J. AGARD

Une méthode statistique d'évaluation des recettes issues de titres de transport

Revue de statistique appliquée, tome 13, nº 4 (1965), p. 69-74

<http://www.numdam.org/item?id=RSA_1965__13_4_69_0>

UNE MÉTHODE STATISTIQUE D'ÉVALUATION DES RECETTES ISSUES DE TITRES DE TRANSPORT

J. AGARD

Chef de Division, Recherche Opérationnelle d'Air France

1 - INTRODUCTION

Une Compagnie de transport prélève sur ses lignes des titres de transport (coupons issus de billets). Ces titres sont ensuite rassemblés dans un service de recettes qui comptabilise mensuellement les résultats obtenus par relation. La tâche de ce service est particulièrement lourde durant la saison de pointe pendant laquelle ses effectifs devraient augmenter de 50 % pour absorber l'accroissement des coupons. La réelle complexité du travail découle surtout de la disparité de nature des titres de transport saisis sur une même relation. Il esiste en effet des réductions de toutes sortes qui exigent de nombreux tris manuels des coupons. L'évaluation des billets à "quote-part" nécessite en outre la consultation de livres tarifaires : ces billets tiennent compte de tarifs dégressifs avec la distance, et il faut attribuer à chaque coupon une quote-part de la valeur du billet.

L'estimation de la recette provenant du transport par relation puis par région d'émission est de ce fait :

- ou terriblement coûteuse en personnel pour obtenir des résultats exacts ;

- ou approximative, si l'on procède à des échantillonnages ou à l'application de tarifs moyens.

Les méthodes d'échantillonnage permettent de réduire le travail de valorisation des coupons, tout en mesurant l'importance de l'erreur commise. Nous exposons ici la méthode retenue à Air France et quelques résultats de son expérimentation.

2 - DONNEES DU PROBLEME

Le service des recettes reçoit les coupons prélevés lors du transport des passagers au cours du dernier mois. Les indications de ces coupons (segment, numéro, classe, lieu d'émission, valeur...) sont codifiées, perforées sur cartes, puis enregistrées sur bandes magnétiques. L'analyse de ces bandes permet de tirer ensuite tous les renseignements statistiques demandés par la direction. Le problème qui nous importe est de donner une valeur à chaque coupon. Un coupon a une valeur x_i au tarif local normal, et une valeur réelle y_i qui sera souvent inférieure à x_i si le coupon est soumis à quote-part ou à diverses réductions. Le calcul de y_i à partir de la valeur indiquée Y_i sur le billet (valeur totale des divers coupons de ce billet) est généralement long et délicat. C'est à ce problème que notre méthode a été appliquée.

Les coupons sont préalablement triés par relation et par classe. Dans un groupe de base relation-classe, tous les coupons i ont même tarif local x_i, mais ils ont des valeurs réelles y_i différentes. Comment estimer les valeurs y_i ?

3 - METHODES D'ESTIMATION

3.1 - Tarif moyen par groupe.

La première méthode venant à l'esprit consiste à calculer la valeur moyenne $\bar{y}$ de chaque groupe sur un échantillon à p %. Quelques essais nous en dissuaderont vite. En effet, la dispersion des tarifs réduits y_i par rapport à la valeur moyenne $\bar{y}$ est importante. Comme le nombre des coupons dans la plupart des groupes est assez faible, il faudrait prélever des échantillons à 50 % et plus pour obtenir une précision satisfaisante. Le gain par rapport à la valorisation exacte de tous les coupons devient négligeable, compte tenu du temps passé à échantillonner les coupons à P % d'après leur numéro.

3.2 - Constitution d'univers.

Pour réduire le nombre des coupons échantillonnés, il faudrait regrouper les groupes relation-classe voisins en univers. Cela est possible pour des relations appartenant à une région géographique où les mêmes réductions sont appliquées. En créant de la sorte une trentaine d'univers, nous avons étudié les moyennes $\bar{y}_j$ et les écarts type σ_j de la valeur des coupons échantillonnés par univers j. Mais la disparité des tarifs locaux des diverses relations à l'intérieur d'un univers fait que σ_j est de l'ordre de $\bar{y}_j$, ce qui conduirait à échantillonner très souvent à plus de 50 % pour obtenir une précision suffisante.

3.3 - Calcul des "absorptions" par univers.

La méthode mise en oeuvre est une extrapolation d'une méthode utilisée lors des facturations entre compagnies aériennes. En effet, chaque compagnie prélève lors du transport des coupons de billets émis par d'autres compagnies qui en ont perçu la valeur. Les coupons ainsi prélevés sont transmis à ces compagnies pour qu'elles en transfèrent la quote-part à la compagnie ayant effectué le transport du passager. Le calcul des quote-parts étant complexe, on se contente d'estimer par échantillonnage l'absorption c'est-à-dire le pourcentage d'abattement à apporter au tarif local de tous les coupons émis par cette compagnie.

Nous avons appliqué cette méthode à l'ensemble des coupons saisis au transport, quel qu'en soit l'émetteur.

4 - PRINCIPE DE LA METHODE

Considérons tous les coupons d'un univers (un univers groupe tous les coupons transportés le même mois dans une même classe sur les relations d'un secteur géographique). Ces coupons ont des valeurs $x_{i,j}$ au tarif local de la relation et certains ont des valeurs y_{ij} après réductions. Nous calculons sur un échantillon les x_{ij} et les y_{ij} et nous estimons ainsi l'absorption à appliquer aux tarifs locaux de cet univers pour avoir une estimation de la recette, c'est-à-dire que nous calculons

$$m_j = \frac{\sum_i y_{i,j}}{\sum_i x_{i,j}} \qquad \text{pour un univers j.}$$

Tous les coupons ayant été triés par relation-classe, on multiplie leur tarif local de leur relation-classe par le coefficient m_j de leur univers pour estimer leur valeur réelle. Lors du traitement automatisé des coupons, le tarif local x et le nom de la relation seront perforés une seule fois et reproduits automatiquement sur tous les coupons de la même relation-classe, et la valeur estimée $m_j x$ sera attribuée à ces coupons, tandis que le numéro du coupon, ainsi que les caractéristiques du centre d'émission sont perforés pour chaque coupon. Des traitements sur ordinateur permettent ensuite d'estimer la recette provenant du transport global, par relation, par classe, par lieu d'émission, par univers.

5 - PRECISION STATISTIQUE

L'erreur sur la recette d'un univers j provient d'une estimation inexacte de m_j à partir des coupons échantillonnés. Cette erreur est :

$$\sum_{k=1}^{N_j} (y_k - m_j x_k)$$

k = $k^{\text{ème}}$ coupon de l'univers j comprenant N_j coupons.

L'espérance mathématique de cette erreur est égale à 0 puisque

$$m_j = \frac{E(y_j)}{E(x_j)}$$

L'écart type de l'erreur sur un coupon est :

$$\sigma_j = \left[\frac{1}{n_j} \sum_{k=1}^{N_j} (y_k - m_j x_k)^2\right]^{1/2}$$

n_j étant le nombre de coupons dans l'échantillon de l'univers j.

L'écart type de l'erreur relative sur la recette d'un univers sera égal à :

$$\pm \frac{\sigma_j}{\bar{y}_j}\left[\frac{1}{n_j} - \frac{1}{N_j}\right]^{1/2}$$

n étant pris supérieur à 50, on pourra admettre que cette erreur est distribuée selon une loi normale centrée normée.

L'erreur sur la recette globale (tous univers) aura un écart type égal à :

$$\sigma = \left[\sum_j N_j \sigma_j^2 \left(1 - \frac{n_j}{N_j}\right)\right]^{1/2}$$

(sommation j étendue à tous les univers).

L'écart type de l'erreur relative sera :

$$\sigma_{rel} = \frac{\sigma}{\sum_j N_j \cdot m_j \bar{x}_j}$$

On peut de la même façon calculer l'erreur annuelle en pondérant les erreurs mensuelles.

6 - EXPERIMENTATION STATISTIQUE

Les calculs statistiques effectués sur les échantillons de quelques mois montrèrent tout d'abord que la précision à 95 % prévisible compte tenu des σ_j empiriques et des effectifs N_j était de l'ordre de 1 à 2 % et les erreurs pouvaient aller jusqu'à 5 % avec un taux d'échantillonnage fixe de 10 %. Mais une précision de 1 ou 2 % sur des univers à tarif x élevé (long-courriers) déterminait la précision globale, compte tenu du poids en recette de ces univers. Il fut donc décidé d'exclure ces univers à faible nombre de coupons et forte recette de la méthode.

De même certains coupons à très fortes réductions qui sont triés et traités à part ont été exclus, ce qui a diminué la dispersion sur la valeur des coefficients m_j.

En adaptant le taux de l'échantillonnage de façon à avoir toujours une cinquantaine de coupons échantillonnés par univers, nous avons pu réduire l'erreur par univers à 1 ou 2 % dans un intervalle de confiance à 95 % des cas, ce qui conduit à une erreur de recette mensuelle inférieure à 5/1 000 dans 95 % des cas.

7 - EXPERIMENTATION PRATIQUE

La méthode précédente nous garantit contre des erreurs globales de recettes, mais elle ne nous dit rien des erreurs résultant de la ventilation des recettes par relation ou par centre d'émission. Avant de mettre la méthode en exploitation, nous avons donc procédé à une expérimentation à blanc. Pour cela nous avons pris les valeurs des coupons enregistrées sur bandes magnétiques sur deux mois antérieurs, nous avons échantillonné et calculé les coefficients m_j des univers. Nous avons ensuite valorisé chaque coupon en multipliant son tarif local x par le m_j de son univers.

Toutes ces opérations ont pu être menées sur calculateur électronique, à l'aide d'un petit programme. Nous avions ainsi les bandes initiales où les coupons avaient des valeurs exactes y_j et des bandes transformées où les coupons avaient des valeurs estimées xm_j.

Nous avons alors effectué avec les programmes d'exploitation existants la ventilation des recettes mensuelles par relation et par centre d'émission, puis un programme de comparaison a permis de sortir les écarts entre les chiffres d'affaire exacts et estimés pour chaque rubrique.

Nous avons ainsi pu mesurer l'importance des erreurs systématiques provenant peut-être de l'existence de réductions moyennes sur une relation ou un centre d'émission différentes des réductions sur l'univers. Un classement a pu montrer que sur 400 relations, il n'y avait que 25 % des cas où l'erreur relative mensuelle dépassait 5 %, que dans 60 % des cas elle était inférieure à 3 %, et dans 30 % des cas inférieure à 1 %. Pour 200 centres d'émission, l'écart relatif de recette était supérieur à 5 % dans 4 % des cas seulement et il était inférieur à 3 % dans 85 % des cas, et inférieur à 1 % dans 60 % des cas.

Les services commerciaux et financiers estimèrent que la précision obtenue convenait aux besoins statistiques et la méthode a été lancée de façon satisfaisante depuis 1963, permettant une économie d'agents de dépouillement, tout en gardant un contrôle sur la précision globale de la recette estimée.

8 - OPTIMISATION DE L'ECHANTILLONNAGE

Soit V_i la somme des valeurs exactes des coupons de l'univers i et $\mathcal{V}_i$ la somme de leur valeur au tarif local, tandis que $E[m_i] = \frac{V_i}{\mathcal{V}_i}$. En fait m_i résulte d'un échantillon sur n_i coupons de l'univers i. Son estimation par échantillonnage entraîne une erreur sur l'univers dont la variance est :

$$E[(V_i - m_i \mathcal{V}_i)^2] = V_i^2 - 2E(m_i) V_i \mathcal{V}_i + E(m_i^2) \mathcal{V}_i^2$$

$$E(m_i^2) = m_i^2 + \frac{\sigma^2_{m_i}}{n_i}$$

avec σ_{m_i} écart type de l'erreur sur le coefficient m_i de l'univers i

$$E\{(V_i - m_i \mathcal{V}_i)^2\} = \frac{\sigma^2_{m_i}}{n_i} \mathcal{V}_i^2 = \sigma_i^2 = \sum_{j=1}^{n_i} (y_j - m_i x_j)^2$$

d'où :

$$\sigma_{m_i} = \frac{\sigma_i \sqrt{n_i}}{\mathcal{V}_i}$$

La variance de l'erreur sur la recette globale est :

$$V(R) = \sum_i \frac{\sigma^2_{m_i} \mathcal{V}_i}{n_i}$$

($\mathcal{V}_i$ étant la recette totale de l'univers i et n_i le nombre de coupons échantillonnés).

Proposons nous de minimiser la variance V(R) pour un nombre fixé N de coupons échantillonnés

$$N = \sum_{i=1}^{i=t} n_i \quad \text{(avec t univers)}$$

Nous obtiendrons l'optimum par application des multiplicateurs de Lagrange pour :

$$\frac{\mathcal{V}_1^2 \sigma^2_{m_1}}{n_1^2} = \frac{\mathcal{V}_2^2 \sigma^2_{m_2}}{n_2^2} \cdots = \frac{\mathcal{V}_t^2 \sigma^2_{m_t}}{n_t^2}$$

soit

$$\frac{n_1}{\mathcal{V}_1 \sigma_{m_1}} \cdots = \frac{n_i}{\mathcal{V}_i \sigma_{m_i}} = \frac{N}{\sum_{i=1}^{i=t} \mathcal{V}_i \sigma_{m_i}}$$

Les taux optimaux d'échantillonnage seront :

$$f_i = \frac{\mathcal{V}_i \sigma_{m_i}}{\sum_i \mathcal{V}_i \sigma_{m_i}} = \left(\frac{n_i}{N}\right)$$

L'application de cette méthode par rapport à un échantillon uniforme à même nombre de coupons prélevés s'est traduit par un gain de précision de 50 %. Pour éviter d'avoir à réaliser des échantillonnages à taux variable suivant l'univers, nous nous sommes contentés d'utiliser les taux à 1 %, 2 %, 5 %, 10 % et 20 % suivant les univers. Les résultats obtenus donnent satisfaction aux utilisateurs et ont permis un gain de plusieurs agents avec une précision accrue.