

METHODE DE SONDAGE POUR L'ETUDE DE L'INFLUENCE
DE NOUVEAUX CRITERES SUR LE TARIF
„AUTOMOBILE”

B. DUBOIS DE MONTREYNAUD

Paris, France

Pour construire un tarif „Automobile” il est nécessaire d'étudier la charge des sinistres causés par les véhicules appartenant à des groupes homogènes divers: il est commode de procéder à un échantillonnage.

On peut construire un échantillon où figurent des proportions constantes ou connues de chacun des groupes constituant la population à étudier.

En pratique l'échantillonnage, doit, pour avoir une efficacité admissible, être particulièrement important dans les groupes à faible fréquence pour donner un nombre significatif de sinistres.

Or, comme pour déterminer la méthode de sélection des échantillons on ne connaît pas le groupe des véhicules sélectionnés, on serait contraint de prendre un échantillon très important.

En France, le Groupement Technique Accidents fait des études sur la totalité des véhicules assurés par les Compagnies adhérentes à la statistique commune et qui représentent un tiers du Parc National.

L'inconvénient de la méthode actuelle est l'impossibilité d'étudier rapidement un nouveau critère — car il faut attendre les mouvements sur le portefeuille à l'occasion desquels on enregistre le critère à étudier; les mouvements, en outre, frappent probablement moins les véhicules à faible fréquence. Le délai nécessaire pour l'étude d'un nouveau critère se compte en années.

Il peut être souhaitable pour suivre l'évolution économique ou sociale d'étudier rapidement dans un délai chiffré en mois, l'influence de certains critères nouveaux, pour éviter sous la pression d'évènements extérieurs d'être amené à prendre des décisions dites „politiques” — cet euphémisme justifiant, dans le doute, des réductions de tarif explicables uniquement par le fait qu'un concurrent dangereux est peut-être sérieux — à ce moment il est inévitable de procéder à un sondage.

Plusieurs méthodes sont théoriquement possibles. On peut rejeter d'emblée la méthode qui consiste à faire une enquête sur des véhicules choisis d'après un critère basé sur le numéro minéralogique, cette méthode étant d'un prix prohibitif car elle oblige la recherche des propriétaires, leur interview, et celle de leurs accidents auprès de leurs organismes assureurs successifs; ces derniers étant les seuls pouvant donner des éléments objectifs tant sur le coût des sinistres que sur leur nombre.

Il est certainement moins onéreux de procéder à un échantillonnage dans le parc des organismes assureurs désireux de faire ces recherches. On pourra étudier le comportement des véhicules sélectionnés d'après leur numéro de police, ce qui donnera certainement un échantillon représentatif de la population à étudier.

Cette méthode, plus économique que la première, nécessite néanmoins l'interview des assurés; son coût reste élevé.

La méthode que nous proposons consiste à prendre comme éléments constitutifs de l'échantillon, l'ensemble des véhicules sinistrés au cours d'une ou plusieurs périodes fixées a priori.

Cette méthode présente incontestablement un plus faible prix de revient, car l'interview du conducteur sera facilitée par le fait qu'il prendra lui-même contact avec son organisme assureur à l'occasion du sinistre. Quant aux éléments que l'on doit puiser dans les archives de la Compagnie relatifs aux sinistres passés, ils n'obligent pas à des recherches supplémentaires, puisqu'à l'occasion des sinistres déclarés il est en général nécessaire de se référer aux éléments statistiques de la police.

Il est bien évident toutefois que l'échantillon ainsi constitué ne sera pas représentatif de la population des véhicules, les catégories à forte fréquence étant en trop forte proportion par rapport aux catégories à faible fréquence; mais l'erreur est si grossière qu'il est facile d'y remédier.

REMARQUE: Nous appellerons groupe homogène un groupe de véhicules tel que tous les véhicules qui le constituent ont la même probabilité de sinistre; par définition les sinistres se répartiront sur les véhicules suivant la Loi de Poisson.

NOTATION

- f = Fréquence
 = Rapport du nombre de sinistres au nombre de véhicules x années
 = Espérance mathématique du nombre de sinistres pour un véhicule x année
 n = Accroissement annuel du nombre de véhicules d'un groupe homogène.
 N = Nombre de véhicules d'un groupe homogène
 N' = Nombre de véhicules x années.
 S = Nombre de sinistres
 σ_t = Nombre de véhicules sinistrés pendant le temps t
 σ_{tn} = Nombre de véhicules subissant n sinistres pendant le temps t
 t = Temps
 t_i = Intervalle de temps séparant deux sinistres consécutifs survenus au véhicule i
 T = Durée de l'échantillonnage
 θ = Durée de la Garantie
 τ = Ancienneté des plus anciens véhicules du groupe.

Considérons un groupe homogène de N véhicules à un instant donne $t = 0$ et de fréquence annuelle f .

Comment se distribue la loi de survenance du premier sinistre de chaque véhicule ?

A chaque instant t , le nombre de véhicules ayant subi n sinistres depuis l'origine $t = 0$ est :

$$\sigma_{tn} = N \frac{f^n t^n}{n!} e^{-ft}$$

et en faisant $n = 0$ nous avons le nombre de véhicules non sinistrés soit :

$$\sigma_{t,0} = N e^{-ft}$$

L'expression $\sigma_t = N - \sigma_{t0} = N (1 - e^{-ft})$ nous donne à chaque instant t le nombre de véhicules qui ne sont plus dans la catégorie des non sinistrés c'est-à-dire qui ont subi au moins leur premier sinistre.

Il apparait que la différence

$$\sigma_{t+dt} - \sigma_t = d\sigma_t = N f e^{-ft} dt$$

est égale au nombre de véhicules subissant leur premier sinistre entre les temps t et $t + dt$.

Nous en déduisons que la loi de répartition des premiers sinistres après un temps t est :

$$\frac{d\sigma_t}{dt} = N f e^{-ft}$$

Nous examinons maintenant le temps moyen écoulé entre l'origine des temps et le premier sinistre.

Le temps moyen est par définition la valeur d'une fraction où nous mettons au numérateur le total des temps écoulés pour chaque premier sinistre entre l'origine des temps et le moment de survenance du premier sinistre, et au dénominateur le nombre de véhicules sinistrés.

Or, quand t tend vers l'infini le nombre de véhicules sinistrés tend vers N le temps moyen cherché est donc :

$$\begin{aligned} \Theta &= \frac{1}{N} \int_0^{\infty} N f t e^{-ft} dt \\ &= \frac{1}{f} \int_0^{\infty} f t e^{-ft} df = \frac{1}{f} \int_0^{\infty} x e^{-x} dx = \frac{1}{f} \int_0^{\infty} -x de^{-x} \\ &= \frac{1}{f} \left[-x e^{-x} \right]_0^{\infty} + \frac{1}{f} \int_0^{\infty} e^{-x} dx \\ &= \frac{1}{f} \left[-x e^{-x} - e^{-x} \right]_0^{\infty} = \frac{1}{f} \end{aligned}$$

Nous pouvons donc affirmer que dans un groupe fermé de véhicules homogènes c'est-à-dire où tous les véhicules ont la même espérance mathématique de f sinistres par an, le temps moyen s'écoulant entre une origine arbitraire et le premier sinistre à survenir sur chacun des véhicules est de $\frac{1}{f}$.

Il est à noter que par raison de symétrie, le temps moyen écoulé depuis le dernier sinistre d'un groupe homogène fermé est également égal à $\frac{1}{f}$.

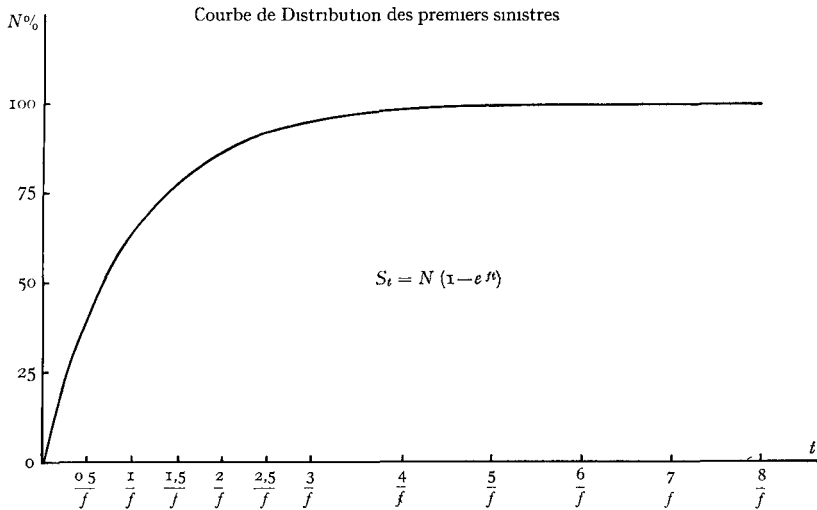


Fig. 1

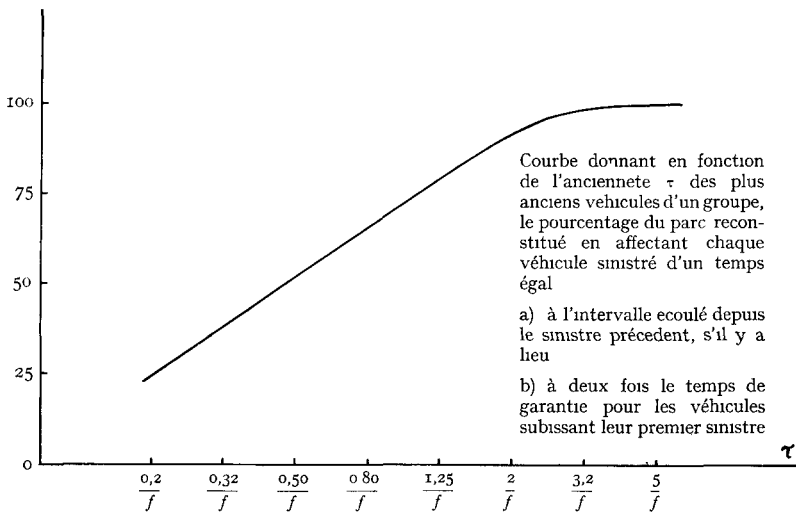


Fig. 2

N.B. — On en déduira que l'intervalle moyen entre le dernier sinistre survenu avant le temps $t = t_0$ et le premier sinistre à venir après le même temps est bien de $\frac{2}{f}$ mais il ne faut pas en conclure

que l'intervalle moyen écoulé entre deux sinistres du même véhicule est $\frac{2}{f}$ mais bien encore $\frac{1}{f}$.

En effet, pour étudier la loi de répartition des intervalles séparant deux sinistres consécutifs nous devons considérer le sous-groupe des véhicules sinistrés le même jour t_0 .

Le sous-groupe reste homogène et sa probabilité de sinistres ultérieurs restant inchangée, sa fréquence de sinistres pour le temps $t > t_0$ reste f .

Nous nous trouvons dans les conditions de l'exemple précédent, le temps moyen s'écoulant pour chacun de ses véhicules entre le sinistre du jour t_0 . et le sinistre suivant reste bien $\frac{1}{f}$.

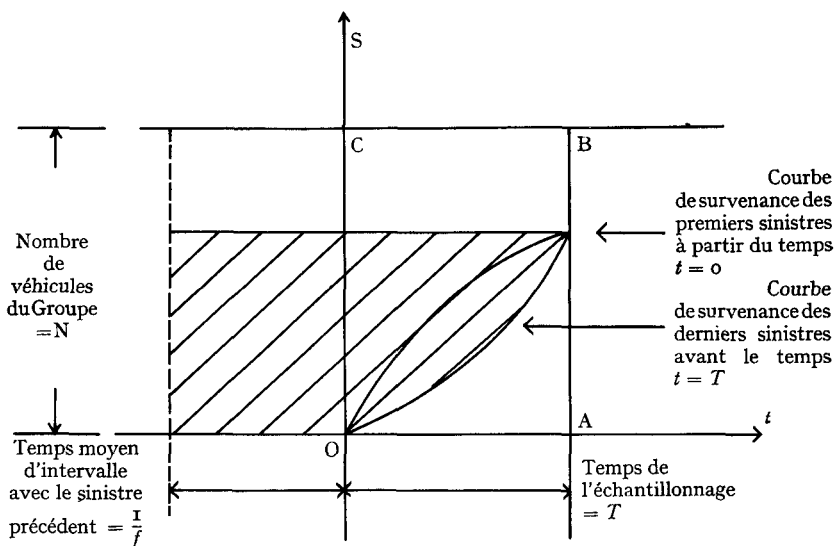


Fig. 3

Le paradoxe vient du fait que nous ne parlons que de premiers sinistres et que le temps moyen séparant deux sinistres est la moyenne des temps séparant des sinistres n et $n - 1$ quelque soit n .

On peut vérifier le raisonnement précédent par la justification graphique suivante:

Nous portons en abscisse le temps et en ordonnée les véhicules

du groupe homogène. Un nombre de véhicules x années est représenté par une aire.

La surface hachurée représente le total des temps écoulés entre chaque sinistre de l'échantillon et le sinistre précédent;

Pour calculer son aire retournons la figure:

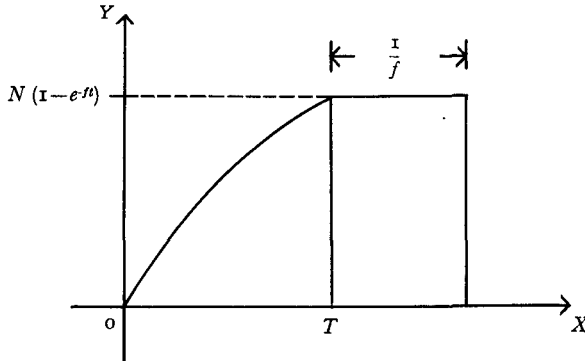


Fig. 4

L'équation de la courbe est

$$y = N (1 - e^{-fx})$$

$$\text{l'aire est } A = \int_0^T N (1 - e^{-fx}) dx + \frac{1}{f} N (1 - e^{-fT})$$

$$A = NT + \int_0^T \frac{N}{f} e^{-fx} dfx + \frac{N}{f} (1 - e^{-fT})$$

$$A = NT + \frac{N}{f} [e^{-fx}]_0^T + \frac{N}{f} (1 - e^{-fT})$$

$$A = NT$$

aire précisément égale au rectangle $O A B C$ représentant le nombre de véhicules x années couverts pendant le temps du sondage.

Nous avons raisonné jusqu'à présent en supposant que le groupe homogène dont fait partie le véhicule sinistré a toujours été dans l'assurance, car dans ce cas le temps moyen entre le sinistre donnant lieu à l'échantillonnage et le sinistre précédent est égal à l'inverse de la fréquence du groupe.

En fait un groupe homogène ayant une ancienneté limitée on trouvera toujours dans un échantillonnage des véhicules n'ayant jamais été sinistrés.

Le problème pour ces véhicules est de trouver la date fictive du sinistre virtuel précédant la prise de garantie.

Considérons dans le temps (en prenant pour origine des temps et pour chaque véhicule le moment du sinistre donnant lieu à l'échantillonnage) la distribution des sinistres précédant celui de l'échantillonnage.

La loi de survenance est celle précédemment établie :

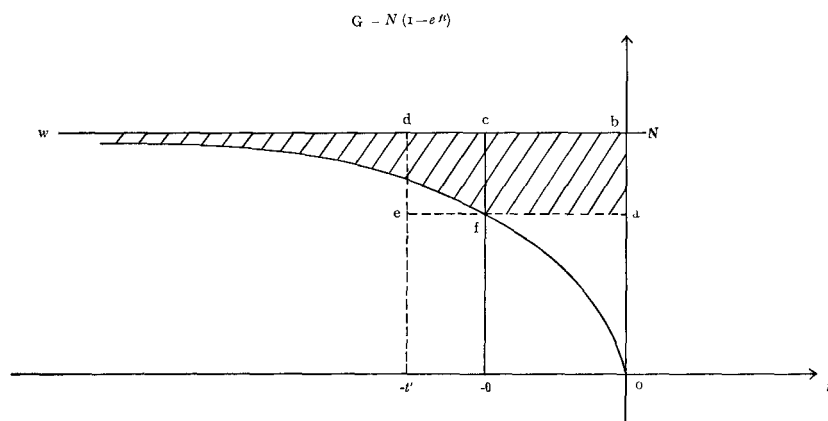


Fig. 5

Considérons un véhicule entré dans l'assurance au temps $-\theta$

a) s'il a été sinistré avant l'échantillonnage, son dernier sinistre précédant celui donnant lieu à l'échantillonnage se trouve en un point de la courbe comprise entre $-\theta$ et o et il n'y a pas de difficulté puisque ce sinistre est déterminé.

b) s'il n'a pas été sinistré, on doit lui affecter un sinistre virtuel sur la partie de courbe située entre $-\theta$ et $-\infty$ et pour chacun des sinistres donnant lieu à échantillonnage, correspondant aux points de segment ab , on fera correspondre un sinistre précédent fictif situé sur l'arc fw

On voit dans cette éventualité qu'il revient au même d'affecter à chacun de ces sinistres un temps t' tel que l'aire A du rectangle $abde$ soit égale à l'aire B du quadrilatère, $abwf$

$$\begin{aligned} \text{or } ab &= Ne^{-f\theta} \\ \text{donc } A &= t' N e^{-f\theta} \\ \text{et } B &= \theta N e^{-f\theta} + \int_{-\infty}^{-\theta} N e^{-ft} dt \\ B &= \theta N e^{-f\theta} + \left[\frac{N}{f} e^{-ft} \right]_{-\infty}^{-\theta} \\ B &= \theta N e^{-f\theta} + \frac{N}{f} e^{-f\theta} \\ \text{et si } A &= B \\ t' N e^{-f\theta} &= \theta N e^{-f\theta} + \frac{N}{f} e^{-f\theta} \\ t' &= \theta + \frac{1}{f} \end{aligned}$$

Or, dans cette expression qui est celle du temps dont il faut affecter chaque véhicule subissant son premier sinistre au cours de l'échantillonnage nous trouvons $\frac{1}{f}$ qui est précisément l'inverse de la fréquence inconnue du groupe dont est extrait le véhicule.

Nous avons démontré précédemment que la moyenne des temps séparant dans un groupe homogène la prise de garantie du premier sinistre survenu à chaque véhicule était précisément égale à l'inverse de la fréquence du groupe considéré.

En première approximation, nous remplacerons dans l'expression précédente $\frac{1}{f}$ par θ

On peut démontrer que cette transformation est valable si dans le parc de véhicules étudié, l'entrée dans l'assurance est linéaire: soit le sous-groupe N_0 des véhicules entrés dans l'assurance au temps $-\theta$

Pendant le temps T du sondage nous sélectionnerons

$N_0 fT$ véhicules dont
 $N_0 e^{-f\theta} fT$ qui n'ont pas été sinistrés antérieurement
et $N_0 (1 - e^{-f\theta}) fT$ qui ont été sinistrés antérieurement.

Les premiers seront affectés du coefficient $\theta + \frac{1}{f}$

Les seconds du temps écoulé entre les deux sinistres consécutifs, soit en moyenne:

$$\frac{\int_0^{\infty} N_{\theta} f \theta e^{-f\theta} d\theta}{N_{\theta} (1 - e^{-f\theta})} = \frac{\frac{1}{f} \left[-f\theta e^{-f\theta} - e^{-f\theta} \right]_0^{\infty}}{1 - e^{-f\theta}} = \frac{1}{f} \frac{1 - f\theta e^{-f\theta} - e^{-f\theta}}{1 - e^{-f\theta}}$$

Nous reconstituerons donc pour le sous-groupe N_{θ} le parc

$$N'_{\theta} = N_{\theta} e^{-f\theta} f T \left(\theta + \frac{1}{f} \right) + N_{\theta} (1 - e^{-f\theta}) f T \frac{1}{f} \frac{1 - f\theta e^{-f\theta} - e^{-f\theta}}{1 - e^{-f\theta}}$$

$$N'_{\theta} = N_{\theta} T f \left[e^{-f\theta} \left(\theta + \frac{1}{f} \right) + \frac{1}{f} (1 - f\theta e^{-f\theta} - e^{-f\theta}) \right]$$

$$N'_{\theta} = N_{\theta} T$$

Supposons N_{θ} constant = $n d\theta$

Nous aurons $\int_0^{\infty} n e^{-f\theta} f T d\theta$ premiers sinistres.

qui reconstitueront avec le coefficient de pondération $\theta + \frac{1}{f}$

$$\begin{aligned} N_0 &= \int_0^{\infty} n e^{-f\theta} f T \left(\theta + \frac{1}{f} \right) d\theta \\ &= n f T \left[\int_0^{\infty} e^{-f\theta} \theta d\theta + \int_0^{\infty} e^{-f\theta} \frac{d\theta}{f} \right] \\ &= \frac{n}{f} T \left[\int_0^{\infty} e^{-f} f \theta df\theta + \int_0^{\infty} e^{-f\theta} df\theta \right] \\ &= \frac{n}{f} T \left(\left[-f\theta e^{-f\theta} - e^{-f\theta} \right]_0^{\infty} - \left[e^{-f\theta} \right]_0^{\infty} \right) \\ &= \frac{2 n T}{f} \end{aligned}$$

Nous aurions obtenu le même résultat en remplaçant $\theta + \frac{1}{f}$ par 2θ

$$\text{car } \int_0^{\infty} n e^{-f\theta} f T 2\theta d\theta = \frac{2 n T}{f}$$

ce qui justifie bien l'hypothèse ci-dessus.

Nous supposons toujours linéaire, l'entrée dans l'assurance des véhicules mais limitée dans le passé au temps τ .

La première expression devient:

$$N_0 = \frac{nT}{f} \left[-f\theta e^{-f\theta} - 2e^{-f\theta} \right]^\tau$$

$$N_0 = \frac{nT}{f} \left(-f\tau e^{-f\tau} - 2e^{-f\tau} + 2 \right)$$

La deuxième expression approchée devient:

$$N'_0 = \frac{nT}{f} \left[-2f\theta e^{-f\theta} - 2e^{-f\theta} \right]^\tau$$

$$N'_0 = \frac{nT}{f} \left(-2f\tau e^{-f\tau} - 2e^{-f\tau} + 2 \right)$$

$$\frac{N'_0}{N_0} = \frac{2 - 2e^{-f\tau} (1 + f\tau)}{2 - e^{-f\tau} (2 + f\tau)} = \frac{1 - e^{-f\tau} (1 + f\tau)}{1 - e^{-f\tau} \left(1 + \frac{f\tau}{2} \right)}$$

$f\tau$	0.1	0.16	0.25	0.40	0.54	1.	1.5	2.5	4
$\frac{N'_0}{N_0}$	0.1	0.15	0.21	0.305	0.45	0.525	0.66	0.875	0.96
$\frac{t'_t}{\theta_t}$	20	13	9.5	6.5	4.5	3.8	3	2.3	2.1

La troisième ligne de ce tableau donne le coefficient dont il faut multiplier le temps d'assurance pour les véhicules subissant leur premier sinistre au cours de l'échantillonnage.

En supposant que l'on néglige cette correction et que l'on se contente d'affecter à chaque véhicule subissant son premier sinistre au cours de l'échantillonnage un temps double de celui de la garantie, et en tenant compte des véhicules déjà sinistrés qui ont un coefficient correct l'erreur commise sur l'évaluation du parc reconstruit est donnée par le tableau suivant:

$f\pi$	0.2	0.32	0.50	0.80	1.25	2	3.2	5
erreur % par défaut	-74	-62	-48	-34	-21	-9	-2	0

On voit que, si un groupe de véhicules a commencé son entrée dans l'assurance depuis un nombre d'années supérieur à $\frac{2}{f}$ l'erreur que l'on commet sur l'évaluation du parc reconstitué suivant la présente méthode est inférieure à 9 %.

Il est à noter que les erreurs qui peuvent en découler sur la détermination des primes ne seront commises que par excès.

En tout état de cause, cette méthode peut être utilisée en toute sécurité pour l'étude de l'influence relative de nouveaux critères sur les primes, si l'on n'a pas une confiance suffisante pour l'utiliser à la détermination des primes absolues.