

# DU LECTEUR D'UN TITRE DE PRESSE A LA DISTRIBUTION DANS LE TEMPS DE SES LECTURES

Gaël CROCHET<sup>1</sup> & Gilles SANTINI<sup>2</sup>

<sup>1</sup> *Médiamétrie, France, gcrochet@mediametrie.fr*

<sup>2</sup> *Vintco, France, gilles.santini@vintco-media.com*

**Résumé.** La Presse en France a pris des mesures fortes et innovantes pour répondre aux défis du numérique et des nouveaux concurrents, jusqu'à repenser son dispositif de mesure d'audience en le recentrant notamment autour de la marque Média (Print + Digital). Ce nouveau dispositif est désormais sorti de terre et les premiers résultats d'audience sont désormais accessibles au marché. Parmi les principales nouveautés, on trouve une évolution majeure dans les outils de médiaplanning Presse du marché. Celle-ci a été rendue possible grâce à la mise en place d'un panel tempo permettant de suivre pendant un temps les comportements de lecture individuels au niveau des numéros des titres lus. Cette nouvelle donnée a permis d'enrichir les modèles de médiaplanning existants en y introduisant une notion de reprise en main ainsi qu'une notion de vitesse d'accumulation des audiences de la Presse papier. Ces innovations ont permis de faire évoluer les modèles en les complétant avec des notions jusqu'ici associées au médiaplanning digital, comblant ainsi une asymétrie notionnelle entre les deux volets qui constituent le média Presse.

**Mots-clés.** Mesure, presse, audience, médiaplanning, sondage, digital

**Abstract.** The media industry has taken strong and innovative measures to meet the challenges of digital technology and new competitors, even going so far as to rethink its audience measurement survey by refocusing it around the Media brand (Print + Digital). This new system is now out of the ground and the first audience results are now available to the market. Among the main new features is a major evolution in the market's press media planning tools. This has been made possible by the introduction of a tempo panel enabling individual reading behaviour to be tracked for a period at the level of the read issue. This new data made it possible to enrich the existing media planning models by introducing a notion of reading repetition as well as a notion of the speed of accumulation of paper print audiences. These innovations have enabled the models to evolve by supplementing them with notions hitherto associated with digital media planning, thus filling in a notional asymmetry between the two components that make up the Print medium.

**Keywords.** Audience measurement, press, planning, survey, digital

## 1. La notion d'audience d'un titre de Presse

L'audience d'un titre de Presse souffrait jusqu'à présent de limites inhérentes à sa définition. La LDP (Lecture Dernière Période) d'un titre résulte en effet de la réponse des interviewés à la question de Date de Dernière Lecture (en regroupant toute les lectures au sein de la période ...), « que ce soit chez vous ou ailleurs, d'un numéro, même ancien, qu'il soit lu, parcouru ou consulté ». Première conséquence, il ne s'agit pas de l'audience d'un numéro sur la période mais de l'audience de tous les

numéros sur cette période dont la durée dépend de la périodicité du titre. En considérant que l'audience réalisée par un numéro paru avant la période est identique à celle que réalisera le numéro paru pendant la période sur une période ultérieure, on peut considérer que l'audience de la Presse papier reposait jusqu'ici sur l'audience moyenne d'un numéro moyen des différents titres, sans tenir compte de la vitesse d'accumulation de l'audience d'un numéro à partir de sa date de parution, ni de la durée nécessaire pour atteindre son audience totale. On ne connaît pas non plus le nombre de (re)prises en main d'un exemplaire, ni le nombre de pages vues ou lues, etc. Or, c'est un atout majeur de la Presse de pouvoir être lue ou consultée à de multiples reprises, pendant et hors période de publication de référence, au gré des désirs, des besoins, ...

Il existe donc un déséquilibre entre les supports de la Presse papier et leurs cousins du Digital, qui disposent pour leur part d'informations plus nombreuses et détaillées sur les contacts : les modèles de médiaplanning pour le digital s'appuient sur des probabilités individuelles de visites sur un jour moyen du mois, ainsi que sur le nombre moyen de pages vues par jour. Certains logiciels proposent d'ailleurs de représenter les dispositifs digitaux à évaluer dans un calendrier et restituent des performances dans le temps – même si les calculs restent basés sur la notion de jour moyen ou jour nommé.

Le dispositif de mesure des audiences de la Presse papier a ainsi été enrichie d'un Panel Tempo de 4000 individus, dispositif complémentaire permettant de suivre dans le temps les lectures de panélistes ainsi que l'identification des numéros lus (via le scan du code barre depuis une application mobile). Sur la base de cette nouvelle donnée, nous avons travaillé à la mise en place d'un nouveau modèle permettant de gommer l'asymétrie notionnelle décrite précédemment. Nous avons ainsi rapproché les concepts sur lesquels repose le Médiaplanning papier de ceux du médiaplanning digital en ajoutant au nombre de lecteurs (AIR : Average Issue Readership), le nombre de leurs lectures (OTS : Opportunity To See) avec une distribution de celles-ci dans le temps.

Des études montrent en effet que l'audience d'un quotidien se crée quasi ponctuellement au moment de la parution du titre et que pour des périodicités plus élevées, l'accumulation de l'audience se fait de façon plus progressive.

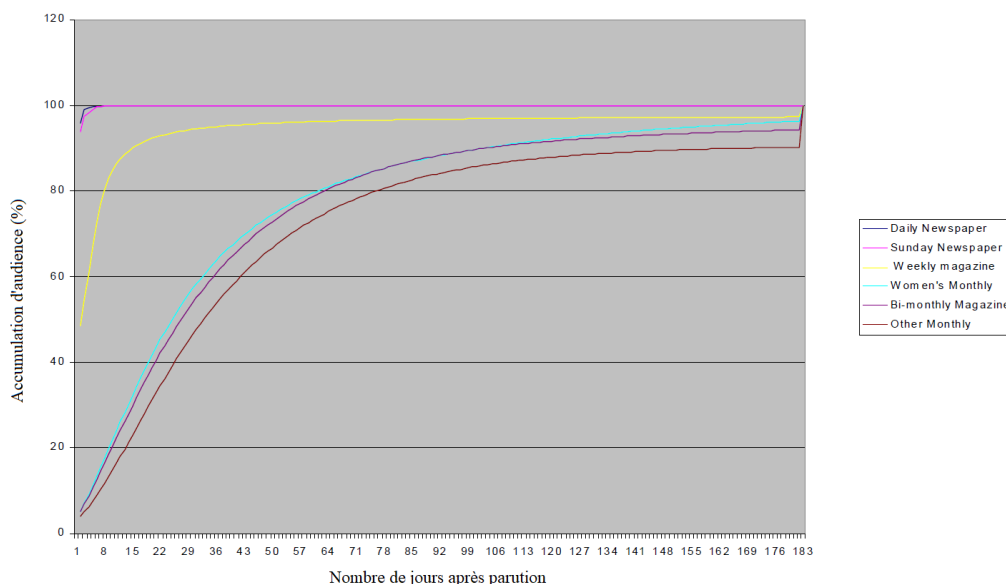


Figure 1 : Courbes d'accumulation d'audience modélisée par périodicité (source ESOMAR WAM 2005 – Time to change time).

## 2. Du lecteur aux lectures d'un titre de presse

### 2.1 Modèle individuel de Lecture d'un numéro d'un magazine

L'observation au cours du temps des actes de lecture d'un numéro d'un magazine par un individu relève de la modélisation d'un processus de comptage. Le plus simple aurait été d'envisager un modèle de type Poissonien mais un tel modèle étant sous-tendu par des hypothèses trop fortes, notamment la stationnarité temporelle et la non-corrélation des lectures dans le temps, nous avons dû nous orienter vers un autre modèle. Il est en effet acquis que la lecture d'un numéro donné d'un magazine subit une obsolescence rapide et donc incompatible avec l'hypothèse de stationnarité.

On a donc été amené à construire un modèle plus complexe donnant à la date écoulée depuis la parution du numéro un rôle plus central.

On formule les quatre hypothèses suivantes :

H<sub>1</sub>- La probabilité d'une lecture au cours de l'intervalle de temps  $(t, t + \Delta t)$  est égale à  $\lambda(t)\Delta t + o(\Delta t)$ , où  $\lambda(t)$  est une fonction intégrable telle que  $\lambda(0) = \lambda$  et  $\lambda(\infty) = 0$  ;

H<sub>2</sub>- La probabilité de plus d'une lecture dans l'intervalle  $(t, t + \Delta t)$  est négligeable d'ordre  $o(\Delta t)$  ;

H<sub>3</sub>- Les hypothèses H<sub>1</sub> et H<sub>2</sub> restent valides lorsque  $\Delta t \rightarrow 0$  ;

H<sub>4</sub>- Les évènements de lecture sont indépendants – i.e. les lectures réalisées à partir d'un instant  $t$  ne dépendent pas de celles réalisées avant  $t$ .

On peut alors écrire que :

$$p_k(t + \Delta t) = p_k(t)(1 - \lambda(t)\Delta t) + p_{k-1}(t)\lambda(t)\Delta t + o(\Delta t),$$

où  $p_k(t)$  désigne la probabilité d'observer  $k$  lectures dans l'intervalle de temps  $(0, t)$ .

En passant à la limite lorsque  $\Delta t \rightarrow 0$ , on trouve l'équation :

$$\frac{dp_0(t)}{dt} = -\lambda(t)p_0(t),$$

$$\frac{dp_k(t)}{dt} = -\lambda(t)[p_k(t) - p_{k-1}(t)].$$

On montre facilement que la solution est donnée par :

$$\left\{ \begin{array}{l} p_0(t) = e^{-\Lambda(t)} \\ p_k(t) = \frac{\Lambda(t)^k}{k!} e^{-\Lambda(t)} \quad k = 1, 2, \dots \end{array} \right.$$

où  $\Lambda(T) = \int_0^T \lambda(t) dt$

Ce processus généralise un processus de Poisson qui est appelé processus de Poisson non-homogène (NHPP). Selon le modèle, la probabilité qu'il y ait au moins une lecture avant la date  $T$  (qui est aussi le complément du fait qu'il n'y ait pas de lecture avant  $T$ ) est égale à :

$$F_T = 1 - p_0(T) = 1 - e^{-\Lambda(T)}.$$

On en déduit l'audience (i.e. la probabilité qu'il y ait au moins une audience à l'infini) :

$$P = F_\infty = 1 - e^{-\Lambda(\infty)}.$$

En exprimant  $F_t$  sur la base de l'audience, on définit la courbe  $C(t)$  d'accumulation de l'audience normée entre 0 et 1. Cette courbe représente la vitesse à laquelle l'audience se réalise avant d'atteindre

son maximum  $P$ .

$$C(t) = \frac{F_t}{P} = \frac{1 - e^{-\Lambda(t)}}{1 - e^{-\Lambda(\infty)}}$$

Nous avons précédemment déclaré un intérêt également pour l'expression du nombre moyen des lectures  $N_T$  ayant été réalisées à cette même date  $T$ , ce qui s'exprime simplement par :

$$N_T = E_T[K] = \sum_{k=0}^{\infty} k \cdot p_k(T) = \sum_{k=0}^{\infty} k \frac{\Lambda(T)^k}{k!} e^{-\Lambda(T)},$$

$$N_T = \Lambda(T).$$

De la même façon que pour l'accumulation de couverture, en exprimant  $N_T$  sur la base du nombre total des lectures cumulées au fil du temps, on exprime la courbe  $N(t)$  d'accumulation des lectures également normée entre 0 et 1. Elle représente la vitesse à laquelle les lectures se font au fil du temps.

$$N(t) = \frac{\Lambda(t)}{\Lambda(\infty)}$$

Par ailleurs, afin d'anticiper la phase de simulation de v.a. suivant un processus NHPP qui sera décrite plus loin, il est utile de s'intéresser à la distribution des temps entre les lectures.

On va introduire la variable aléatoire  $\Theta_T$  désignant le temps d'attente entre une lecture ayant eu lieu à l'instant  $T$  et la suivante. On va s'intéresser à la fonction de répartition de  $\Theta_T$ , que l'on va noter  $G_T(\theta)$  :

$$G_T(\theta) = P[\Theta_T \leq \theta] = P\{\text{Au moins une lecture dans } ]T, T + \theta[ | \text{une lecture à la date } T\}$$

$$= 1 - P\{0 \text{ lecture dans } ]T, T + \theta[ | \text{une lecture à la date } T\}.$$

Comme les évènements de lecture sont par hypothèse indépendants, on peut écrire :

$$G_T(\theta) = 1 - P\{0 \text{ lecture dans } ]T, T + \theta[\}$$

$$= 1 - P\{\text{Nb lectures dans } ]0, T[ = \text{Nb lectures dans } ]0, T + \theta[\}$$

$$= 1 - P\{\text{Nb lectures dans } ]T, T + \theta[ = 0\}.$$

Or, les nombres de lectures dans  $]0, T[$  et dans  $]0, T + \theta[$  suivent des lois de Poisson de paramètres respectifs  $\Lambda(T)$  et  $\Lambda(T + \theta)$  donc le nombre de lectures dans  $]T, T + \theta[$  (obtenue par la différence des deux nombres de lectures précédents) suit également une loi de Poisson, de paramètre  $\Lambda(T + \theta) - \Lambda(T)$ .

La probabilité que ce nombre de lectures soit nul s'exprime donc par :

$$G_T(\theta) = 1 - \frac{[\Lambda(T + \theta) - \Lambda(T)]^0}{0!} e^{-(\Lambda(T + \theta) - \Lambda(T))} = 1 - e^{-(\Lambda(T + \theta) - \Lambda(T))}$$

$$= 1 - e^{-\left(\int_0^{T+\theta} \lambda(t) dt - \int_0^T \lambda(t) dt\right)} = 1 - e^{-\int_T^{T+\theta} \lambda(t) dt} = 1 - e^{-\int_0^\theta \lambda(T+t) dt},$$

$$G_T(\theta) = 1 - e^{-\int_0^\theta \lambda(T+t) dt}$$

L'ensemble de ces résultats permet de complètement connaître le processus NHPP sous réserve de spécifier la forme de la fonction  $\lambda(t)$  usuellement appelée intensité du processus.

Celle-ci doit être une fonction tendant vers zéro à l'infini et il est logique qu'elle soit monotone décroissant à partir d'une certaine date pour refléter l'obsolescence d'un numéro d'un magazine. De plus il est intéressant d'examiner la fonction  $\Lambda(T)$  qui donne le nombre d'événements sur la période  $]0, T]$ .

Si on considère une forme du type  $\lambda(t) = \frac{c}{1+t}$ , souvent utilisée en pratique car elle donne des expressions analytiques simples, on trouve pour  $\Lambda(T)$  :

$$\Lambda(T) = \int_0^T \frac{c}{1+t} dt = c[\log(1+t)]_0^T = c \cdot \log(1+T),$$

ce qui n'est pas acceptable dans le cas présent car  $\Lambda(T) \rightarrow \infty$  quand  $T \rightarrow \infty$ .

On est donc conduit à considérer des formes plus complexes de  $\lambda(t)$ .

On a imaginé utiliser une forme de type Gamma, mais celle-ci ne permet pas d'obtenir des expressions analytiques simples pour la distribution des temps inter-lectures.

On retiendra une décroissance exponentielle qui est un cas simple de loi Gamma :

$$\lambda(t) = \lambda e^{-ct}.$$

On a alors  $\Lambda(T) = \lambda(1 - e^{-cT})$  et  $\Lambda(\infty) = \lambda$  ce qui paraît pour le coup acceptable.

Les quantités précédentes deviennent :

$F_T = 1 - e^{-\lambda(1-e^{-cT})}$	$N_T = \lambda(1 - e^{-cT})$
$P = 1 - e^{-\lambda}$	$N = \lambda$
$C(t) = \frac{1 - e^{-\lambda(1-e^{-cT})}}{1 - e^{-\lambda}}$	$N(t) = (1 - e^{-ct})$

$G_T(\theta)$  s'exprime lui comme :

$$G_T(\theta) = 1 - e^{-\int_0^\theta \lambda(T+t) dt} = 1 - e^{-\int_0^\theta \lambda c e^{-c(T+t)} dt} = 1 - e^{-\lambda c e^{-cT} \int_0^\theta e^{-ct} dt} = 1 - e^{\lambda e^{-cT} [e^{-ct}]_0^\theta},$$

$$G_T(\theta) = 1 - e^{\lambda e^{-cT}(e^{-c\theta}-1)}$$

Cette expression sera reprise plus loin lors de la simulation des lectures (§2.4).

## 2.2 Modèle Bayésien de Lecture d'un numéro d'un magazine

Nous avons jusqu'ici décrit le modèle au niveau d'un individu, auquel est associée une probabilité  $P$  que ce dernier ait au moins une lecture sur le numéro du magazine considéré.

Si on cherche maintenant à décrire le modèle sur l'ensemble de la population,  $P$  peut être vue comme une variable aléatoire suivant une loi Béta de paramètres  $(\alpha, \beta)$  de densité

$$f(p) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}.$$

Au niveau de l'individu, le paramètre  $\lambda$  est lié à  $P$  par la relation :

$$\lambda = \Lambda(\infty) = -\ln(1-P) \Leftrightarrow P = 1 - e^{-\lambda}.$$

A l'instar de  $P$ ,  $\lambda$  peut alors être vue comme une variable aléatoire sur  $[0, \infty[$  de densité  $g(\lambda)$  avec :

$$g(\lambda) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} (1 - e^{-\lambda})^{\alpha-1} e^{-\lambda\beta}.$$

Dans ces conditions, le nombre de lectures  $\tilde{N}$  sur l'ensemble de la population peut être explicité comme suit :

$$\begin{aligned} \tilde{N} = E_\lambda[N_\lambda] = E_\lambda[\lambda] &= \int_0^\infty \lambda g(\lambda) d\lambda = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty \lambda e^{-\lambda\beta} (1 - e^{-\lambda})^{\alpha-1} d\lambda \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} I_1(\alpha, \beta) \end{aligned}$$

$$\text{où } I_k(\alpha, \beta) = \int_0^\infty \lambda^k e^{-\lambda\beta} (1 - e^{-\lambda})^{\alpha-1} d\lambda$$

On utilisera dans la suite la notation  $\alpha^{\bar{l}} = \alpha(\alpha + 1)(\alpha + l - 1)$ . Après quelques développements, simplifications et changements de variables opportuns, on peut montrer que :

$$I_k(\alpha, \beta) = k! \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \sum_{l=k}^{\infty} \frac{s(l, k)}{l!} \frac{\alpha^{\bar{l}}}{(\alpha + \beta)^{\bar{l}}},$$

avec  $s(l, k)$  le nombre de Stirling de première espèce.

De plus,  $s(l, 1) = (l - 1)!$  donc pour  $k = 1$ , on a :

$$I_1(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \sum_{l=1}^{\infty} \frac{1}{l} \frac{\alpha^{\bar{l}}}{(\alpha + \beta)^{\bar{l}}}.$$

En posant  $x - y = \alpha$  et  $y = \beta$ , on peut écrire  $I_1(\alpha, \beta)$  en fonction d'une fonction  $\psi$  appelée digamma telle que  $\sum_{l=1}^{\infty} \frac{1}{l} \frac{(x-y)^{\bar{l}}}{x^{\bar{l}}} = \psi(x) - \psi(y)$ .

On a ainsi :

$$I_1(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} [\psi(x) - \psi(y)]$$

et

$$\boxed{\tilde{N} = \psi(\alpha + \beta) - \psi(\beta)}$$

Par ailleurs, on peut écrire l'audience  $\tilde{P}$  sur l'ensemble de la population de la façon suivante :

$$\begin{aligned} \tilde{P} = E_\lambda[P_\lambda] = E_\lambda[1 - e^{-\lambda}] &= \int_0^\infty (1 - e^{-\lambda}) g(\lambda) d\lambda = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty e^{-\lambda\beta} (1 - e^{-\lambda})^\alpha d\lambda \\ &= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} I_0(\alpha + 1, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + 1)} \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha)} \sum_{l=0}^{\infty} \frac{s(l, 0)}{l!} \frac{(\alpha + 1)^{\bar{l}}}{(\alpha + \beta + 1)^{\bar{l}}} \\ &= \frac{\alpha}{\alpha + \beta} \left( \frac{s(0, 0)}{0!} \frac{(\alpha + 1)^{\bar{0}}}{(\alpha + \beta + 1)^{\bar{0}}} + \sum_{l=1}^{\infty} \frac{s(l, 0)}{l!} \frac{(\alpha + 1)^{\bar{l}}}{(\alpha + \beta + 1)^{\bar{l}}} \right). \end{aligned}$$

Comme  $s(0, 0) = 1$  et  $s(l, 0) = 0 \forall l = 1, 2, \dots$ , on peut écrire que :

$$\boxed{\tilde{P} = \frac{\alpha}{\alpha + \beta}}$$

et le nombre de lectures à la date  $T$  sur l'ensemble de la population :

$$\begin{aligned}\tilde{N}_T &= \int_0^\infty N_T(\lambda) g(\lambda) d\lambda \\ &= \int_0^T \int_0^\infty \lambda(t) g(\lambda) d\lambda dt = \int_0^T c e^{-ct} \int_0^\infty \lambda g(\lambda) d\lambda dt = \tilde{N} \int_0^T c e^{-ct} dt = -\tilde{N}[e^{-ct}]_0^T,\end{aligned}$$

soit

$$\boxed{\tilde{N}_T = (1 - e^{-cT}) \tilde{N}}$$

La probabilité  $\tilde{F}_T$  que la première lecture intervienne avant la date  $T$  est égale à :

$$\begin{aligned}\tilde{F}_T &= E_\lambda[1 - e^{-\Lambda(T|\lambda)}] = 1 - \int_0^\infty e^{-\lambda(1-e^{-cT})} g(\lambda) d\lambda \\ &= 1 - \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty e^{-\lambda(1-e^{-cT})} (1 - e^{-\lambda})^{\alpha-1} e^{-\lambda\beta} d\lambda \\ &= 1 - \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty e^{-\lambda(\beta+1-e^{-cT})} (1 - e^{-\lambda})^{\alpha-1} d\lambda = 1 - \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} I_0(\alpha, \beta'), \\ \tilde{F}_T &= 1 - \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha)\Gamma(\beta')}{\Gamma(\alpha + \beta')} \sum_{l=0}^\infty \frac{s(l, 0)}{l!} \frac{\alpha^l}{(\alpha + \beta')^l},\end{aligned}$$

soit

$$\boxed{\tilde{F}_T = 1 - \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + 1 - e^{-cT})} \cdot \frac{\Gamma(\beta + 1 - e^{-cT})}{\Gamma(\beta)}}$$

A noter que quand  $t \rightarrow \infty$ , on a  $e^{-ct} \rightarrow 0$  et donc

$$\tilde{F}_t \rightarrow 1 - \frac{\beta}{\alpha + \beta} = \frac{\alpha}{\alpha + \beta} = \tilde{P}.$$

Connaissant l'audience  $\tilde{P}$  totale

$$\tilde{P} = \frac{\alpha}{\alpha + \beta},$$

on en déduit la courbe de croissance de l'audience sur la population :

$$\boxed{C(t) = \frac{\tilde{F}_t}{\tilde{P}} = \frac{1 - \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha + \beta + 1 - e^{-cT})} \cdot \frac{\Gamma(\beta + 1 - e^{-cT})}{\Gamma(\beta)}}{\frac{\alpha}{\alpha + \beta}}}$$

qui tend donc vers 1 lorsque  $t$  tend vers l'infini.

Par ailleurs, on a vu précédemment que le nombre de lectures à la date  $T$  est égal à :

$$\tilde{N}_T = (1 - e^{-cT}) \cdot \tilde{N}.$$

On en déduit la courbe d'accumulation des lectures sur la population :

$$N(t) = \frac{\tilde{N}_t}{\tilde{N}} = 1 - e^{-ct}$$

qui tend également vers 1 lorsque  $t$  tend vers l'infini.

$C(t)$  et  $N(t)$  forment, en faisant varier les paramètres  $(\alpha, \beta, c)$ , un réseau de courbes variées qui s'adaptent bien à tous les types de magazines grâce au paramètre  $c$  (que l'on pourrait interpréter comme la vitesse d'accumulation de l'audience du numéro du magazine considéré) et à tous les modes de lecture grâce aux paramètres  $(\alpha, \beta)$ .

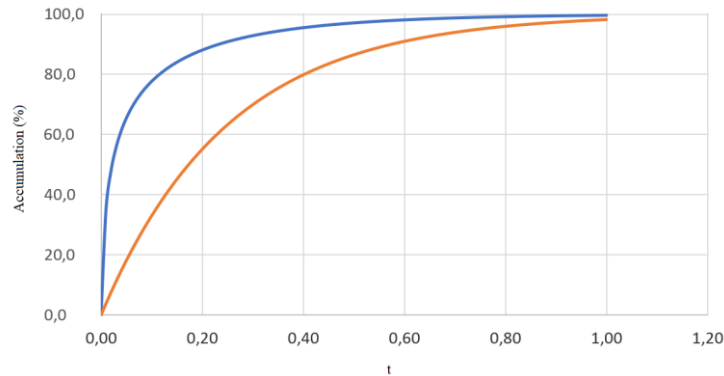


Figure 2 : exemple de courbe d'accumulation des lecteurs ( $C(t)$  en orange) et de leurs lectures ( $N(t)$  en bleu) en fonction du temps.

### 2.3 Estimation des paramètres du modèle

Le panel tempo fournit pour chacun des panélistes, pendant sa période de panélisation, ses lectures de la presse papier avec identification du titre et du numéro lu.

Connaissant la date de parution de chacun des exemplaires et considérant qu'un exemplaire peut se substituer à un autre, il est possible d'identifier des fenêtres  $[t_1, t_2]$ , définies en temps relatif après la date de parution, pendant lesquelles il y a eu au moins une lecture. Le modèle nous permet d'écrire que le nombre moyen de lectures dans une fenêtre où il y a au moins une lecture est en moyenne :

$$N_{t_1, t_2} = \frac{\lambda(e^{-ct_1} - e^{-ct_2})}{1 - e^{-\lambda(e^{-ct_1} - e^{-ct_2})}}$$

Ce nombre dépend de deux paramètres  $(\lambda, c)$  qui peuvent être estimés par minimisation de l'écart quadratique avec la valeur tirée de l'observation panel. Toutefois, cette estimation n'est pas faite globalement mais par segments de population afin de disposer d'une variabilité des paramètres comme nous incite à le faire le modèle Bayésien – ce qui est semblable à ce qui est fait en matière de traitement des audiences moyennes des titres lorsque l'on considère qu'il existe des segments de population dans lesquels la probabilité de lecture est constante.

Les segments utilisés sont dérivés des caractéristiques sexe croisé par âge et de la déclaration de fréquence de lecture faite par les panélistes lors du recrutement.

On pourrait déduire le nombre moyen de reprises en main d'un numéro de la connaissance des paramètres ci-dessus mais le panel tempo disposant d'une information supplémentaire, à savoir si une lecture est la première lecture du numéro, il est apparu utile de l'utiliser. Le modèle nous permet d'écrire que si  $\varpi$  est le pourcentage de fois où il n'y a pas de lecture dans l'intervalle  $[0, T]$  alors :

$$\lambda = \frac{-\ln(\varpi)}{1 - e^{-cT}}. \text{ Sachant que le nombre théorique } r \text{ de reprises en main est } r = \frac{N}{P} = \frac{\lambda}{1 - e^{-\lambda}}, \text{ ayant mesuré}$$



sur le panel une valeur  $r^*$ , on en déduit une correction de la valeur théorique  $g = \frac{r^*}{r}$  que l'on applique à chacun des segments utilisés pour l'estimation.

## 2.4 Calculs en médiaplanning

Les agences de publicité ont l'habitude d'évaluer les performances d'un plan media presse à l'aide de logiciels de simulation qui exploitent des dataset permettant de disposer pour un référentiel d'individus représentatifs de probabilités de lecture d'un numéro moyen.

Le travail effectué permet d'enrichir ces dataset en ajoutant un comportement temporel. Toutefois, de même qu'il a fallu affecter à chacun des individus du dataset une probabilité de lecture, il est nécessaire de lui affecter un comportement temporel. Cela est fait en simulant des patterns de lecture.

Compte tenu de la forme particulière de  $\lambda(t)$ , il est possible de mettre en œuvre une méthode rapide basée sur le principe de la simulation au niveau de l'individu, par inversion de la distribution cumulée des fréquences.

On a en effet établi précédemment que la fonction de répartition associée à  $\Theta_T$  pouvait s'exprimer de la façon suivante :

$$G_T(\theta) = 1 - e^{\lambda e^{-cT}(e^{-c\theta}-1)}.$$

On peut alors appliquer la méthode de simulation par inversion de la fonction de répartition, méthode reposant sur le principe suivant :

*Si  $X$  est une variable aléatoire dont la fonction de répartition  $F$  définit une bijection de  $]a, b[$  sur  $[0,1]$  et si  $U$  suit la loi uniforme sur  $[0,1]$  alors  $F^{-1}(U)$  suit la même loi que  $X$ .*

On peut donc appliquer ce principe en posant  $X = \Theta_T$  et  $F = G_T$  et appliquer la méthode d'inversion à partir de l'expression de la fonction de répartition inverse  $G_T^{-1}$ . En notant  $u = G_T(\theta)$ , on exprime  $\theta$  tel que :

$$\theta = G_T^{-1}(u) = c^{-1} \ln \left( \frac{1}{r(T, u)} \right)$$

où  $r(T, u) = 1 + \frac{\ln(1-u)}{ab e^{-cT}}$ .

On génère alors le processus des temps entre les lectures puis on répercute ce processus au niveau des dates de lecture comme suit.

$U_j$  désigne une variable aléatoire uniforme sur  $[0,1]$ . Les  $\{U_j\}$  sont tirées de façon indépendante.

Les temps entre les lectures sont alors successivement :

$$\begin{cases} \Theta_1 = c^{-1} \ln \left( \frac{1}{r(0, U_1)} \right) \\ \Theta_2 = c^{-1} \ln \left( \frac{1}{r(\Theta_1, U_2)} \right) \\ \Theta_3 = c^{-1} \ln \left( \frac{1}{r(\sum_{k=1}^2 \Theta_k, U_3)} \right) \\ \vdots \\ \Theta_j = c^{-1} \ln \left( \frac{1}{r(\sum_{k=1}^{j-1} \Theta_k, U_j)} \right) \end{cases}$$

Soit encore les dates de lecture suivantes :

$$\begin{cases} T_1 = T_0 + \Theta_1 \\ T_2 = T_1 + \Theta_2 \\ T_3 = T_2 + \Theta_3 \\ \vdots \\ T_j = T_{j-1} + \Theta_j \end{cases}$$

où  $T_0=0$  désigne le jour de parution du numéro (moyen) concerné par la modélisation et les dates  $T_j$ , préalablement arrondies à l'entier inférieur, les numéros de jours de lecture suivant la parution (0 = jour de parution, 1 le lendemain de la parution, etc.)

## 2.5 Quelques résultats du modèle

L'application du modèle sur plusieurs types de titre a permis de mettre en évidence des comportements différents d'une périodicité à l'autre ou même d'un titre à un autre.

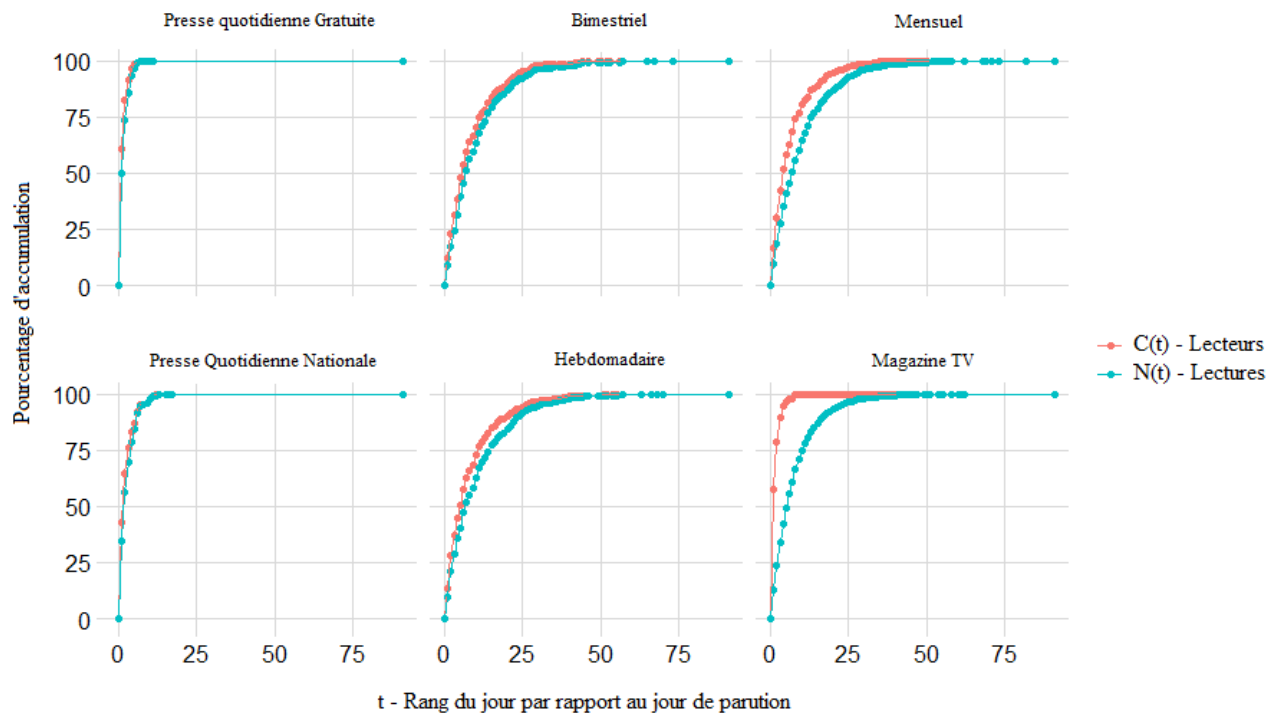


Figure 3 : Exemples de courbes d'accumulation des lecteurs ( $C(t)$  en rouge) et des lectures ( $N(t)$  en bleu) à partir de simulations issues du modèle NHPP.

On peut en effet voir que la Presse Quotidienne est consommée de façon plus concentrée après le jour de parution du titre et que le phénomène est encore accentué pour la Presse Gratuite par rapport à la Presse Quotidienne Nationale.

Les courbes des magazines sont quant à elles plus similaires d'une périodicité à une autre même si on peut voir que la Presse TV atteint très vite son public de lecteurs tout en étant reprise en main au-delà de la période de parution du numéro du magazine. D'un point de vue des courbes d'accumulation des lecteurs, ces titres de la Presse TV sont plus proches des quotidiens que des autres magazines, ce qui semble en phase avec l'idée que l'on se fait de leur consommation.

## Bibliographie

Saint-Joanis, G. et Crochet, G. (2018). *La mesure unifiée des marques de Presse*, 10ème Colloque Francophone sur les Sondages, Lyon.

Santini, G. (2003), *Mathematical models & methods for media research*, G.S. IT Services, Paris.