
Premier décile	12
Neuvième décile	125

Nombre	65531
Moyenne	1158,55
Médiane	47
Écart moyen	1930,46
Écart type	21241,12
Variance	451185020,10
Coefficient de variation	1833%
Asymétrie	64,02
Aplatissement	5883,36
Minimum	0
Maximum	2668621
Étendue	2668621
Premier quartile	21
Troisième quartile	175
Int. semi-interquartile	77
Premier décile	14
Neuvième décile	685

## Semaine 20055209

## Ventes Internet

Nombre	14580
Moyenne	3,80
Médiane	1,1
Écart moyen	3,93
Écart type	18,11
Variance	327,82
Coefficient de variation	476%
Asymétrie	35,11
Aplatissement	1813,78
Minimum	1
Maximum	1202
Étendue	1201
Premier quartile	1,1
Troisième quartile	2,2
Int. semi-interquartile	0,55
Premier décile	1,1
Neuvième décile	5,5

## Ventes Hors Internet

Nombre	59616
Moyenne	48,76
Médiane	2
Écart moyen	82,30
Écart type	879,69
Variance	773846,50
Coefficient de variation	1804%
Asymétrie	68,05
Aplatissement	7458,43
Minimum	1
Maximum	122205,12
Étendue	122204,12
Premier quartile	1
Troisième quartile	7
Int. semi-interquartile	3
Premier décile	1
Neuvième décile	23,8

## Semaine 20055213

## Ventes Internet

Nombre	12250
Moyenne	3,72
Médiane	1,1
Écart moyen	3,94
Écart type	42,58
Variance	1813,23
Coefficient de variation	1144%
Asymétrie	96,94
Aplatissement	10163,53
Minimum	1
Maximum	4500,1
Étendue	4499,1
Premier quartile	1,1
Troisième quartile	2,2
Int. semi-interquartile	0,55
Premier décile	1,1
Neuvième décile	5,4

## Ventes Hors Internet

Nombre	55632
Moyenne	31,71
Médiane	2
Écart moyen	51,45
Écart type	588,41
Variance	346227,07
Coefficient de variation	1855%
Asymétrie	108,42
Aplatissement	17000,14
Minimum	1
Maximum	102207,64
Étendue	102206,64
Premier quartile	1
Troisième quartile	6
Int. semi-interquartile	2,5
Premier décile	1
Neuvième décile	20

## Semaine 20055251

## Ventes Internet

Nombre	18093
Moyenne	5,08
Médiane	1,12
Écart moyen	5,77
Écart type	28,79
Variance	829,00
Coefficient de variation	567%
Asymétrie	28,00
Aplatissement	1070,82
Minimum	1
Maximum	1583,4
Étendue	1582,4
Premier quartile	1,12
Troisième quartile	3,24
Int. semi-interquartile	1,06
Premier décile	1
Neuvième décile	6,72

## Ventes Hors Internet

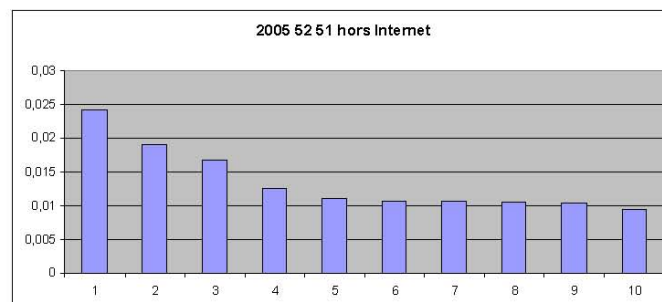
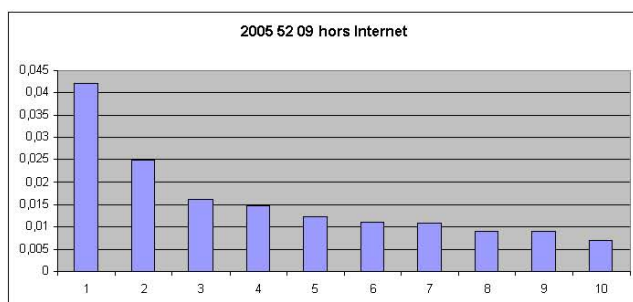
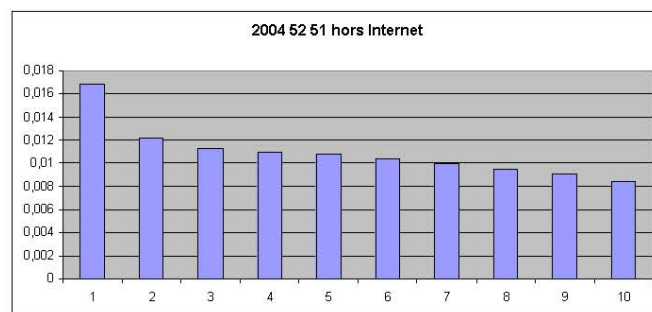
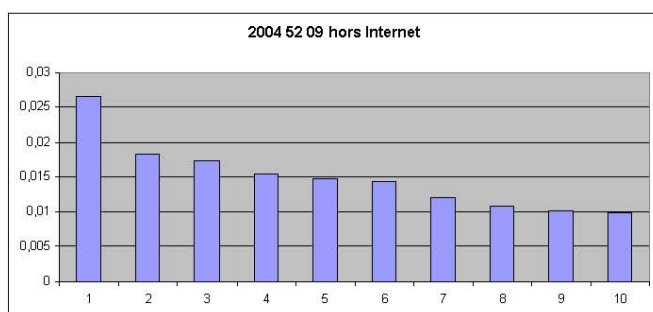
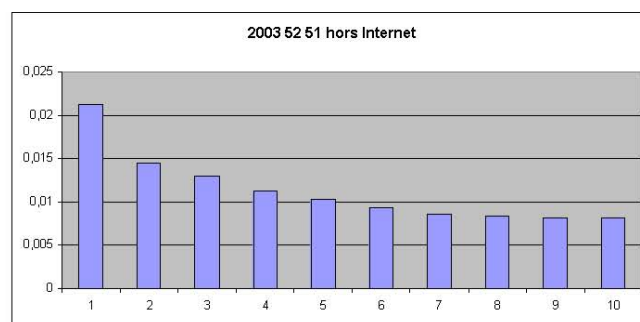
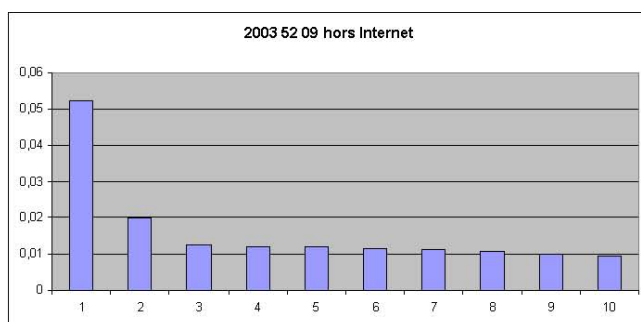
Nombre	65534
Moyenne	72,50
Médiane	3
Écart moyen	121,47
Écart type	1230,70
Variance	1514614,42
Coefficient de variation	1697%
Asymétrie	49,68
Aplatissement	3458,46
Minimum	1
Maximum	126606,86
Étendue	126605,86
Premier quartile	1
Troisième quartile	10,43
Int. semi-interquartile	4,715
Premier décile	1
Neuvième décile	40

### A.3. Représentation de la tête de distribution : top 10 des ventes de CD sur et hors Internet

Les graphes suivants présentent le pourcentage des ventes (en volume) réalisé par chacun des dix titres les plus vendeurs.

Internet	Semaine 09	Semaine 51	Hors Internet	Semaine 09	Semaine 51
2003	15.25%	8.35%	2003	16.20%	11.29%
2004	8.47%	9.77%	2004	15.00%	10.91%
2005	9.57%	10.13%	2005	15.70%	13.52%

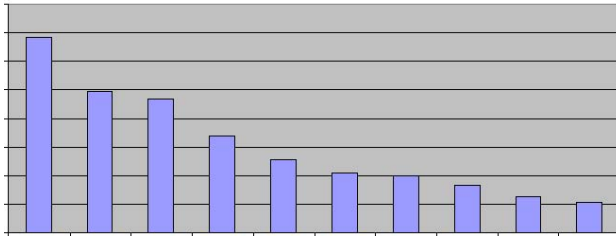
#### Marchés traditionnels



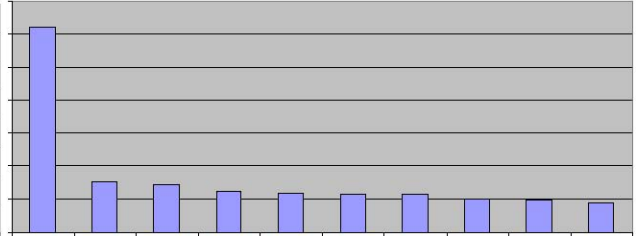


# Internet

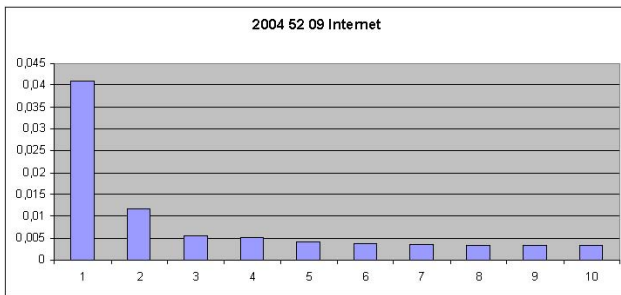
2003 52 09 Internet



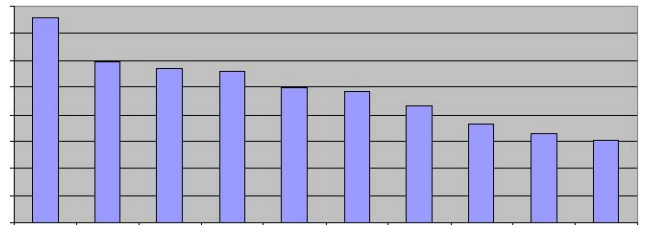
2003 52 51 Internet



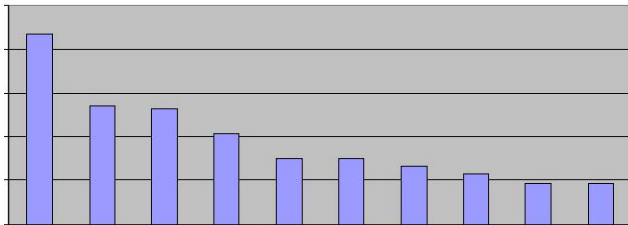
2004 52 09 Internet



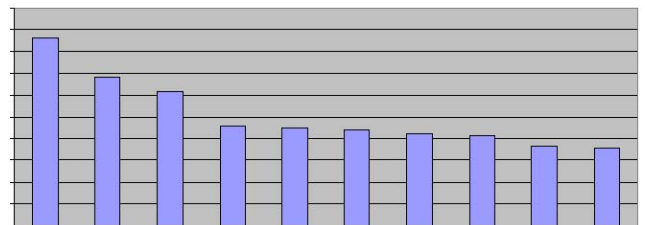
2004 52 51 Internet



2005 52 09 Internet



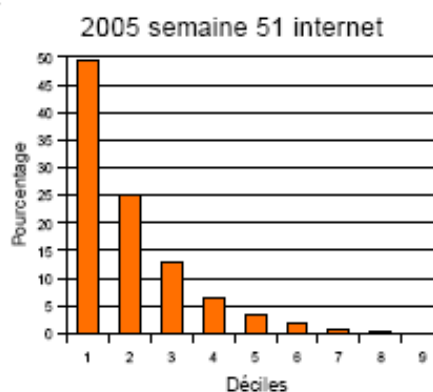
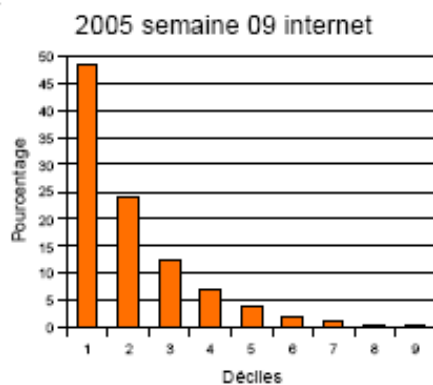
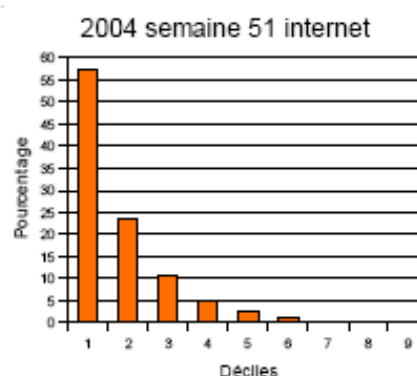
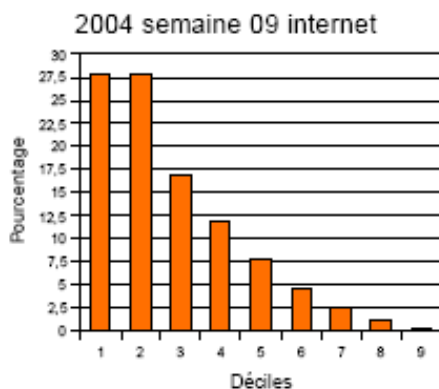
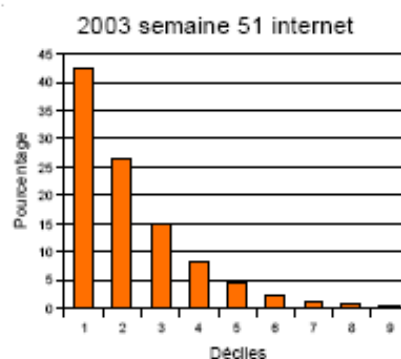
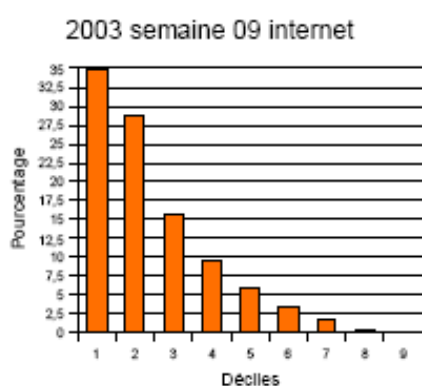
2005 52 51 Internet



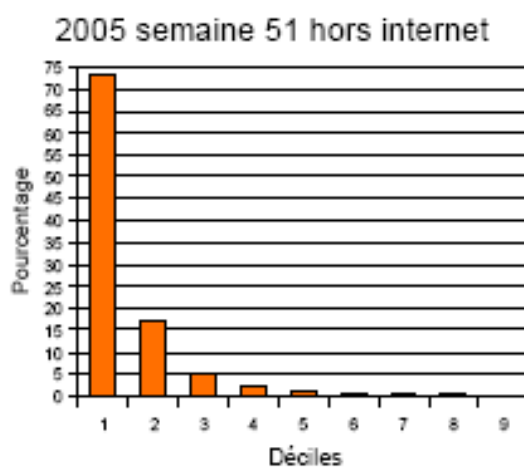
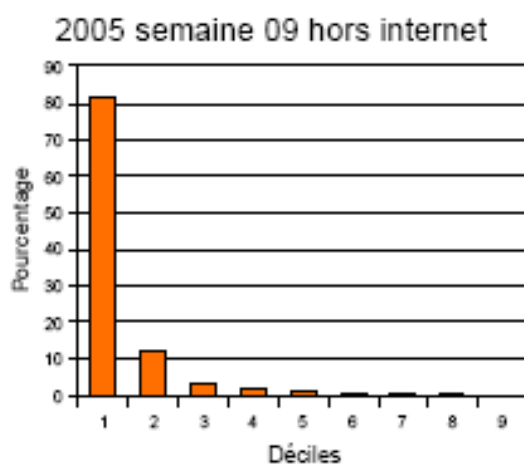
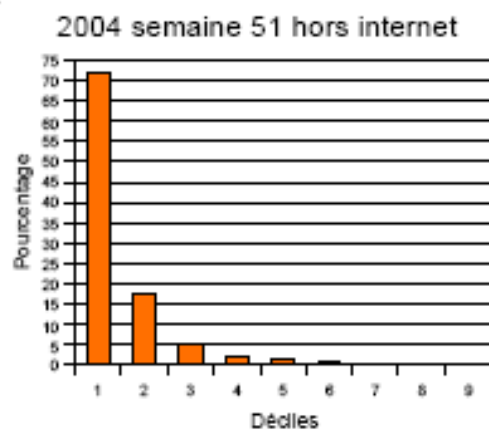
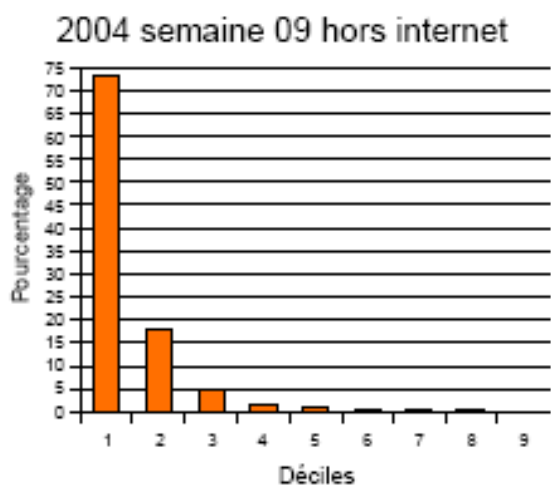
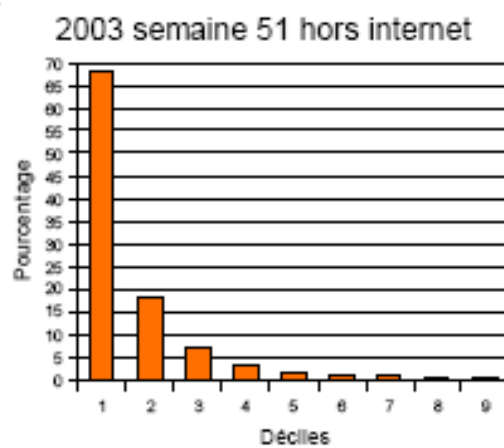
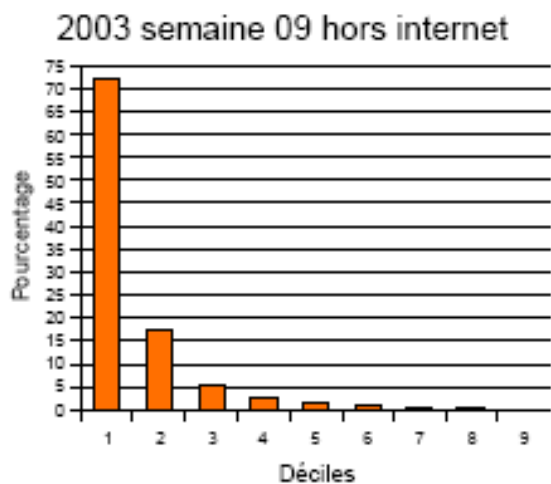
## A.4. Représentation de la queue de distribution : ventes de CD sur et hors Internet

Les graphes suivants présentent les ventes en déciles avec le pourcentage d'artistes concernés dans chaque décile. Afin de pouvoir analyser la queue de distribution, le décile 1 du graphe correspond en fait au 10e de la distribution, et ainsi de suite. Le 1er décile (décile 10 sur le graphe n'est pas représenté car il n'est pas visualisable). Ainsi, sur le premier graphe, les 10% des plus petites ventes sont faites par 35% des artistes).

### Internet



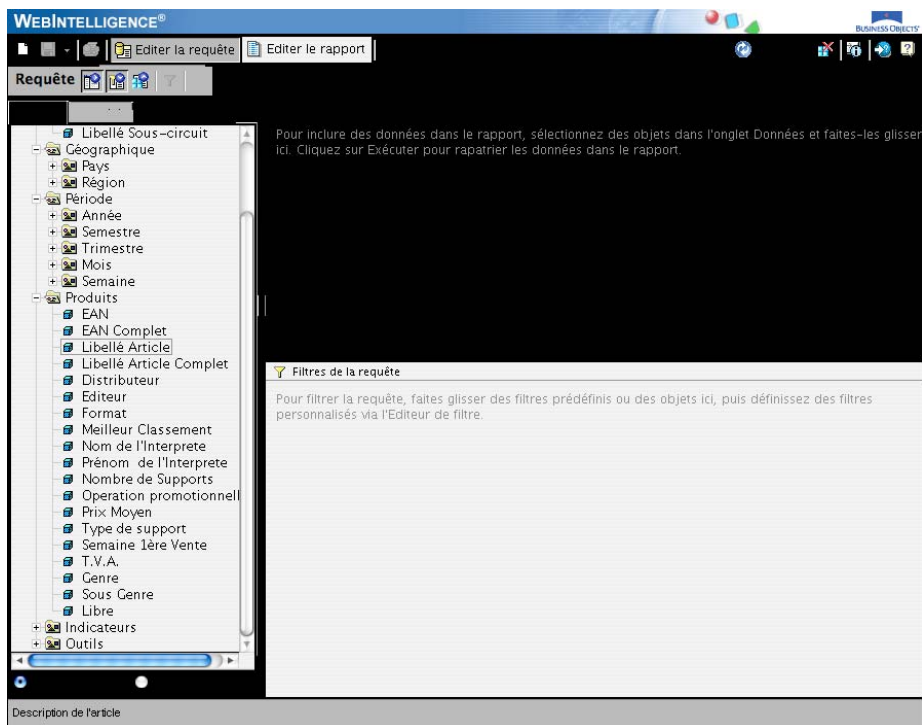
## Marchés traditionnels



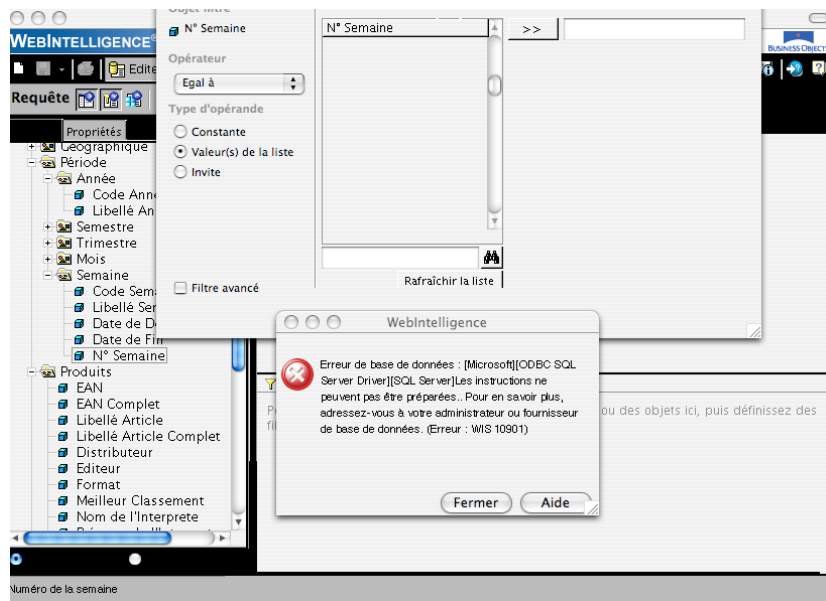
## A. 5. Données GfK – Le logiciel de consultation fourni en ligne

Nous fournissons ci-dessous un exemple des résultats que l'on peut obtenir à l'aide des bases GfK, une fois la prise en main du logiciel effectuée et les principales fonctionnalités maîtrisées.

On présente ci-dessous le cadre de travail : dans le bandeau gauche, on trouve les différents objets et filtres ; il suffit de les faire glisser dans la partie droite pour créer une requête avant de la lancer.



Dans la mesure où la disponibilité du logiciel repose sur la programmation Java, son utilisation ne bénéficie pas d'une stabilité exemplaire. Ainsi les traitements complexes ou longs sont régulièrement confrontés à des arrêts intempestifs (bugs) ou à des temps de latence excessivement élevés.



## A.6. Rappels statistiques et techniques

### *Les critères de mesure d'asymétrie*

Dans une distribution, la *médiane* est la valeur en deçà de laquelle se trouvent 50 % des données. La *moyenne* est la valeur correspondant à la somme des valeurs données, éventuellement pondérée, divisée par le nombre de données utilisées, ou par la somme des poids si les données sont pondérées.

L'*écart absolu moyen* est la mesure de dispersion indiquant la moyenne des valeurs absolues des écarts de chaque valeur par rapport à la moyenne. Le coefficient de variation est l'écart type sur la moyenne en pourcentage.

L'*asymétrie* est caractérisée par le fait que la médiane et la moyenne de la distribution ne sont pas confondues. Quand la moyenne est inférieure à la médiane, la distribution est plus étalée vers la gauche, quand la moyenne est supérieure à la médiane, la distribution est plus étalée vers la droite. Pour rendre compte de ce phénomène, il existe plusieurs indicateurs d'asymétrie. Le coefficient d'asymétrie de Pearson et celui de Fisher qui sont parmi ceux qui sont les plus utilisés en statistique.

- Le *coefficient d'asymétrie de Pearson* compare les positions respectives du mode et de la moyenne en prenant en considération la dispersion. Il s'écrit comme suit :

$$AS_1 = \frac{3(\bar{X} - Mode)}{\sigma}$$

X et  $\sigma$  étant, respectivement, la moyenne et l'écart type de la distribution. Le mode est la valeur dont l'effectif est le plus grand. Lorsque le coefficient d'asymétrie est nul, la distribution est symétrique. La distribution présente une asymétrie AS1 positive ou bien négative selon que le coefficient d'asymétrie est positif ou négatif.

- Le *coefficient d'asymétrie de Fisher* est déterminé à partir des moments et fondé sur un résultat théorique selon lequel les moments centrés d'ordre impair sont toujours nuls pour les distributions symétriques. Il est défini comme suit :

$$\gamma_2 = \frac{\mu_3}{\sigma^3}$$

où  $\mu^3$  indique le moment centré d'ordre 3 de la distribution. Une distribution est considérée comme symétrique lorsque le coefficient d'asymétrie correspondant est nul. Si le moment centré d'ordre 3 est positif, la distribution sera asymétrique avec un étalement à droite. Dans le cas contraire, on aura une distribution asymétrique avec un étalement vers la gauche.

- La comparaison de ces indicateurs sur trois exemples différents nous montre, dans la perspective de l'étude menée, que ces coefficients donnent sensiblement la même indication. Nous avons donc retenu de façon privilégiée le coefficient de Fisher car il permet, à l'ordre 4, d'obtenir d'autres résultats intéressants.

En effet, une *distribution statistique* peut être plus ou moins aplatie selon qu'il y a ou non une forte concentration des observations autour des valeurs centrales de la distribution. La

distribution peut être considérée comme normale (mésokurtique), plate et fortement aplatie (platykurtique), ou bien aiguë et faiblement aplatie (leptokurtique). L'indicateur de Fisher permet d'obtenir un coefficient d'aplatissement déterminé à partir du moment centré d'ordre 4 et mesuré de la manière suivante :

$$AP = \frac{m_4}{s^4} - 3$$

Pour une distribution normale, le coefficient d'aplatissement s'annule. Lorsqu'il est négatif, la distribution est platykurtique. La distribution est, au contraire, leptokurtique dans le cas où ce coefficient serait positif.

### Lexique et précisions sur les valeurs mesurées sous Excel

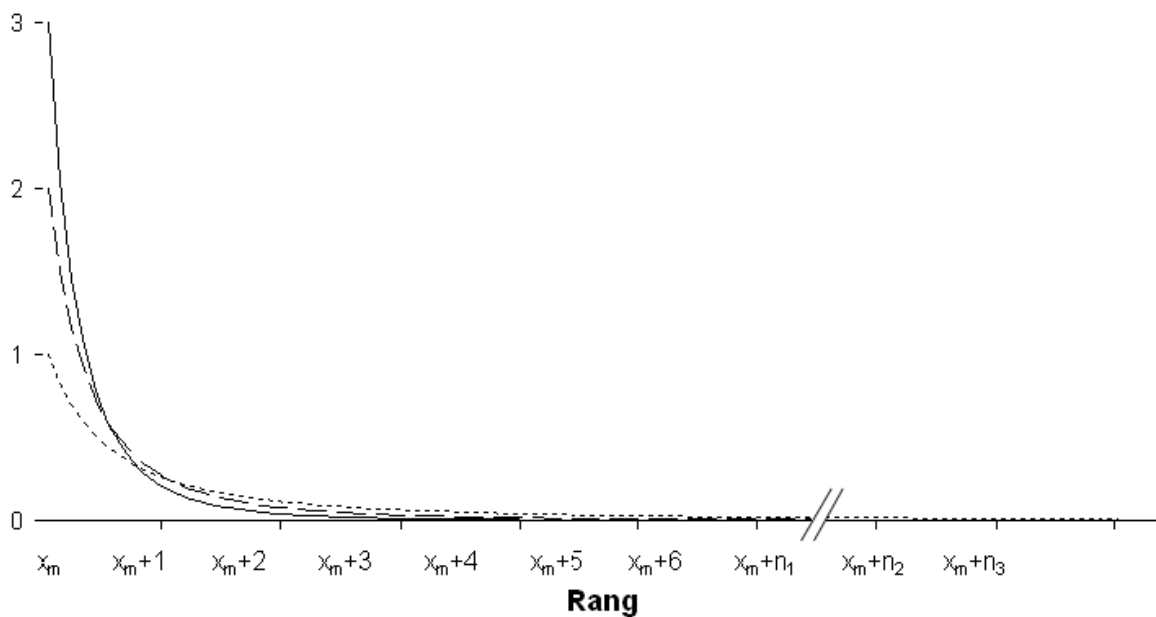
► D'une variable aléatoire  $X$ , on dit qu'elle suit une *loi de Pareto* de paramètres  $(x_{min}; k)$ , si :

$$P(X > x) = \left(\frac{x}{x_{min}}\right)^{-k} \text{ avec } x \geq x_{min}$$

Cette expression correspond à la probabilité que  $X$  soit supérieure à la valeur d'un réel  $x$ , pour tout  $x \geq x_{min}$ , où  $x_{min}$  est la valeur minimale positive que peut prendre  $X$  et  $k$  un réel positif usuellement nommé indice de Pareto. La densité de probabilité associée à la loi de Pareto est la suivante :

$$f(x; k, m) = k \frac{x_m^k}{x^{k+1}} \text{ pour } x \geq x_{min}$$

Son tracé suit :



► Une distribution exponentielle s'exprime par contre selon qu'elle est définie sur l'intervalle  $[0; +1]$  ;

- sa fonction de probabilité s'écrit  $1/b \cdot e^{-x/b}$ , b étant un paramètre d'échelle strictement positif ;
- moyenne  $\mu = b$ , médiane =  $b \cdot \log(2)$  et variance  $\sigma = b^2$

- L'écart type (en anglais *standard deviation*) est un paramètre classique défini en général comme la racine carrée de la variance

$$s^2 = \frac{1}{n} \sum_{i=0}^n (x_i - \bar{x})^2$$

L'utilisation des données sous Excel amène à retenir, en réalité, pour valeur de la variance :

$$s^2 = \frac{1}{n-1} \sum_{i=0}^n (x_i - \bar{x})^2$$

L'écart-type est donc donnée par

$$s = \sqrt{\frac{1}{n-1} \sum_{i=0}^n (x_i - \bar{x})^2}$$

et l'erreur type par :

$$\sqrt{\frac{1}{n(n-1)} \sum_{i=0}^n (x_i - \bar{x})^2}$$

- Le coefficient d'aplatissement est appelé, en anglais, (degré de) *kurtosis* (du grec *kurtos*, « arrondi, convexe »), mot introduit en 1905 par Karl Pearson (1857-1936). En principe, on appelle coefficient de Pearson (pour l'aplatissement) le nombre

$$\beta_2 = \frac{m_4}{s^4}$$

où  $m_4$  est le moment centré d'ordre 4, le moment centré d'ordre k étant défini par

$$m_k = \frac{1}{n} \sum_{i=0}^n (x_i - \bar{x})^k$$

et on le nomme coefficient de Fisher le nombre AP =  $\beta_2 - 3$  défini précédemment (une distribution à peu près normale a en effet un  $\beta_2$  proche de 3). Le logiciel Excel adopte en fait pour valeur du *kurtosis*

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left( \frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

- Le premier quartile est défini comme « le plus petit élément q des valeurs des termes de la série, ordonnées par ordre croissant, tel qu'au moins 25 % des données soient inférieures ou égales à q » ; le troisième quartile est défini de même comme « le plus petit élément q' des valeurs des termes de la série, ordonnées par ordre croissant, tel qu'au moins 75 % des données soient inférieures ou égales à q' ». L'écart interquartile, différence entre les 3<sup>e</sup> et 1<sup>er</sup> quartiles, mesure l'étendue au centre.

- Le coefficient d'asymétrie est en principe donné par

$$\gamma_2 = \frac{\mu_3}{\sigma^3}$$

mais le logiciel Excel prend en fait pour coefficient d'asymétrie

$$\frac{n}{(n-1)(n-2)} \sum \left( \frac{x_i - \bar{x}}{s} \right)^3$$



