

MEDIAPLANNING & HYBRIDATION APPLIQUE A L'INTERNET MOBILE

Gaël Crochet¹ & Gilles Santini²

¹ *Médiamétrie, 70 rue Rivay, 92532 Levallois-Perret, France, gcrochet@mediametrie.fr*

² *Vintco SARL, 8 rue Jean Goujon 75008 PARIS, France, gilles@santini.fr*

Résumé. Sur le marché français de l'Internet Mobile, Médiamétrie produit mensuellement des données individuelles probabilisées à des fins d'exploitation médiaplanning et les met à disposition de sociétés éditrices de logiciels spécialisés. Celles-ci mettent ainsi à disposition du marché des outils d'évaluation de performances de divers dispositifs de campagne sur le média. Médiamétrie a ainsi mis au point un dispositif *hybride* innovant qui exploite à la fois des données panel via de la modélisation de comportements individuels (basée sur la loi béta-binomiale largement utilisée dans les médias) et des données agrégées jouant un rôle de contraintes à respecter en terme de niveaux d'audience et de volumes d'exposition. La richesse de l'approche mise en place par Médiamétrie réside autant dans la modélisation qui est effectuée que dans les algorithmes de calage nécessaires aux respects des contraintes.

Mots-clés. Médiamétrie, Vintco, médiaplanning, hybridation, Internet Mobile, mesure, modélisation

Abstract. Mediametrie provides the French mobile advertising market with individual modeled data on a monthly basis. This data is fed into mediaplanning tools (developed by specialized software houses) that evaluate the potential performance of mobile media campaigns. Mediametrie has developed an innovative hybrid approach combining both individual data (by modeling individual media consumption based on beta-binomial distribution widely used within the media industry) and aggregated data (used as additional constraints on audience and volume levels). The richness of Mediametrie's approach resides in the audience modeling that is applied as well as the calibration algorithms used to stay within defined constraints.

Keywords. Médiamétrie, Vintco, mediaplanning, hybrid, mobile market, measures, models...

1 Introduction

L'objectif de cette note méthodologique est de décrire l'approche retenue par Médiamétrie dans le cadre d'une mise à disposition, sur le marché français, de données individuelles probabilisées permettant de procéder à des calculs de médiaplanning via des logiciels dédiés au média Internet Mobile.

La mesure d'audience de l'internet mobile est « hybride ». Nous disposons en effet d'un dispositif de mesure basé sur deux sources mensuelles (dans la suite de ce document, T désignera le mois considéré) :

- données exhaustives fournies par les principaux opérateurs télécoms français (données agrégées sur l'ensemble de la population mesurée)
- données issues d'un panel représentatif de l'ensemble des mobinautes français (on parle ici d'une donnée individuelle qualifiée retraçant l'ensemble des connexions individuelles sur le média).

Ces deux sources nous permettent de qualifier les performances exhaustives des sites et applications

issues des log¹ par des informations sur le profil sociodémographique des mobinautes.

Dans le cadre du médiaplanning, le défi imposé par le marché est double : un, utiliser cette donnée panel pour modéliser le comportement des individus sur chacune des entités² du Média et deux, imposer les niveaux d'audience et de volumes issus des données exhaustives.

Pour chaque entité mesurée et retenue pour figurer dans le fichier diffusé, une probabilité individuelle d'exposition journalière est calculée ainsi qu'une mesure du volume d'exposition correspondant, à partir des données panel, ces deux indicateurs étant ensuite calés de façon à être cohérent avec les données exhaustives. Le processus de calcul de ces deux quantités individuelles, pour chaque entité, se résume en 4 grandes phases décrites dans les paragraphes suivants :

- Modélisation de la distribution de fréquence d'expositions observée
 - Attribution des probabilités d'exposition aux individus du panel
 - Hiérarchisation et Méthode des voisinages
 - Calage des probabilités et volumes d'exposition sur les niveaux fournis par les données exhaustives
- } Donnée panel
} Donnée exhaustive

2 Modélisation de la loi de probabilité individuelle d'exposition

On détermine a priori une segmentation pour chaque entité. Il s'agit d'un découpage sociodémographique dont le niveau de détail dépend du nombre de visiteurs de l'entité à modéliser pendant la période T .

On fait l'hypothèse que, pour une entité donnée, sur un segment donné, les probabilités individuelles d'exposition sont, dans la réalité, réparties continûment et qu'elles peuvent être modélisées par une loi Bêta avec segment à zéro. La fonction de densité de cette loi est donnée par l'expression :

$$f(t) = z\delta_0 + (1 - z) \frac{1}{B(\alpha, \beta)} t^{\alpha-1} (1 - t)^{\beta-1} \quad (1)$$

où δ_0 représente la masse de Dirac en zéro. Le reste de la formule représente une loi Bêta de paramètres α et β .

Pour chaque entité à modéliser, on calcule la distribution de fréquence d'exposition (le nombre d'expositions correspond au nombre de jours où l'individu a eu au moins un contact avec l'entité) observée à l'issue de la période mesurée (28, 29, 30 ou 31 jours selon les mois). On estime les paramètres z , α et β par une méthode d'optimisation qui minimise l'écart entre la distribution de fréquence théorique et la distribution de fréquence observée.

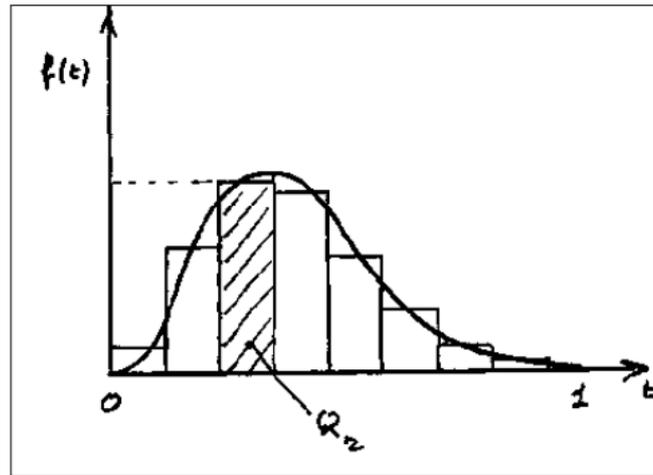
3 Processus d'affectation des probabilités d'exposition aux individus du panel

Nous avons lors du paragraphe précédent déterminé la loi régissant les probabilités individuelles d'exposition à une entité au sein d'un segment de population. Les probabilités à affecter doivent approcher au mieux une distribution de type Beta de densité connue (1) avec z , α et β estimés.

¹ Fichier opérateur regroupant l'ensemble des connexions 3G (ainsi que leurs caractéristiques) effectuées par chaque mobinaute client chez l'opérateur de téléphonie mobile

² Brand, Brand support, Channel, Custom-roll-up

On procède par discrétisation en découpant l'intervalle des valeurs possibles pour $p \in [0,1]$ en n segments égaux (par exemple $n = 100$).



Les bornes de ces intervalles sont donc :

$$p_0 = 0, p_1 = \frac{1}{n}, p_2 = \frac{2}{n}, \dots, p_k = \frac{k}{n}, \dots, p_n = 1$$

La distribution cumulée des p s'écrit :

$$F_{z,\alpha,\beta}(p) = \int_0^p f(t) dt = z + \frac{(1-z)}{B(\alpha,\beta)} \int_0^p t^{\alpha-1} (1-t)^{\beta-1} dt$$

et s'exprime simplement en fonction d'une fonction mathématique standard facilement calculable (intégrale Bêta d'Euler tronquée en x) :

$$F_{z,\alpha,\beta}(p) = z + \frac{(1-z)}{B(\alpha,\beta)} I_{\alpha,\beta}(p)$$

$$\text{où } I_{\alpha,\beta}(x) = \int_0^x t^{\alpha-1} (1-t)^{\beta-1} dt$$

et $B(\alpha, \beta)$ est une constante calculable pour α et β fixés.

Pour simuler correctement la distribution recherchée il faut affecter dans l'intervalle $[0, p_1]$ une quantité de probabilités égale à $Q_1 = (1-z) \frac{I_{\alpha,\beta}(p_1)}{B(\alpha,\beta)}$ et dans l'intervalle $[p_1, p_2]$ une quantité égale à $Q_2 = (1-z) \frac{I_{\alpha,\beta}(p_2) - I_{\alpha,\beta}(p_1)}{B(\alpha,\beta)}$

D'une façon générale il faut affecter dans l'intervalle $[p_{k-1}, p_k]$ une quantité :

$$Q_k = (1-z) \frac{I_{\alpha,\beta}(p_k) - I_{\alpha,\beta}(p_{k-1})}{B(\alpha,\beta)}$$

Dans un intervalle on connaît la moyenne des probabilités qui vaut :

$$M_k = \frac{\int_{p_{k-1}}^{p_k} t f(t) dt}{\int_{p_{k-1}}^{p_k} f(t) dt} = \frac{I_{\alpha+1,\beta}(p_k) - I_{\alpha+1,\beta}(p_{k-1})}{I_{\alpha,\beta}(p_k) - I_{\alpha,\beta}(p_{k-1})}$$

Le principe de la méthode consiste à affecter successivement les z valeurs de probabilités nulles et pour $k = 1, 2, \dots, n$ Q_k probabilités égales à M_k . Pour cela, il va falloir hiérarchiser les individus du moins probable ou plus probable (selon une probabilité calculée dans le paragraphe suivant) et attribuer successivement les probabilités nulles puis M_1 aux individus les moins probables jusqu'aux probabilités M_n pour les individus les plus probables. On attribue ainsi à chaque individu i du segment, une probabilité notée p_i d'être exposé à l'entité considérée.

4 Hiérarchisation et Méthode des voisinages

L'étape 3 nécessite donc de hiérarchiser les individus selon une certaine probabilité d'exposition à l'entité considérée. Médiamétrie a mis en place une méthode de calcul par voisinage qui va permettre de calculer de façon similaire une probabilité d'exposition mensuelle (qui sera donc utilisée pour la hiérarchisation nécessaire dans l'étape 3) et un volume d'exposition associé pour chaque individu du panel et chaque entité à considérer.

4.1 Groupes

On définit dans le panel un certain nombre de groupes d'individus présentant un niveau d'exposition à l'internet mobile croissant.

On isole tout d'abord les individus ayant un niveau d'exposition nul sur la donnée mesurée : ces derniers recevront par convention une probabilité et un volume d'exposition nuls.

On répartit les individus dans chacun des déciles de cette quantité :

$$\forall i \rightarrow d_i = 0, 1, 2, \dots, 10 \quad d_0 \text{ désignant par convention le groupe des non exposés.}$$

Par la suite on travaillera séparément dans chaque groupe et on notera I_G l'effectif du groupe G .

4.2 Distances

On calcule pour chaque individu d'un groupe G sa distance $d(i, j)$ aux autres individus du groupe sur la base des caractéristiques disponibles.

On introduit les quantités : $\delta_{ij} = e^{-\lambda d(i, j)}$ ($\delta_{ii} = 1$) et on fixe $\lambda = \frac{I_G (I_G - 1) \ln(2)}{2 \times \sum_{1 \leq i < j \leq I_G} d(i, j)}$ de manière à ce

que cette quantité prenne la valeur $\frac{1}{2}$ pour une distance égale à la moyenne.

4.3 Probabilités d'exposition sur T (utilisé comme critère de hiérarchisation dans le §3)

Supposons que sur une période T un individu i de poids ω_i ait eu v_i contacts avec l'entité e , on associe à cet individu une observation $x_i = 1$ si $v_i \neq 0$ et $x_i = 0$ sinon.

On calcule la probabilité d'exposition de i comme suit :

$$p_{i,T}^o = \sum_{j \in G} \bar{\omega}_j \frac{\delta_{ij}}{\delta_j} x_j \quad \text{où } \bar{\omega}_j = \frac{\omega_j}{\sum_{i \in G} \omega_i} \text{ et } \delta_j = \sum_{i \in G} \bar{\omega}_i \delta_{ij}$$

Il est facile de voir que $0 \leq p_{i,T}^o \leq 1$ et que en sommant sur tous les individus du groupe :

$$\sum_{i \in G} \bar{\omega}_i p_{i,T}^o = \sum_{i \in G} \bar{\omega}_i \left(\sum_{j \in G} \bar{\omega}_j \frac{\delta_{ij}}{\delta_j} x_j \right) = \sum_{j \in G} \frac{\bar{\omega}_j x_j}{\delta_j} \left(\sum_{i \in G} \bar{\omega}_i \delta_{ij} \right) = \sum_{j \in G} \bar{\omega}_j x_j = \sum_{i \in G} \bar{\omega}_i x_i$$

on retrouve l'exposition totale du groupe.

4.4 Volumes d'exposition sur T

On peut de façon analogue calculer pour chaque individu un volume d'exposition conditionnelle au fait qu'il soit exposé sur la période T :

$$u_{i,T} = \frac{1}{p_{i,T}} \sum_{j \in G} \bar{\omega}_j \frac{\delta_{ij}}{\delta_j} v_j \quad \text{si } p_{i,T} > 0$$

$$u_{i,T} = 0 \quad \text{si } p_{i,T} = 0$$

On a alors :

$$\sum_{i \in G} \bar{\omega}_i u_{i,T} p_{i,T} = \sum_{i \in G} \bar{\omega}_i v_i$$

Le volume d'exposition total du groupe est ainsi bien respecté.

5 Calage des probabilités et volumes d'exposition sur les niveaux fournis par les données exhaustives

5.1 Contraintes liées aux données exhaustives

Les données exhaustives en provenance des opérateurs fournissent pour chaque entité e un niveau d'audience de référence $A^*(e)$ et un nombre de pages/sessions associé $V^*(e)$.

Sachant que l'on a

$$\sum_i \omega_i p_{i,T} = A(e) \neq A^*(e)$$

$$\text{et } \sum_i \omega_i u_{i,T} p_{i,T} = V(e) \neq V^*(e),$$

il convient de modifier les couples $(p_{i,T}, u_{i,T})$ pour rapprocher les valeurs calculées des niveaux issus de l'exhaustif.

5.2 Ajustement des audiences

Pour ajuster les probabilités on procède de façon itérative en transformant à chaque pas de l'itération les valeurs par une transformation homographique (transformée d'Agostini).

$$p_{i,T} \rightarrow p_{i,T}' = 1 - (1 - p_{i,T})^\delta$$

Une description détaillée de cet algorithme est présente dans SANTINI [1]. Il garantit une convergence telle que :

$$\sum_i \omega_i p_{i,T}^* = A^*(e)$$

5.3 Ajustement des volumes

Pour ajuster les volumes on procède par simple correction arithmétique :

Si on note $\sum_i \omega_i u_{i,T} p_{i,T}^* = V'(e)$ le volume d'exposition obtenu une fois les probabilités calées de façon à atteindre l'audience exhaustive $A^*(e)$, il suffit, pour retrouver le volume d'exposition exhaustif $V^*(e)$, de modifier les $u_{i,T}$ par un facteur $\gamma = \frac{V^*(e)}{V'(e)}$:

$$u_{i,T} \rightarrow u_{i,T}^* = \gamma u_{i,T}$$

On a ainsi ce que l'on cherche, à savoir :

$$\sum_i \omega_i u_{i,T}^* p_{i,T}^* = \gamma \sum_i \omega_i u_{i,T} p_{i,T}^* = \gamma V'(e) = V^*(e)$$

5.4 Données journalières

Dans la mesure où les plans média Internet mobile s'étalent en général sur moins d'un mois, il est convenu d'exprimer les données individuelles $u_{i,T}^*$ et $p_{i,T}^*$ sur une base journalière. Faisant l'hypothèse d'une combinaison binomiale au niveau individuel, on passera de la probabilité d'exposition sur la période T (contenant N_J jours) à la probabilité jour par la formule :

$$p_{i,J}^* = 1 - (1 - p_{i,T}^*)^{\frac{1}{N_J}}$$

De même, en ce qui concerne les volumes jour d'exposition (conditionné au fait que l'individu i soit exposé) :

$$u_{i,J}^* = \frac{u_{i,T}^* p_{i,T}^*}{N_J p_{i,J}^*}$$

Les $p_{i,J}^*$ et $u_{i,J}^*$ (pour chaque individu du panel et chaque entité à diffuser) seront les informations transmises au marché pour être intégrées dans les logiciels de médiaplanning du média.

Bibliographie

[1] Gilles SANTINI (2003), *Mathematical models & Methods for Media Research*, G.S. IT Services, Paris.