

MESURE HYBRIDE DE L'AUDIENCE TV

Lorie Dudoignon¹ & Jean Logeart²

¹ *Médiamétrie, 70 rue Rivay, 92532 Levallois-Perret Cedex, ldudoignon@mediametrie.fr*

² *Médiamétrie, 70 rue Rivay, 92532 Levallois-Perret Cedex, jlogeart@mediametrie.fr*

Résumé. Le développement des technologies numériques et des nouveaux moyens de communication offre à Médiamétrie de nouvelles perspectives en matière de mesure d'audience. Les voies de retour permettent aujourd'hui de disposer de résultats exhaustifs sur le nombre de téléphones, d'ordinateurs ou de décodeurs TV connectés. Ces voies de retour ont, en contrepartie, des limites : elles mesurent des machines sans tenir compte du ou des individus les utilisant. Elles sont néanmoins une formidable opportunité pour Médiamétrie d'enrichissement de ses dispositifs de mesure traditionnellement basées sur des données d'enquête. C'est ce mélange de deux mesures complémentaires, de natures et de granularités différentes que nous qualifierons de mesure hybride. Cette communication a pour but de présenter la solution retenue par Médiamétrie pour la mesure hybride en télévision et les enjeux scientifiques et techniques liés à sa mise en œuvre.

Mots-clés. Enquête par sondage, Voie de retour, Modèle de Markov Caché

1 Introduction

Le développement des technologies numériques et des nouveaux moyens de communication offre à Médiamétrie de nouvelles perspectives en matière de mesure d'audience. Les voies de retour permettent aujourd'hui de disposer de résultats exhaustifs sur le nombre de téléphones, d'ordinateurs ou de décodeurs TV connectés. Ces voies de retour ont, en contrepartie, des limites : elles mesurent des machines sans tenir compte du ou des individus les utilisant. Elles sont néanmoins une formidable opportunité pour Médiamétrie d'enrichissement de ses dispositifs de mesure traditionnellement basées sur des données d'enquête. C'est ce mélange de deux mesures complémentaires, de natures et de granularités différentes que nous qualifierons de mesure hybride.

Cette communication a pour but de présenter la solution retenue par Médiamétrie pour la mesure hybride en télévision et les enjeux scientifiques et techniques liés à sa mise en œuvre.

2 Une mesure hybride pour la télévision ?

2.1 Contexte de la mesure d'audience TV

Le panel Médiamat

Le panel Médiamat de Médiamétrie constitue la mesure de référence. Cette mesure est fondée sur un panel de près de 5000 foyers équipés d'au moins une TV en France Métropolitaine.

Tous les postes actifs sont mesurés – c'est-à-dire ceux qui servent au moins une fois par mois pour regarder la télévision. Sur chacun de ces postes, un audimètre détecte automatiquement l'allumage de la télévision, les changements de chaîne et la vision d'enregistrement.

Les individus du foyer doivent participer activement au panel et déclarer leur présence devant le poste de télévision via une télécommande en appuyant sur la touche qui leur est attribuée.

Les données stockées par les audimètres sont collectées en continu.

La voie de retour en TV

La voie de retour en TV est techniquement possible dans deux cas :

- les décodeurs numériques de l'ADSL, du Câble et du Satellite lorsqu'ils sont connectés à Internet
- les TV connectées (encore assez rares)

Les logs de connexion sont alors disponibles auprès de l'opérateur (si celui-ci les récupère) et permettent de savoir sur quelle chaîne ou service est allumé le décodeur. Tout usage fait en dehors du décodeur n'est pas mesuré (si on passe par le tuner TV par exemple).

2.2 Quelle approche pour la TV ?

Deux approches sont possibles pour les mesures hybrides selon le besoin du marché :

- Une première qu'on appellera **Panel-UP** où la mesure voie de retour vient enrichir l'information issue de l'enquête média (le plus souvent un panel).
Dans cette approche la mesure voie de retour va être considérée comme une information auxiliaire que l'on intègre dans le redressement de l'enquête.
- La seconde approche qu'on appellera **Log-UP** consiste en une qualification des données voie de retour.
On construit un modèle à partir des données de l'enquête qui nous permet d'estimer le profil des consommateurs du média.

Le marché souhaite aujourd'hui utiliser la mesure voie de retour pour valoriser des « petites » chaînes. C'est donc l'approche Log-UP qui est retenue car elle permet d'apporter des informations complémentaires à des chaînes à faibles audiences et qui n'ont donc pas beaucoup de visibilité dans l'enquête média.

3 Méthode retenue par Médiamétrie

L'idée est de « transformer » la donnée d'usage (remontée par les décodeurs numériques connectés) en donnée d'audience individuelle grâce à une modélisation en plusieurs étapes :

Etape 1 : passage du décodeur au poste TV

Il s'agit de repérer et d'écarter les usages du décodeur qui semblent plus correspondre au cas où le décodeur reste allumé alors que la TV est éteinte.

Etape 2 : individualisation des audiences poste obtenues à l'étape 1

C'est cette seconde étape qui présente le plus de difficultés et que nous nous proposons de détailler. L'approche que nous avons retenue est une modélisation basée sur la connaissance du profil socio-démographique des « décodeurs » à individualiser (nombre de personnes au foyer, sexe, âge et CSP et lien de parenté des individus).

Les individus du foyer étant « connus », il nous reste à déterminer à chaque instant qui regarde la TV (quand celle-ci est allumée). Pour cela, nous utilisons les Modèles de Markov Cachés.

3.1 Modèles pour individualisation des audiences

L'idée générale est la suivante :

- (1) Les états inconnus (les individus qui regardent la TV) sont modélisés par une chaîne de Markov $\{X_n\}_{1 \leq n \leq N}$ à valeur dans un espace d'états fini X .

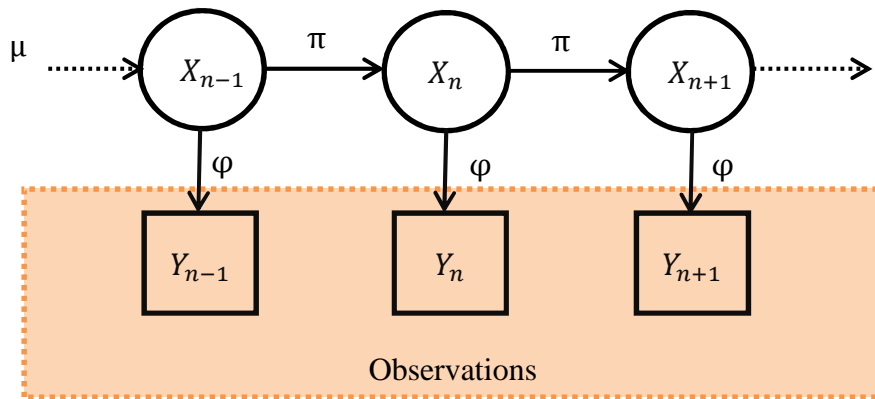
On définit les probabilités de transition : $\pi_{x,x'}^{(n)} = P(X_n = x' | X_{n-1} = x)$ et la loi initiale : $\mu_x = P(X_1 = x)$.

Dans notre cas, nous allons considérer une chaîne de Markov homogène, i.e. la probabilité $\pi^{(n)}$ ne dépend pas de l'indice n .

- (2) Cette chaîne n'est pas directement observée mais on dispose d'observations $\{Y_n\}_{1 \leq n \leq N}$ à valeur dans un espace d'états fini Y . On suppose que ces observations sont réalisées à travers un canal sans mémoire, c'est à dire que conditionnellement aux états X_n , les observations Y_n sont mutuellement indépendantes, et que chaque observation Y_n ne dépend que de l'état X_n au même instant.

$\varphi = (\varphi_x^y)$ est la matrice des probabilités d'observation : $\varphi_x^y = P(Y_n = y | X_n = x)$.

Les probabilités d'observations ne dépendent pas de l'indice n .



3.2 Mise en œuvre

Découpage du temps (indice n) :

La restitution des résultats pour les clients se fera au niveau du quart d'heure. Le découpage temporel pour la modélisation doit être au moins aussi fin que celui de la restitution. Nous avons retenu un pas de 5 minutes.

Définition de l'espace d'état Y :

Nous ne pouvons conserver le détail par chaîne. Nous avons donc regroupé les chaînes par thématiques.

Définition de l'espace d'état X :

Nous considérons tous les états possibles du foyer. Les individus ne sont pas traités indépendamment les uns des autres. Ainsi, nous pouvons conserver les corrélations entre individus d'un même foyer (et globalement les niveaux d'audience conjointe) ce qui est indispensable pour la télévision qui est un média familial.

Les états possibles dépendent de la taille du foyer (NPF pour nombre de personnes au foyer) mais aussi de la composition du foyer :

- Pour un foyer NPF=1, un seul état est possible et la modélisation est inutile (c'est l'individu du foyer qui regarde la TV).
- Pour un foyer NPF=2, par exemple un couple, 3 états sont possibles : le chef de ménage seul, le conjoint seul, le couple.

- Pour un foyer NPF=3, par exemple un couple + 1 enfant, 7 états sont possibles : le chef de ménage seul, le conjoint seul, l'enfant seul, le chef + l'enfant, le conjoint + l'enfant, le couple, le couple + l'enfant.
- ...

On peut assez facilement démontrer que pour un foyer de taille k , le nombre d'états possibles est de 2^{k-1} .

Nous avons mis en place une typologie des foyers qui décrit toutes les compositions foyer à prendre en compte :

- NPF1
- NPF 2 – Couple
- NPF2 – Parent + enfant
- NPF3 – Couple + 1 enfant
- NPF3 – Parent + 2 enfants
- NPF3 – 3 adultes
- ...

A chacun de ces types de foyer correspond un modèle (i.e. un paramétrage $M=(\mu,\pi,\phi)$ par type de foyer).

Tous les paramètres sont estimés à partir des données du panel Médiamat (mesure de référence de l'audience TV, en France) qui en ce sens nous sert d'échantillon d'apprentissage. Les présences devant la TV peuvent ensuite être estimées grâce à l'algorithme de Viterbi.

Bibliographie

- [1] Rabiner, L. R. (1989), A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, 77(2), 257–286.
- [2] Rabiner, L. R. And Juand, B.-H. (1993), *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs.