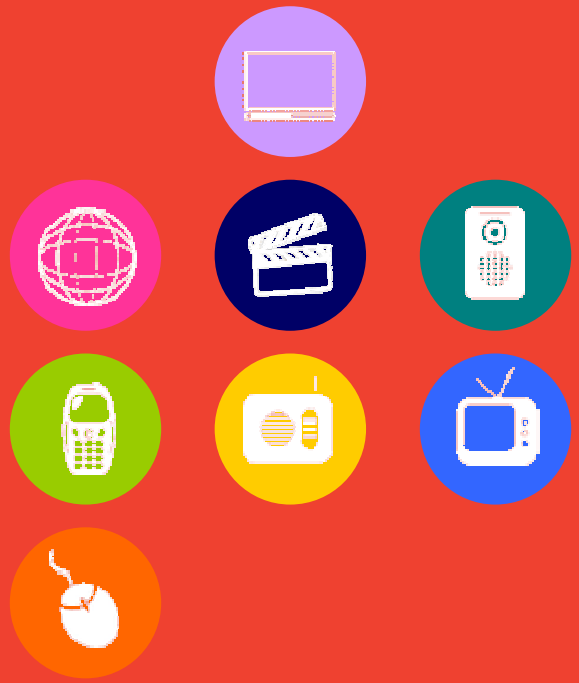
La pratique des sondages à Médiamétrie: état des lieux et perspectives



Journée « Méthodes Avancées pour l'Analyse de Sondages Complexes »

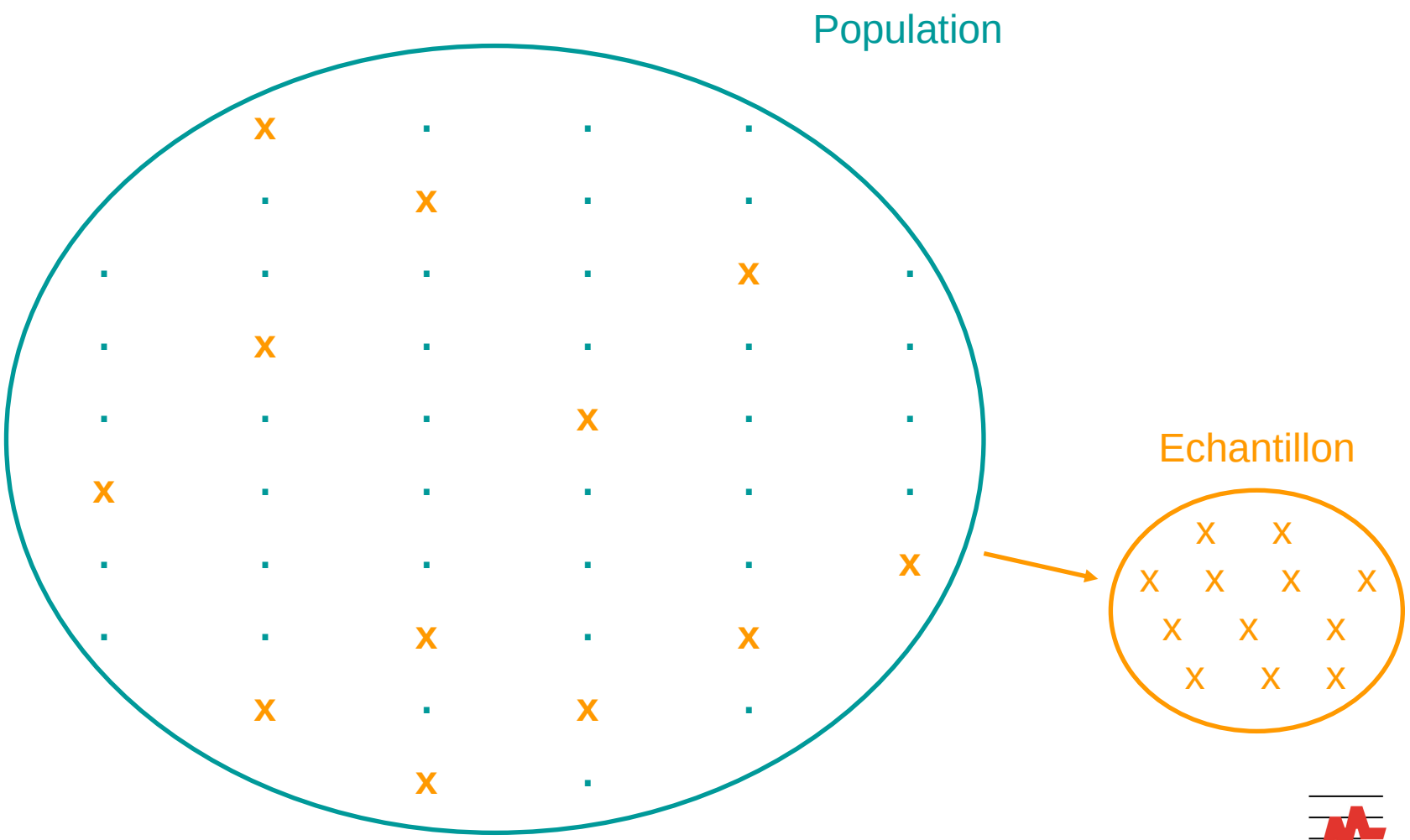
Dijon, le 9 novembre 2010

mediametrie

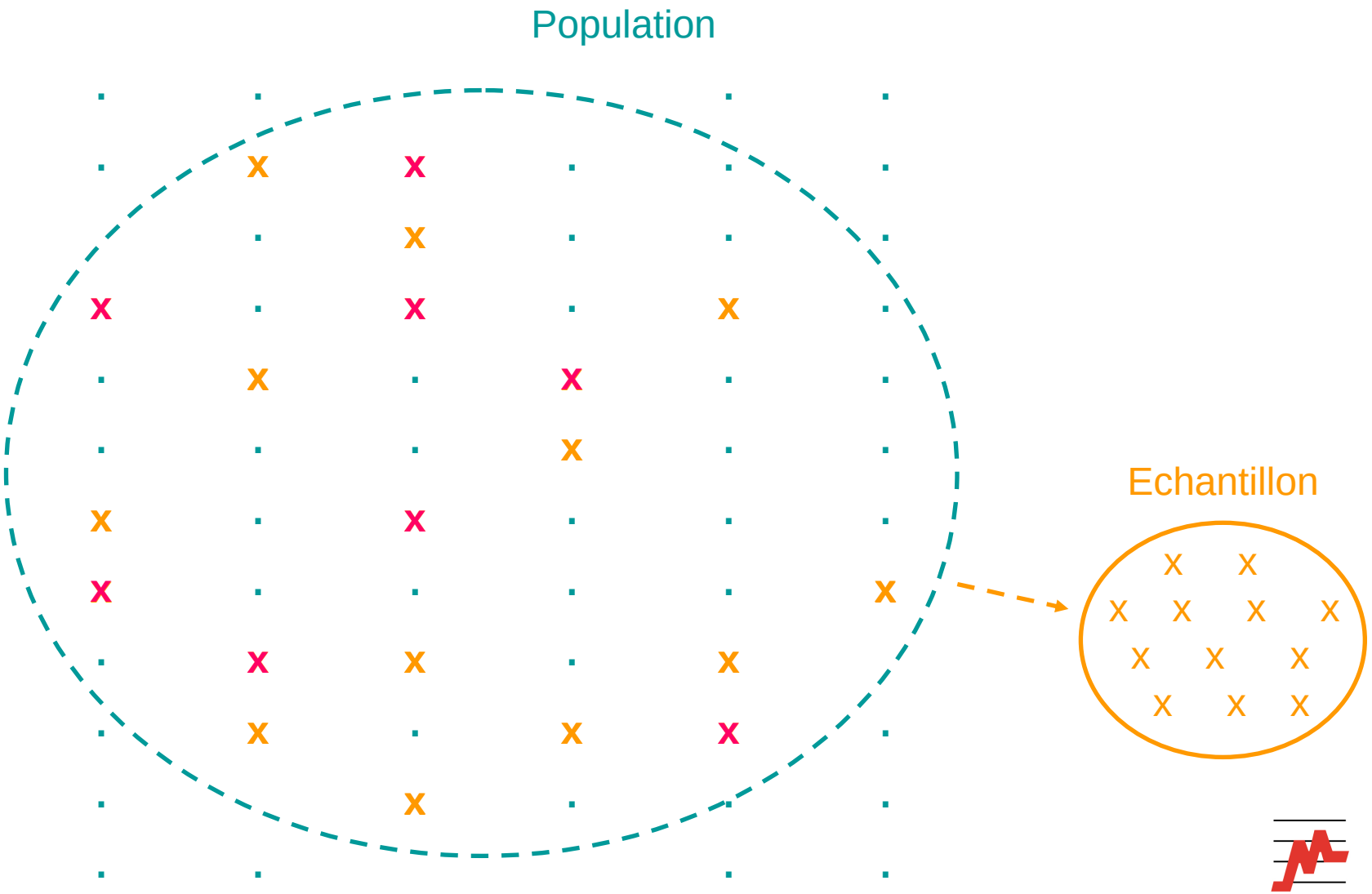
De la théorie à la pratique

mediametrie

Les fondamentaux de la théorie des sondages



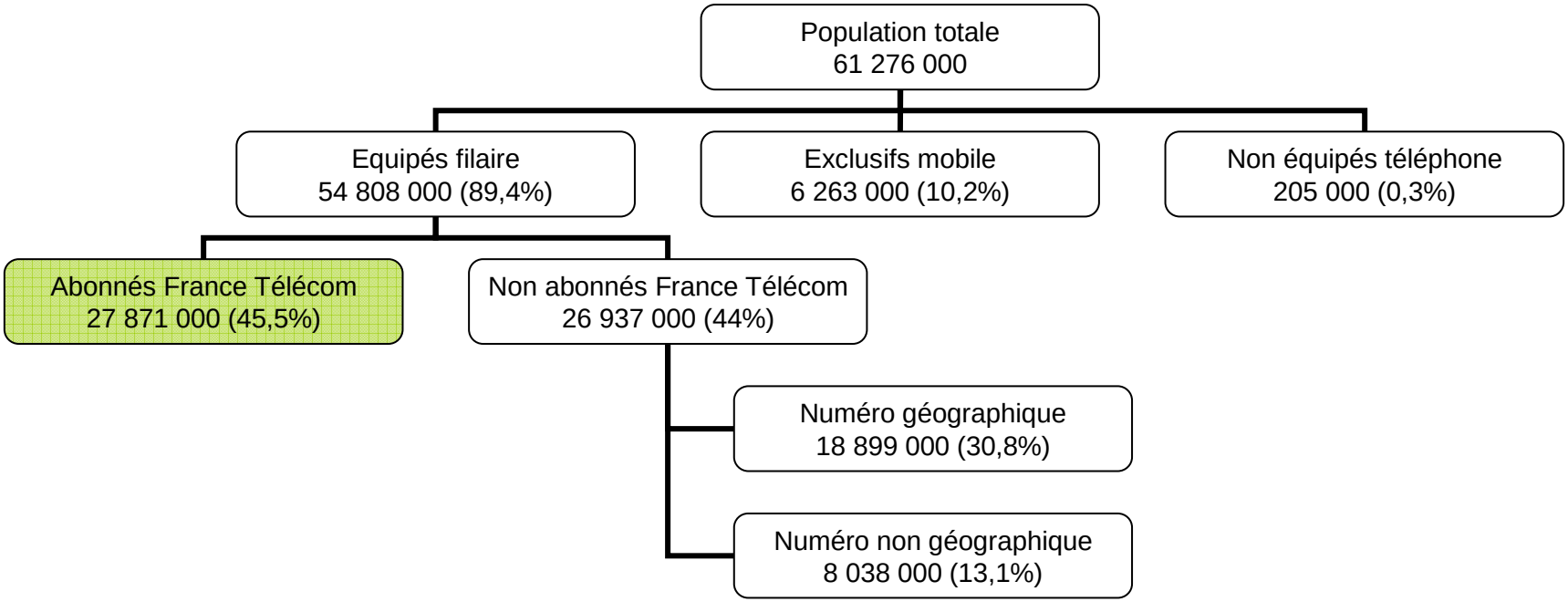
La pratique des sondages en institut



Les difficultés actuelles

Absence de base de sondage exhaustive

- Annuaire France Telecom de plus en plus incomplet



Source : Médiamétrie – Référence des Equipements Multimédia – Janvier-Juin 2010

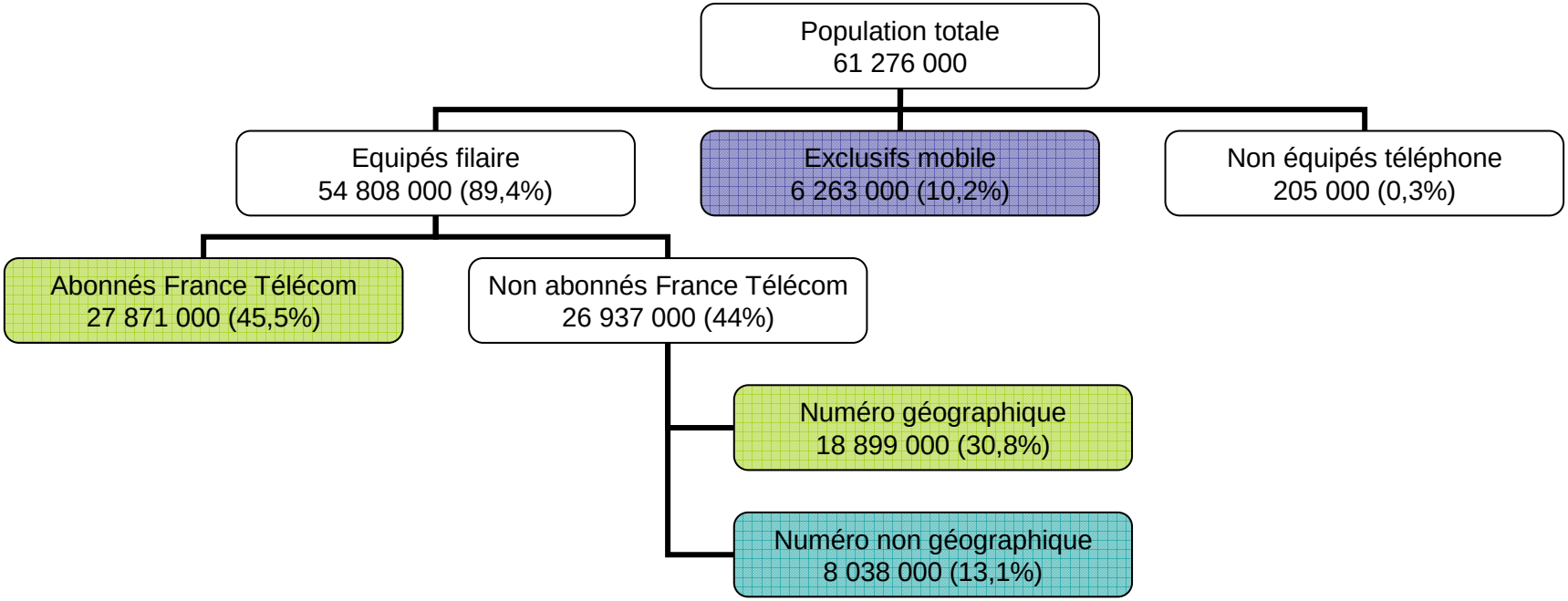
Les solutions mises en place

Constitution de la base de sondage dans le cadre des enquêtes téléphoniques

- Numéros de téléphone fixe géographiques
  - Base de départ : annuaire France Télécom
  - Déclinaison des numéros afin de joindre les numéros en liste rouge
- Numéros de téléphone mobile (depuis septembre 2003)
  - Objectif : joindre les exclusifs mobile
  - Base de départ : numéros générés aléatoirement
  - Qualification préalable de ces numéros afin d'identifier les individus équipés exclusivement d'un téléphone mobile
- Numéros de téléphone fixe non géographiques (depuis janvier 2009)
  - Objectif : joindre les équipés filaire en dégroupage total
  - Base de départ : numéros générés aléatoirement
  - Qualification préalable de ces numéros afin d'identifier ceux qui ne sont pas joignables sur un numéro géographique

Les solutions mises en place

Population couverte



- 99,7% de la population est aujourd'hui couverte

Source : Médiamétrie – Référence des Equipements Multimédia – Janvier-Juin 2010

Les difficultés actuelles

Baisse progressive du taux de réponse aux enquêtes

- Les individus sont de plus en plus mobiles
- Le téléphone fixe n'est plus le principal moyen de communication
- Les sollicitations sont de plus en plus nombreuses

⇨ Aujourd'hui, entre 10 et 20% des appels aboutissent à une interview

Les solutions mises en place

La problématique du taux de réponse

- Recours à la méthode des quotas
  - Sélection « raisonnée » de l'échantillon selon les caractéristiques des personnes interrogées de manière à respecter une structure prédéfinie
  - Respect des proportions données au sein des interviews réalisées (% d'hommes, de femmes, % par tranche d'âge …)
  - Choix des critères de quotas lié à l'information recueillie
- Adapter le mode de recueil à l'interviewé
  - 126 000 Radio : en cas de RDV, on propose aux individus un rappel sur téléphone mobile
  - Panel Radio : lancement d'une version Web du carnet d'écoute papier en septembre 2007 de manière à renforcer le taux de maintien des panélistes en cours d'étude, principalement des cibles les plus « abandonnistes »

Problématiques actuelles

mediametrie

Tout savoir sur tout le monde…

mediametrie

De problématiques mono-média à une problématique pluri-média

Chaque média dispose aujourd'hui d'un dispositif de mesure d'audience spécifique

- Enquête 126 000 Radio pour la Radio
- Panel Médiamat pour la TV
- Panel Médiamétrie//NetRatings pour Internet
- Mesure d'audience de l'internet mobile

Or, aujourd'hui, les stratégies deviennent pluri-média

- L'objectif est d'avoir une estimation de l'ensemble des contacts médias, tout support confondu

De problématiques mono-média à une problématique pluri-média

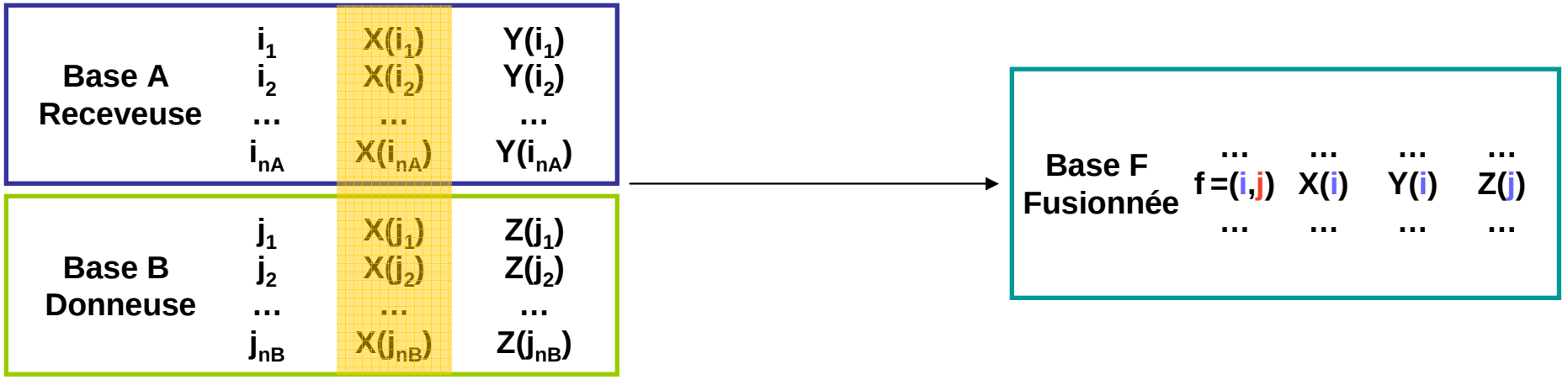
Comment répondre à cette problématique ?

- Option 1 : Constituer un panel sur lequel on recueillerait les comportements d'audience sur l'ensemble des médias
  - Difficilement concevable étant donnés les taux de réponse

- Option 2 : Recueillir, auprès des mêmes individus, la consommation globale des différents médias
  - Ne répond que partiellement à la problématique
  - En particulier, ne permet pas de mesurer la duplication entre supports d'un même groupe média (par exemple, NRJ, NRJ12 et nrj.fr)

De problématiques mono-média à une problématique pluri-média

- Option 3 : Procéder à des fusions statistiques des différentes enquêtes
  - Principe : injecter au fichier dit « receveur » un certain nombre de variables présentes dans le fichier « donneur »



Les fusions de données

Deux types d'approche

- Approche par modèle
  - Chaque variable spécifique est étudiée de manière indépendante
  - Limite : les corrélations observées dans le fichier donneur sont difficilement maintenues dans le fichier receveur

- Appariement
  - Détecter, pour chaque individu du fichier receveur, un ou plusieurs « sosies » au sein du fichier donneur
  - ⇨ C'est cette approche qui a été choisie pour l'étude Cross-Médias

Les fusions de données

Difficultés mathématiques

- Choix d'une fonction de distance
- Choix des critères sur lesquels on détermine la similarité entre 2 individus
- Possibilité d'utiliser plusieurs fois le même donneur ou receveur ?
- Tous les individus doivent-ils apparaître dans la base fusionnée ?
- Conservation des résultats initiaux (moyennes des variables Y calculées sur A et des variables Z calculées sur B)

Limites en terme d'analyse

- Précautions à prendre sur l'analyse et l'interprétation des résultats

Avoir des résultats cohérents…

mediametrie

Des informations auxiliaires de plus en plus nombreuses

Disposer d'information auxiliaire est a priori un atout

- Utiliser une information auxiliaire permet d'améliorer la précision des estimateurs
- Cette information auxiliaire peut être utilisée au niveau de l'échantillonnage ou au niveau du calcul des estimateurs

Jusqu'à présent, les seules informations auxiliaires dont on disposait pour nos enquêtes provenaient de l'INSEE

- Elles étaient utilisées en amont (nb d'interviews par commune, quotas) et en aval (redressement)

Des informations auxiliaires de plus en plus nombreuses

Aujourd'hui, les données des voies de retour sont de plus en plus nombreuses

- Sur Internet, on dispose d'une mesure exhaustive du nombre de pages vues par site
- Sur la TV par ADSL, on connait, à chaque instant, le nombre de boîtiers allumés sur chaque chaîne

Comment utiliser ces informations complémentaires ?

- La question n'est pas tant celle de l'amélioration de la précision des enquêtes que celle de la mise en cohérence des résultats de sources différentes

Des informations auxiliaires de plus en plus nombreuses

La méthode classique : le calage sur marges

- Le principe consiste à modifier les poids de sondage de manière à faire coïncider les totaux observés sur l'échantillon avec ceux calculés sur la population
- L'utilisation de la macro CALMAR, mise au point par O. Sautory (INSEE) sur l'approche générale formulée par J.C. Deville et C.E. Särndal (1992), permet de minimiser la distance entre les poids de redressement et les poids de sondage

Des informations auxiliaires de plus en plus nombreuses

Le problème de la multiplicité des variables auxiliaires

- Combien de variables auxiliaires peut-on faire intervenir ?
- Au-delà de la question de la convergence de l'algorithme se pose celle de la dispersion des poids de redressement
- Dans le cas d'une trop grande dispersion des poids de redressement, l'estimateur redressé peut perdre son efficacité

Trouver un compromis acceptable !

Un cas pratique : la mesure d'audience de l'internet mobile

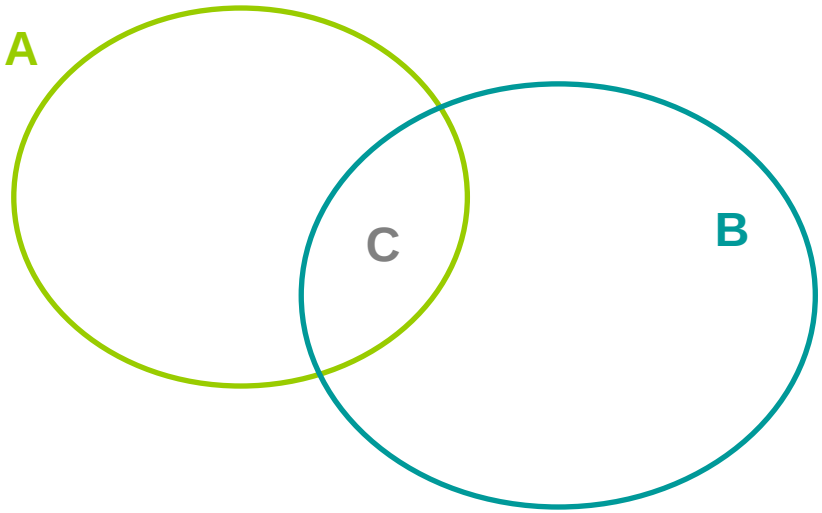
Deux sources d'information

- Les logs des opérateurs de téléphonie
  - Nombre d'individus connectés par site/application
  - Nombre de pages vues par site
- Un panel de 10 000 individus
  - Profil des visiteurs des sites/applications
  - Duplication entre sites/applications

Comment combiner ces deux sources ?

Un cas pratique : la mesure d'audience de l'internet mobile

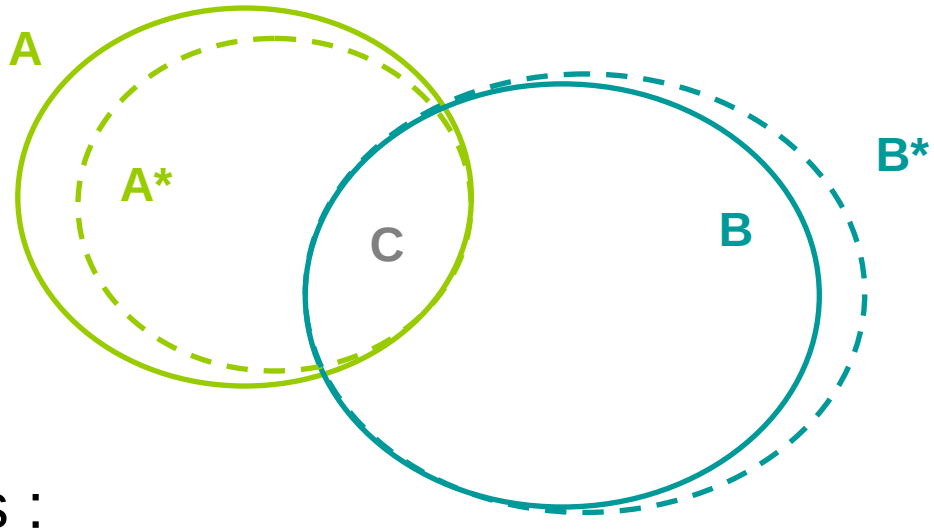
L'estimation de la duplication



- Notations :
  - A, B le nombre de visiteurs respectifs des sites A et B dans le panel
  - C le nombre de visiteurs communs aux sites A et B dans le panel

Un cas pratique : la mesure d'audience de l'internet mobile

L'estimation de la duplication



- Notations :
  - A*, B* le nombre de visiteurs respectifs des sites A et B dans la mesure exhaustive

- Question : estimation de C*

Un cas pratique : la mesure d'audience de l'internet mobile

Les hypothèses de départ

- Dans le cas où A et B sont indépendants, c'est-à-dire

$$C = \frac{A \times B}{N} \quad \text{, alors on veut que} \quad C^* = \frac{A^* \times B^*}{N} = C_{ind}$$

- Dans le cas où A est inclus dans B, alors

$$C^* = \min\left(\frac{C \times B^*}{B}, A^*\right) = C_{inc}$$

- Dans les cas intermédiaires, on est entre les deux…

Un cas pratique : la mesure d'audience de l'internet mobile

Notations

- $K = \min\left(\frac{C}{A \times B/N}, \frac{A \times B/N}{C}\right)$
- $L = \frac{C}{A}$

- Alors C* est estimée par : $C^* = \frac{C_{ind} \times (1-L) + C_{inc} \times (1-K)}{1-L+1-K}$

Conclusion et perspectives

mediametrie

La statistique d’enquête aujourd’hui et demain

Constats

- L’observation seule ne permet plus de répondre aux problématiques actuelles
- On s’éloigne progressivement du cadre « classique » de la statistique d’enquête pour aller vers des systèmes « hybrides » mixant observation et modélisation

Risques

- Ne pas négliger pour autant la qualité de l’échantillonnage
- Le recours à la modélisation ne permet pas de corriger les défauts de l’observation