JUST IN TIME DATA MODELLING

Dina Raimondi, TMPR
Gilles Santini, IMS/IPSOS

Almost all the media research community agrees on the fact that the specific role devoted to each media in the construction of a mix media strategy should be passed, to build media strategies based on specificities and differences.

We used to consider the archetypal role of media, we could summarize as : Print to build image, Tv to bring awareness, Radio to produce traffic, Billboard for visibility. The role of each media has been well known and identified, and particularly the one of print, due to the unique link existing between a reader and his magazine, based on share of interests, ideas, way of life, leisures etc...

The choice of the media itself is a part of communication, the creative aspect of advertising takes this into account as well as the building of a media-mix.

This is not so simple, as each campain is unique, has specific objectives, and as we use more and more media in a roundabout way, far from their original function, to define appropriate media mix.

At the downstreams end of the communication process, advertising efficiency monitoring tools measure a posteriori the results of a particular strategy in terms of effectiveness criteria, as recall, awareness, recognition, advertising evaluation, brand evaluation, etc...

Media Observer, developped by TMPR, evaluates precisely for each campaign the growth in effectiveness criteria due to mix media, in any case of combinations.

The results are very different from one campaign to another, according to the objectives and the associated media mix, as there is not one single rule.

In all cases, we observed that media mix brings a significative plus.

Even if an average value has no signification, as it is based on many specific advertising campaigns, we will give, as a global example, the amount of contribution of print in increasing effectiveness in cases of a TV+print media mix :

- +8.5% on aided recall
- +12% on awareness
- +8.6% on brand evaluation
- +17.6% on brand strength.

Source : Media Observer audits, average 1991-1997;

The effect of print is particularly sensitive on criteria regarding brand which are the most difficult to increase, here is another proof of the indepth work that print is able to do in the mind of consumers, due to the caution that magazines give to advertising, associated to a high receptivity and involvement of the reader.

Post campaign evaluation is easy, but we do not always have a priori the means to find the perfect tuning in media mix, except by using complementarities based on a similarity -or dissimilarity - of audiences, as TV + TV magazines on large targets, or TV + News magazines on upper professional targets etc...

Validation of these media mix is not easy, as audience informations are not available on the same database, and multi-media planning is not possible.

We considered once that fusion could be a solution to obtain multimedia databases, but, without putting on question the interest of fusion on the analytical point of view, to describe behaviours, and media frequentation, we must agree on the fact that media planning remains monomedia, and even in countries where fusion is institutionnalized as Germany, for operationnality reasons, we work media by media.

Furthermore, we should consider the strong constraints on information collection who are indispensable to process a good fusion:

- donor and recipient sample should be comparable and bridge variables genuinely identical,
- the two samples should be balanced,
- bridge variables should include socio-demographics, media habits or centers of interest, consumption or geo-demographics,
- these variables should be collected in an appropriate way.

The basic idea is that, for media planning, we need for each individual his frequency of exposure to every media, when fusion is more adequate to provide a global vision. We work here with something more specific and more quantitative, as exposures to a media vs a more qualitative description of a phenomenon.

Furthermore, in this case, the level of expectancy is higher in terms of precision as these quantitative elements must be significant on the statistical level.

To this double requirement, the statistician answers by the fact that

- a modelling should be done,
- this modelling should be specific to each case story.

This is why we have set up a new methodology for this purpose.

IMS and TMPR defined an alternative route in attempting "just in time modelling" of the data from the various sources in a combined way.

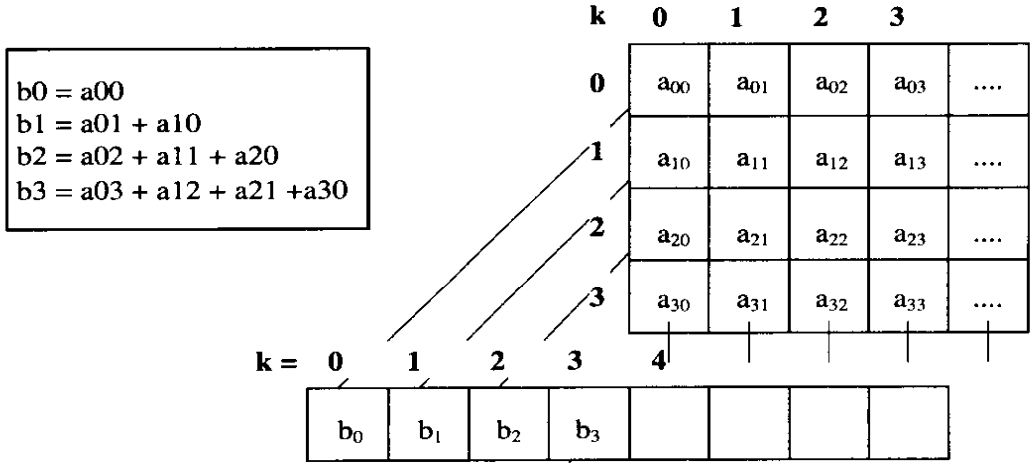
We will now present the methodology involved.

Media Planning

State of the art mediaplanning models provide results that are consistent across complementary targets (i.e. : the distribution on two disjoint targets is the sum of the distribution estimated for each of them separately) and that are free of the decline of the reach paradox (i.e. : the reach decreases when an insertion is added to the schedule).

In order to do so modern models operate at individual level and at one stage or another calculate for each respondent in the target the probability distribution of the number of exposure to the schedule under consideration. In doing so, it is generally accepted that an hypothesis of independence can be applied to the random variables representative of the individual media vehicle.

Such a simplifying assumption leads to the nice property that the frequency distribution for the combination of two media vehicles is the diagonal product of the frequency distribution of exposure to each of them.



If one avoids at this stage the question of the relative value of the exposures obtain from one media compared to the other (this is not a statistical problem but an economic one), it is legitimate to extend the above hypothesis of independence to mix media evaluation.

Due to the additive form of the equation that gives raise at individual level to the frequency distribution on each media it is easy to get the mix media individual frequency distribution by diagonal sums of cross products of the two mono media ones. Such operation is mathematically known as the convolution of these two distribution. Clearly, this is true at individual level only : the

global frequency distribution does not share the independence of the media property and cannot be calculated directly by convolution of the mono media frequency distribution.

Pooling

Considering the above, one needs to know for the same respondent i and a mixed media campain the probability distribution of exposure to each media separately before any cross media evaluation can be calculated.

This is an easy task if two or more media are studied jointly (i.e. : the behaviour of i is known from the audience study for all medias).

As we have already pointed out this is generally not the case in practice. In order to overcome such difficulty one has to pool information from disjoint audience sources. The idea behind the work reported here has been to do it in consideration of the limited task which is to get for each respondent the two probabilities distributions.

To do so and guarantee statistical efficiency, we are ready to pay the price of a specific analysis for each particular situation (i.e. : each different target and each different schedule).

Distribution of exposure segmentation

To achieve the goal set above, we need to model the data base in order to separate respondents who have similar probability distribution of exposure from the other ones. Segmentation trees are good at such a job specially within an already restrained subset of cases (i.e. : the target). The problem is that usual segmentation techniques only handle scalar variable (i.e. : yes/no or a quantity) not frequency distribution. Hopefully the statistician Jack of al Trades tool (i.e. : the chi-square test) provides help to get out of this limitation.

The distance between two frequency distribution f'_k and f''_k $k = 0,1,.....,n$ can be measured as the chi-square of the following contingency table frequency level

	GROUP 1	GROUP 2	
0			
1			
k	x_1^k	x_2^k	x_0^k
n			
	n_1	n_2	n

$f_1^k = \frac{x_1^k}{n_1}$ is the probability of exposure at level k in group 1 and respectively f_2^k in group 2 and f_0^k for the pooled two groups.

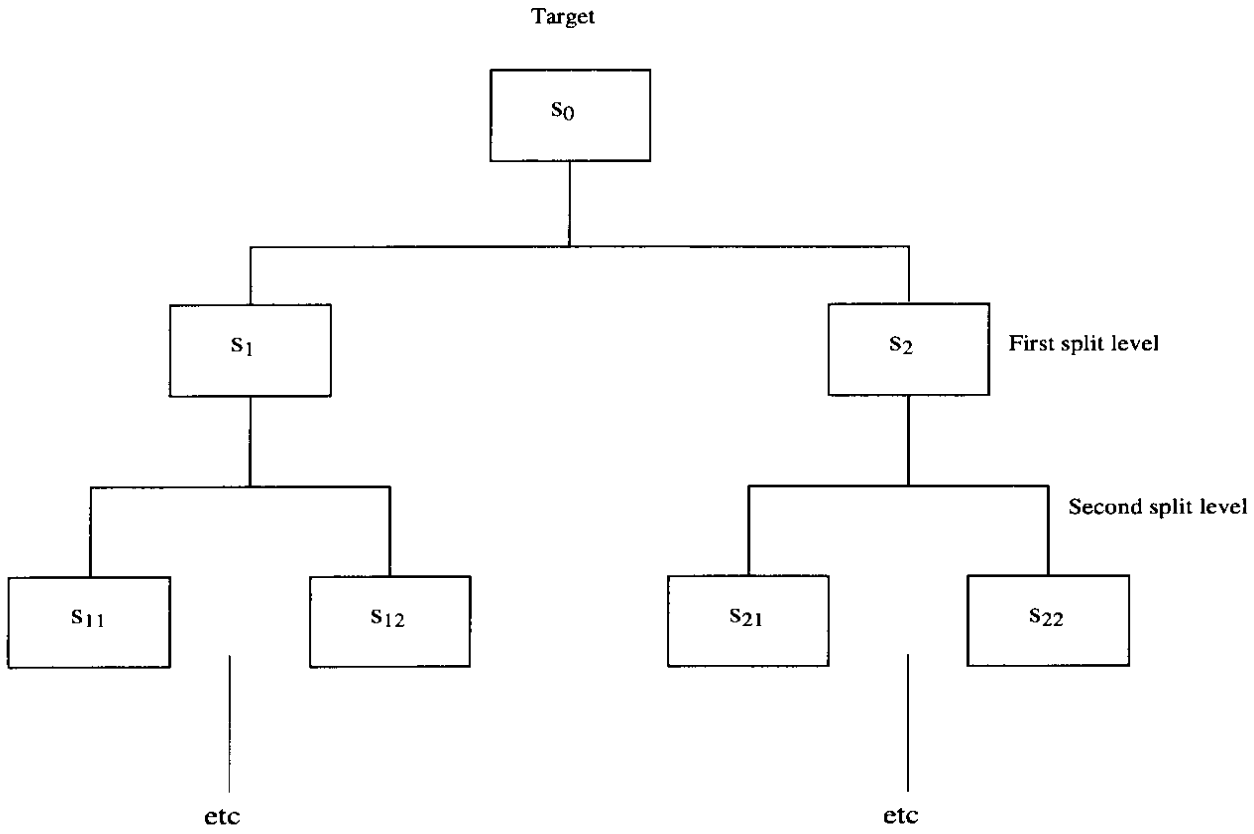
An elementary calculation shows that the chi-square distance can be written as :

$$dchi_2\left(\bar{f}_1, \bar{f}_2\right) = \sum_{k=1}^{n} \frac{\left(f_1^k - f_2^k\right)^2}{f_0^k}$$

If $dchi_2$ is large the two distributions are different.

This statistical indicator is a generalisation of the one used for standard AID segmentation. In such a case it is known as the Sonquist and Morgan criterium.

From here on segmenting the distribution of exposures is straightforward : we split hierarchically the target in sub groups and sub group of sub group maximising at each level the homogeneity within groups and the dissimilarity between groups. The process ends when the chi-square distance between group is not statistically significant or if the number of cases in one group does not allow further split.



At each level, the split is done on the basis of the best splitting variable. Such variable is chosen by exhaustive search within a set of candidate variables known in both audience studies (generally the main sociodemos).

With proper programming techniques the distribution of exposures segmentation can be performed very quickly, *just in time*.

Distribution of exposures transfer

Once a distribution of exposures segmentation process is performed we are able to know to which terminal segment belongs each individual of both files. It is then straightforward to ascribe to each respondent in file 1 the frequency distribution of a respondent within the same segment in file 2. This ascription is done by a hot deck type procedure precisely a random drawing with replacement which avoids regression to the mean and retains within segment variability.

Knowing for each individual in file 1 both its probability of exposures to media 1 and to media 2, allows to proceed easily towards calculation by convolution of the mix media frequency of exposure distribution.

A multimedia mediaplanning software has been developped by IMS for TMPR.

The name of this software is MediaWizard, main benefits of this system are :

- to be multimedia, and work with the current media audience databases,
- to integrate effectiveness data.

The objective is to build, evaluate and optimize multi media strategies, taking into acccount the contribution of each media to the global utility.

It allows the user to think off the beaten tracks, without risks linked to the implementation of a media mix out of the traditional field of practice, by evaluating precisely any combination, and select within innovatives strategies, the one with the best potential.

This enables the user to introduce print media in strategies in the most effective way, quantify in terms of coverage and repetition the global result, and have the best tuning between groups of vehicles.

We will soon illustrate the above by a short live demonstration of the software : presentation of a working implementation on a concrete case, by courtesy of OMD, the worlwide network of media agencies created by BBDO and DDB.

We will present two functions of the software :

- the media scene view gives an exhaustive display of all media offered, shows for each media the available groups of vehicles, their performances in audience and costs, and allows target definition.
  It should be noted that definitions of groups and criteria are the one specific to each media survey.
- the optimal budgeting evaluates, for a particular target, a budget, a choice of groups within two media, TV and Print, and under an effectiveness constraint, all hypothesis of mix-media the user requests.

As a conclusion, the "just in time data modelling" technique allows to consider simultaneously, on a quantitative point of view, all media, to treat all of them on equal terms, to have a precise audience valuation of multimedia strategy, and by so doing, give back to print media the place it deserves in the communication process.