

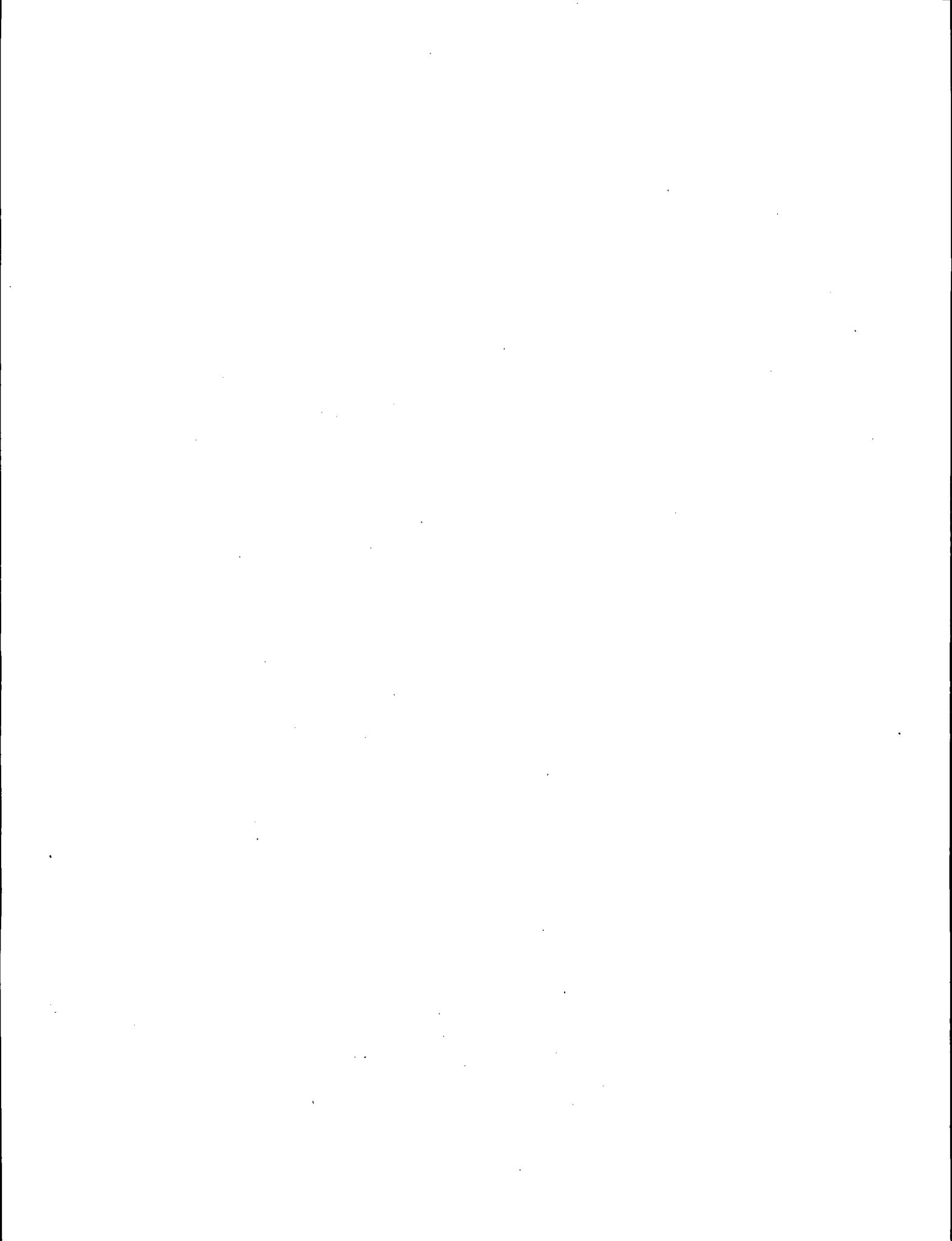
ORGANISME
NATIONAL DE
SÉCURITÉ
ROUTIÈRE

L'ESTIMATION
STATISTIQUE
DE LA PROPENSION
AUX ACCIDENTS

cahiers d'etudes no 8

Bulletin no 24
Décembre 1969

CDAT
15062



L'ESTIMATION STATISTIQUE DE LA PROPENSION AUX ACCIDENTS

D. SROUR

Ingénieur E.C.P. - I.S.U.P.
Docteur de 3ème cycle

RESUME

Des études très nombreuses ont été effectuées sur le problème de la prédisposition aux accidents. Elles se sont le plus souvent bornées à un aspect psychologique ou psychométrique, consistant à sélectionner les conducteurs en fonction de leur "réponses" à des tests. Ces tests ont été choisis de façon à prendre en compte l'ensemble des caractéristiques humaines intervenant dans la survenance des accidents : données biographiques, mesures physiologiques, capacités motrices ou intellectuelles, traits caractériels, etc. L'aspect "prévention" a en général été négligé : les tests opérés ne permettaient d'écarter qu'une trop faible proportion de conducteurs, soi-disant "mauvais", sans pour autant réduire de façon sensible le bilan des accidents.

Les tentatives mathématiques ont été différentes : il ne s'agissait pas de "prévoir" les mauvais conducteurs, mais d'essayer de savoir, compte tenu de l'implication de certains individus dans des accidents, si l'on pouvait vraiment alléguer un quelconque facteur "prédisposition". On avait pour cela besoin, d'abord d'un critère de différenciation des individus entre eux (nombre d'accidents pour une période de temps, degré de responsabilité, etc.), ensuite d'une référence devant servir à la comparaison des soi-disant "prédisposés" aux individus "normaux".

Parti d'objectifs précis, l'observateur statisticien s'est malheureusement trop vite trouvé arrêté ; son art s'est borné à des constatations du type suivant : "deux individus ayant chacun été impliqué dans un sinistre au cours d'une année d'observation sont équivalents ; peut être sont-ils légèrement moins prédisposés qu'un individu ayant encouru 2 accidents pendant la même période de temps avec une exposition au risque identique".

On a pu récemment montrer que l'outil statistique permettait non seulement une classification des individus mais en outre une individualisation de leur risque propre. Encore fallait-il choisir comme critère de différenciation une variable non plus discrète, (comme le nombre d'accidents), mais continue, l'intervalle de temps entre 2 accidents successifs par exemple. La démarche scientifique reste néanmoins la même que celle qui a guidé les tentatives mathématiques, c'est-à-dire essentiellement expérimentale : observer et essayer d'inférer ; aucune supposition n'est faite quant à la valeur dans l'avenir des conclusions éventuelles ce qui ne veut nullement dire qu'une telle position soit a priori exclue.

A l'heure où nous rédigeons ce résumé, les méthodes préconisées dans cette étude ont déjà donné lieu à certaines applications, peu nombreuses, il est vrai. Mais pour la première fois (réf. 29) l'accord est bon entre les méthodes statistiques et psychométriques ; cela ne fait que renforcer la nécessité de poursuivre les recherches dans ces deux domaines.

PREMIERE PARTIE

INTRODUCTION : PRESENTATION DE L'ETUDE

GENERALITES

L'insuffisance des résultats auxquels on aurait pu a priori s'attendre, vu l'immense déploiement d'efforts, partout dans le monde, afin de recueillir le maximum de données sur les accidents, s'explique partiellement par les lacunes laissées en certains domaines, tel celui de l' "Exposition au Risque".

L'accident n'est pourtant que l'aboutissement de chaînes de probabilité qui caractérisent une situation non moins permanente que complexe. Une représentation classique consiste à faire correspondre à cette situation un espace à plusieurs dimensions, divisé en trois sous-espaces caractérisant la Route, le Véhicule, le Conducteur.

C'est du troisième élément de cette "densité de danger" : le Conducteur, qu'il s'agira dans le présent rapport.

LE CONDUCTEUR : SA PREDISPOSITION AUX ACCIDENTS

L'idée d'accidents de tous genres (circulation, aviation, accidents du travail, absentéisme. . .) ne se produisant pas au hasard, mais engageant une responsabilité humaine, n'est pas toute récente puisqu'elle voyait déjà ses prémices dans les célèbres articles de GREENWOOD et WOODS et GREENWOOD et YULE ("the incidence of industrial accidents upon individual with special reference to multiple accidents" 1919 - "an inquiry into the nature of frequency distributions representative of multiple happenings" 1920).

Mis à part l'intérêt purement mathématique de telles théories, on conçoit l'importance qu'elles pouvaient avoir dans la lutte contre toutes les sources d'accidents. Or si le premier point a abouti à des classifications relativement étroites de conducteurs automobiles, dont l'utilité en ce qui concerne les assurances par exemple ne peut être contestée, il semble que le second n'ait pas donné les résultats escomptés a priori. En effet une prévention d'accidents basée uniquement sur l'identification "mauvais conducteurs" a vite fait d'atteindre ses limites d'efficacité puisque cette classe d'individus ne tient finalement pas une aussi grande place qu'on veut volontiers le croire. Le problème primordial restant donc l'immense majorité des "conducteurs moyens", (GOLDSTEIN : "Human variables in traffic accidents")

on concevra facilement le caractère extensif d'individualisation donné à cette étude.

L'ECHEC DES TENTATIVES PSYCHOMETRIQUES

Si parmi les méthodes statistiques destinées à démontrer l'influence de la prédisposition aux accidents, on excepte pour le moment les recherches de corrélation entre les nombres d'accidents pendant différentes périodes, on est porté à rechercher les causes dans le comportement individuel.

Les recherches de corrélations, en effet, il faut bien l'avouer, n'ont point conduit à des résultats significativement différents de ceux qu'on obtiendrait dans l'hypothèse du hasard (Mc FARLAND : Opinions courantes américaines sur le concept de la prédisposition aux accidents), et c'est d'ailleurs bien pour cette raison qu'une prévention routière centrée sur les récidivistes et menée à l'aide de dossiers "passé du mauvais conducteur" devenait inefficace (PRESTON - HARRIS : Psychology of drivers in traffic accidents).

Mais avant d'en aboutir aux études psychologiques des caractéristiques individuelles, qu'est-ce que le comportement ?

"Le comportement est le reflet d'habitudes prises par l'homme" : cette phrase a elle seule ne justifie-t-elle pas les thèses récentes (1960) sur le thème "l'homme conduit comme il vit" et selon lesquelles c'est dans les habitudes les plus normales qu'il faut chercher les causes d'accidents ?

"Mais ce n'est pas que cela : c'est en même temps le reflet de la maîtrise de soi, de la faculté que possède un être de freiner ses impulsions" : il semblerait donc qu'en plus d'une étude générale de comportement humain on doive accorder une attention toute particulière aux caractéristiques individuelles directement en rapport avec les reflexes de conduite nécessaires.

"Si l'on peut classer les êtres suivant les habitudes de leur milieu peut-on dire pour cela que la plus ou moins grande maîtrise de soi caractérise un milieu" ? : y-a-t-il donc des corrélations suffisantes permettant de structurer les individus en sous-groupes relatifs à leurs caractéristiques de comportement habituel ?

Ces trois phrases d'ARIBAUD dans "Quelques remarques sur l'assurance automobile" résument les tentatives faites jusqu'alors dans le domaine psychométrique. Il semble intéressant d'en donner un bref aperçu.

I – LES “ATTITUDES SCALES” (tests de comportement individuel)

On a vu depuis une vingtaine d'années se développer plusieurs de ces tests se proposant par l'analyse de certaines caractéristiques individuelles et l'attribution de notes relatives à une référence précise, de chiffrer la prédisposition aux accidents aussi bien qu'aux infractions routières (ex. : Siebrecht attitude scale, Rosenzweig picture frustration study, Drivers attitude scale, Mc Guire safe driver scale, Guilford-Zimmerman temperament survey, Dynamic factors opinion survey, Guilford-Holley Inventory, Allport-Vernon-Findzey study of values, etc.).

Or non seulement ces tests ne donnaient des corrélations valables que pour une faible proportion des caractéristiques humaines étudiées, mais il arrivait en plus que le même échantillon de conducteurs testé suivant deux échelles différentes (ex. La “Driver attitude scale” ou DAS et la “Mc Guire safe driver scale” ou MSDS) ne donnât pas des corrélations suffisantes.

Fallait-il alors en déduire l'incapacité de telles méthodes? Probablement pas puisque les deux tests pouvaient d'abord ne pas avoir la même valeur dans la prédiction des accidents ensuite et inversement donner des prévisions comparables en mesurant des facteurs différents mais **dont la contribution à la responsabilité serait la même.**

L'idée s'imposait alors d'élargir le champ d'investigation des tests en étudiant aussi bien les fonctions physiologiques, les capacités sensorielles et psychomotrices que les données biographiques et les traits caractéristiques de la personnalité. Il ne faut donc pas s'étonner des études exhaustives telles que celles de SCHUSTER et GUILFORD (The psychometric prediction of problem drivers) couvrant le large éventail précité à partir de plusieurs choix de caractères tirés de tests plus spécialisés.

Mais comment expliquer alors que de telles théories n'aient pas toujours conduit à des résultats sinon complets du moins suffisants? A ce propos il ne faut pas oublier que ce que cherchent à mesurer ces tests, ce sont **les caractères stables de l'individu** c'est-à-dire **des tendances persistantes.** Or indépendamment de l'influence du véhicule et de l'environnement routier sur les accidents, et si l'on s'en rapporte au seul individu, il faut outre ces caractères stables tenir compte de ce que l'on pourrait appeler **les caractéristiques transitoires ou momentanées** ainsi que de certains facteurs (perte de contrôle par exemple) qui n'ont pratiquement pas de liaison avec la prédisposition aux accidents. C'est

sur le concept de “liaison” que nous mettrons l'accent dans ce qui suit.

II – LES ETUDES DE REGRESSION (analyse corrélative)

A mi-chemin entre la psychologie appliquée et les mathématiques, la psychométrie a largement mis à profit les méthodes statistiques de régression multiple.

Rappelons à ce sujet, qu'il s'agit d'écrire les variations à prévoir (nombre d'accidents ou infractions par exemple) sous la forme de combinaisons linéaires des caractéristiques humaines contribuant “suffisamment” à expliquer le phénomène.

Une étude très complète en a été donnée (The psychometric prediction of problem drivers by D.H. SCHUSTER and J.L. GUILFORD) et a conduit à partir de 40 variables étudiées à des expressions telles que :

$$\text{Accident} = 0,06 \text{ viol.} + 0,12 \text{ Em}/2 + 0,17 \text{ car.} \\ \text{cond.} + 0,15 \text{ Yrs drv.} + 0,21 \text{ Ann. Mil.} + 0,04 \text{ P} \\ + 0,08$$

avec

Viol. : number of moving violations in preceding two years (ne sont pas prises en compte les contraventions dues au stationnement interdit).

Em/2 : number of employers in preceding two years.

Car. Cond : mechanical condition of car usually drive.

Yrs driv. : number of years testes had been driving.

Ann. Mil. : annual mileage driven.

P. : personal relations.

Des remarques particulièrement suggestives peuvent en être déduites ; les variables étant en effet classées par contribution décroissante à la corrélation il faut noter :

- l'importance considérable du facteur “infractions routières” qui justifie la programme de lutte anti-accidents à partir de la répression des contrevenants.
- le peu d'intérêt relatif du facteur “kilométrage annuel moyen” et cela, contrairement à la croyance généralement répandue qui considère ce caractère comme prépondérant. Il ne faut pas s'en étonner puisque si ce facteur augmente d'un côté l'exposition au risque, il apporte de l'autre une contribution favorable puisque le conducteur gagne en expérience (GOLDSTEIN “that the relation ship between miles driven and violations or accidents is not very high”).

- l'absence du facteur "accidents" dans le groupe de variables de prévision ce qui montre bien que la corrélation est pratiquement insignifiante entre 2 nombres d'accidents pendant des périodes consécutives.
- l'absence aussi du facteur "caractéristiques humaines physiques" (ex. : acuité visuelle, temps de réaction, réflexes, etc.).
- l'importance par contre du facteur psychologique Em/2 censé donner une idée de l'instabilité de l'individu.

Il est en outre réconfortant de noter que ces conclusions sont en accord avec celles de GOLDSTEIN (l'auteur mettant en particulier l'accent sur la contribution notable des facteurs tels que : irresponsabilité sociale, agressivité) ainsi qu'avec celles de BRODY ("Personal characteristics of chronic violators") où la comparaison de deux groupes de conducteurs - les "bons" et les "mauvais" - choisis de façon aussi homogène que possible (même exposition au risque, même revenu, même éducation, etc.) aboutit :

- à l'importance du facteur "stabilité professionnelle"
- à la contribution insignifiante des fonctions psychophysiques (acuité visuelle, temps de réaction, champ de vision, etc.), par opposition à celle prépondérante des fonctions sociopsychologiques (intolérance, agressivité, impulsivité, manque de responsabilité, etc.).

En conclusion donc, si les essais de régression multiple ont été en général satisfaisants, ceux de psychologie proprement dite ont été relativement décevants : insuffisance, subjectivité, inadaptivité des tests, attitude défensive des récidivistes testés nuisant à la sincérité exigée en telle matière, etc. les raisons en sont multiples. Mais elles se résument toutes par le fait que "l'humain est difficile à mettre en équations".

Les tentatives mathématiques

L'origine en est les célèbres articles précités (GREENWOOD, WOODS, YULE 1919-1920) dont l'originalité consistait à remarquer la non-conformité de la représentation d'une population hétérogène d'agents par des survenances poissoniennes d'accidents à paramètre constant (ARBOUS et KERRICH : Accident statistics and the concept of accident-proneness).

L'idée qui consistait à isoler l'inclination d'un individu aux accidents (le paramètre de POISSON suivant lui-même une loi de PEARSON du type III) porta alors de gros espoirs.

On sait aujourd'hui la déception qui s'ensuivit : malgré la découverte de la méthode du maximum de vraisemblance en 1921 (FISHER : On the mathematical foundations of theoretical statistics) les paramètres de la loi binomiale négative (résultant de la combinaison POISSON-PEARSON III) furent longtemps estimés avec fort peu d'efficacité par la "Méthode des moments" dont le tort le plus grave était d'ailleurs d'entraîner par le test du χ^2 , le rejet fréquent d'une hypothèse vraie (HALDANE : The fitting of binomial distributions). Le problème ne se pose plus actuellement puisque la méthode du maximum de vraisemblance a été appliquée avec succès à la loi binomiale négative (SICHEL : The estimation of the parameters of a negative binomial distribution).

Si l'on en arrive maintenant aux écueils pratiques de telles études, il faut avouer que les hypothèses simplificatrices trop souvent admises sans aucun soin préalable de justification (exemple : constance de la responsabilité individuelle dans le temps), la restriction de l'analyse à une fraction seulement des données disponibles, l'exclusion tout à fait arbitraire enfin des groupes d'individus à taux d'accidents élevés, etc. ont longtemps entraîné, d'abord un gaspillage dans l'utilisation des méthodes statistiques conventionnelles, ensuite - et ce qui est plus grave - des distorsions dans les résultats.

Pour progresser donc dans l'étude du facteur individuel d'accident, il est nécessaire d'établir des méthodes statistiques qui, outre leur support par des bases mathématiques indispensables :

- soient centrée sur l'individu (le "passé d'accidents" ou "accident record" d'une personne ne devrait pas avoir d'influence sur l'analyse statistique des accidents de toute autre personne).
- utilisent autant que possible le maximum de données disponibles.
- soient capables de tester l'hypothèse d'une prédisposition constante.
- permettent enfin une évaluation statistique "sûre" de ce "potentiel individuel d'accident".

Partant de ces conclusions fondamentales SICHEL a été amené à proposer un modèle d'estimation de la réticence (et non de la propension) aux accidents par l'intervalle moyen entre les différents accidents dans lesquels un même individu a été impliqué (The statistical estimation of individual accident liability). Appliqué à un cas pratique de conducteurs d'autobus le modèle semble avoir donné des résultats valables sur le plan de la pré-

vention (SHAW-LYNETTE et SICHEL : The reduction of accidents in a transport company by the determination of the accident liability of individual drivers. SPANGENBERG : The use of projective tests in the selection of bus drivers).

Il paraît alors intéressant :

- de donner un aperçu de l'ensemble des méthodes mathématiques et statistiques appliquées à l'étude de la prédisposition ;
- de rappeler les résultats principaux du modèle de l'intervalle de temps moyen et de préciser son champ d'application.
- de proposer un autre estimateur de la réticence tenant compte lui, de toute l'information disponible (1).
- d'affiner le modèle par l'introduction des facteurs : "responsabilité individuelle d'implication" et "indice de gravité de l'accident" (en l'occurrence le coût ou la désutilité). De la notion d'intervalle de temps moyen, (\bar{T}) on aboutirait ainsi graduellement à celle "d'intervalle virtuel de temps moyen, à responsabilité maximale (\bar{T}^*)" puis à celle "d'intervalle virtuel de temps moyen à responsabilité maximale et par unité monétaire de désutilité (\bar{T}^{**})". Plus qu'un simple classement, ou une structuration péniblement homogène, on peut donc facilement parler là d'individualisation.
- d'étudier l'apport de telles méthodes pour la création d'un fichier des conducteurs.

DEUXIEME PARTIE

METHODES MATHÉMATIQUES ET STATISTIQUES

INTRODUCTION : SPECIFICATION DU LANGAGE UTILISÉ

L'imbrication du monde actuel avec l'Univers du Risque, la présence bien définie de l'individu en tant que membre d'une société et sa liaison inextricable avec le monde environnant, font que de plus en plus, des accidents se produisent.

(1) Ce point sera clarifié dans la suite.

Quel que soit son mode de participation à cet Univers, (ouvrier de production, conducteur ou simplement piéton) l'individu possède un "potentiel d'accidents", qui est la résultante de 2 composantes : l'une centrée sur son individualité propre, l'autre afférente aux inévitables risques extérieurs du monde dans lequel il évolue.

La confusion qui peut naître du vocabulaire jusqu'alors utilisé pour ce genre d'études, fait que déjà, des précisions s'imposent :

- le "potentiel d'accidents" dont on vient de parler pourra être indifféremment dénommé : propension ou susceptibilité aux accidents (en anglais : liability) ;
- la première composante de ce potentiel est dite : facteur individuel d'accident ou prédisposition aux accidents (en anglais : proneness) ;
- quant à la seconde on la désigne couramment par : risques extérieurs ou environnants, composante environnante ou universelle, etc.

L'implication d'un agent exerçant une certaine activité dans l'Univers du Risque ne pouvant être saisie que par ses "performances d'accidents" sur des périodes d'exposition, **le facteur individuel ne sera valablement dégagé, que si l'étude porte sur un groupe de personnes supposées subir en moyenne les mêmes risques extérieurs.** En particulier, tous les tableaux qui seront présentés dans ce qui suit, supposeront vraie l'équivalence de la composante universelle pour les divers individus considérés. Mais on voit déjà les controverses abondantes que peut faire naître l'acceptation ou inversement le refus de l'égalité du risque extérieur, pour la population considérée. Ceci explique d'ailleurs en partie les positions très diverses et souvent opposées des spécialistes en la matière.

D'un autre côté, et dans le cas heureux où l'on atteint le facteur individuel, il faut noter qu'il est souvent intéressant d'opérer la subdivision en 3 prédispositions : stable (1), transitoire et instantanée correspondant respectivement : aux caractères profonds inséparables de l'individu, à son état du moment (euphorique, soucieux, etc.) bref à un état où "l'individu ne dispose pas de l'énergie optimale à l'accomplissement de son activité" (2), ou enfin aux affections d'état spontanées

(1) Ce mot est pris ici dans le sens : "inséparable de l'individu".
(2) Phénomène dont le signal d'alarme est constitué par ce qu'on appelle les "presque accidents" (near-accidents ou almost-accidents).

et de brève durée (presque instantanées) (exemple : conducteur qui relâche son attention pendant quelques secondes fatales ; sautes d'humeur entraînant des accidents. . .).

La prédisposition stable se rapporte aussi bien aux facteurs pouvant changer avec le temps (âge, expérience. . .) qu'à ceux restant constants pendant l'exposition au risque (prédisposition stable constante) ; ceci entraîne naturellement – à moins que l'on n'utilise des coefficients de corrections calculés à l'aide des statistiques importantes – de nouvelles difficultés de classement de la population étudiée en fonction du critère "prédisposition stable variable" en vue d'aboutir à la **prédisposition stable constante** qui nous intéresse le plus.

Ces précisions successives, en vue d'arriver en matière de prédisposition à un facteur relativement "sensible" et permettant éventuellement des considérations prédictives, restreignent du fait de leurs exigences d'homogénéité des classes (homogénéité des risques extérieurs, des prédispositions variables) le champ d'étude du statisticien et ne permettent pas à cause des faibles effectifs concernés, d'aboutir à des conclusions universellement valables.

On ne s'étonnera donc plus des diverses positions en la matière, certains auteurs niant sinon la prédisposition aux accidents, du moins sa valeur en tant que critère de classement (1), d'autres au contraire, l'acceptant comme caractéristique inséparable de l'individu et se hasardant même jusqu'à lui imposer des conditions de constance dans le temps. Actuellement, il semble que l'on s'oriente vers l'admission courante de l'hypothèse de prédisposition, cela cependant avec des restrictions très larges :

- la courbe d'inégalité donnant la proportion d'accident en fonction de la proportion d'agents doit être significativement plus prononcée que celle obtenue dans l'hypothèse de répartition au hasard (2).
- il est hors de question d'admettre une prédisposition constante en fonction du temps.

(1) A moins que l'on ne considère de longues périodes successives, la corrélation entre les 2 nombres successifs d'accidents relatifs à un même individu est en effet faible et ne justifie donc pas que l'on puisse classer un individu pour une période future à partir de son "passé d'accidents".

(2) La courbe de concentration (courbe de Gini) est en effet assez prononcée même dans le cas d'absence de prédisposition. C'est donc en regard à cette "concentration rémanente" que doit être jugée la courbe d'inégalité relative au cas étudié. Ce point sera repris plus loin en détail.

- la modification de classement consécutive à l'observation d'un individu pendant une certaine période d'exposition (exemple du changement de tarification d'un assuré après enregistrement de son passé d'accidents) doit être considérée moins comme justifiée par la valeur prédictive du passé d'accidents que comme une "punition" ou une "récompense" (serait-ce une morale de force ?).

Pour terminer, il faudrait attirer l'attention sur un point généralement négligé dans la littérature : la **prédisposition et les risques extérieurs ne peuvent être considérés comme indépendants**. D'où la nécessité d'introduire un troisième terme représentant les interactions entre ces 2 facteurs : certaines personnes peuvent en effet être particulièrement prédisposées dans le cas de risques extérieurs précis (1).

Mais les conclusions partielles obtenues par certains auteurs affirmant que les individus prédisposés l'étaient, quel que soit l'univers du risque environnant, ont relégué au second plan, l'étude pourtant très importante de la de la définition de la prédisposition de façon absolue, ou inversement, relativement au risque extérieur encouru.

Dans tout ce qui suit, nous supposons donc que la prédisposition est étudiée **en fonction de risques extérieurs bien définis**.

I – L'HYPOTHESE DE L'HOMOGENEITE DE LA POPULATION EXPOSEE AU RISQUE

Dans le cas où l'on suppose que tous les individus exposés au risque sont "également" sujets à des accidents éventuels ("pure chance distribution of accidents" autrement dit : absence de prédisposition individuelle inhérente mais évidemment, présence inévitable des risques extérieurs d'accidents), on peut démontrer que l'on a affaire à une survenance poissonnienne d'événements.

Peut-être est-il d'abord nécessaire de rappeler que l'on s'intéresse à des événements :

- ayant peu de chance d'arriver en une circonstance précise (unlikely to occur on any given occasion).
- ayant les mêmes probabilités de se produire en des "circonstances équivalentes" (equally likely on all occasions).

(1) Par exemple, un conducteur automobile ayant subitement un malaise aux yeux et étant ébloui par les phares d'une autre voiture venant à cet instant précis en direction opposée, illustre parfaitement bien l'interaction : prédisposition instantanée - risques extérieurs.

- pouvant survenir assez souvent sur une longue durée d'exposition (with many opportunities for occurring).

Ceci étant dit, considérons une population exposée au risque pendant un intervalle de temps (0, T) que l'on aura préalablement subdivisé en k sous-intervalles

$$\delta t_i = t_{i+1} - t_i \text{ avec } t_0 = 0 \text{ et } t_k = T.$$

Dans ce qui suit, on appellera :

- X_i la variable aléatoire : "nombre d'accidents par individu pendant la durée δt_i " (il est évident que l'on peut parler d'une telle notion puisque l'on suppose les différentes personnes également sujettes aux accidents).

$$- p_i(x_i) = \text{Prob}(X_i = x_i)$$

$$+ p(\vec{x}) = p(x_0, x_1, \dots, x_{k-1}) = \text{Prob}(X_0 = x_0 ; X_1 = x_1 ; \dots ; X_{k-1} = x_{k-1}).$$

L'hypothèse implicite de "pure chance" est équivalente à une indépendance interne de la succession des variables aléatoires considérées, indépendance qui s'écrit :

$$p(\vec{x}) = \prod_{i=0}^{k-1} p_i(x_i) \quad \text{et qui s'énonce :}$$

"Pour un même individu, les nombres successifs d'accidents x_0, x_1, \dots, x_{j-1} survenus pendant les j premières périodes n'influent en rien sur les nombres futurs $x_j, x_{j+1}, \dots, x_{k-1}$, à attendre dans les k-j périodes subsidiaires".

Si l'on considère donc un intervalle quelconque δt_i , supposé assez petit pour que l'on doive vraisemblablement s'attendre à :

- plusieurs personnes ayant 0 accident
- peu de personnes ayant 1 accident
- très peu en ayant plus d'un.

on peut écrire pour un individu quelconque :

$$p_i(0) = 1 - f(t_i) \delta t_i - o(\delta t_i)$$

$$p_i(1) = f(t_i) \delta t_i + o(\delta t_i)$$

$$p_i(x_i) = o(\delta t_i)^{x_i - 1} \text{ pour } x_i > 1$$

où $f(t_i)$ est très proche de la densité de probabilité d'avoir 1 accident dans la période δt_i .

Le calcul de la fonction caractéristique du nombre total X d'accidents sur l'intervalle (0, T) permet d'aboutir à la conclusion suivante : l'hypothèse d'homogénéité de la population quant à une prédisposition aux accidents (qui équivaut en fait à une absence de prédisposition c'est-à-dire à une survenance des accidents au hasard) entraîne automatiquement une distribution poissonnienne du nombre d'accidents par individu pour une période quelconque d'exposition au risque.

Pour qui connaît le processus de Poisson, les hypothèses implicites préalables avaient déjà permis de se douter du résultat puisque :

- Le "unlikely to occur on any given occasion" correspond à une faible probabilité d'accident pendant un petit intervalle de temps t (probabilité équivalente à $o(\Delta t)$).
- le "equally likely to occur on all occasions" correspond au processus sans mémoire, les probabilités d'arrivées d'événements dépendant de la longueur des intervalles considérés mais non du repérage de ces intervalles.
- le "with many opportunities for occurring" enfin, à une moyenne d'accidents $\lambda(T)$ croissante avec la période d'exposition T.

II - L'INSUFFISANCE DE L'HYPOTHESE D'HOMOGENEITE DE LA POPULATION

Mais il ne semble pas que l'hypothèse d'homogénéité de la population ait été vérifiée par les observations pratiques d'accidents. A ce sujet, retenons que pour tester une telle hypothèse, on disposait de 2 tests du χ^2 ...

- le test du χ^2 a priori consistant à se donner une valeur pour λ , à en déduire les fréquences théoriques et à effectuer le test classique.

– le test du χ^2 a posteriori consistant à estimer λ par la moyenne des observations et à effectuer le test en conséquence.

Le deuxième de ces tests est évidemment plus en faveur d'un non-rejet de l'hypothèse de distribution au hasard (intuitivement, il consiste en effet à tester seulement la forme de la loi alors que le premier teste simultanément et la forme et la moyenne de la loi).

Bien que les conclusions d'un test du χ^2 ne puissent ainsi être considérées comme strictes que pour des tests a priori, les tests a posteriori ont souvent été effectués et la non-adéquation obtenue n'en a que plus incité à l'abandon de l'hypothèse d'homogénéité.

Plusieurs exemples de cette non-adéquation peuvent être donnés. Les deux premiers sont pris au domaine des assurances automobiles (DELAPORTE : Un problème de tarification de l'assurance accidents d'automobile examiné par la statistique mathématique) les deux derniers à celui des accidents du travail (GREENWOOD and WOODS : the incidence of industrial accidents upon individuals).

EXEMPLE 1 : Pour une classe du tarif automobile (usage, promenade) et un modèle particulier de véhicule, on a observé en 1958 les nombres suivants de véhicules ayant eu 0, 1, 2, 3... sinistres dans l'année.

● λ est estimé par
$$\frac{\sum n_j x_j}{\sum n_j} = \frac{565}{1744} = 0.324$$
 et les

nombres théoriques sont calculés par :

$$n'_j = \frac{1744 \times e^{-0.324} (0.324)^j}{j!}$$

● Pour le calcul du χ^2 et à cause des faibles effectifs théoriques dans certains cas, on a opéré à un regroupement des 4 dernières classes.

On voit que le χ^2 obtenu dépasse largement le χ^2 tolérable à 2 degrés de liberté (4 classes et 1 paramètre estimé) même pour un seuil 1 ‰ largement inférieur au seuil habituel de 5 % (λ limite de 13.82).

Nombre de sinistres	Nombre observé de véhicules	Nombre de sinistres observés	Nombre théorique (ajusté par la loi de Poisson)	Contribution au χ^2
x_j	n_j	$n_j x_j$	n'_j	$(n_j - n'_j)^2 / n'_j$
0	1316	0	1261.3	2.37
1	323	323	408.7	17.97
2	81	162	66.2	24.75
3	18	54	7.2	33.65
4	4	16	0.6	
5	2	10	0.0	
≥ 6	0	0	0.0	
TOTAUX	1744	565	1744.0	78.74

EXEMPLE 2 : Pour une autre classe de tarif automobile (usage affaires) et un modèle particulier du véhicule, la répartition des nombres probables de sinistres par véhicule en 1958, a été la suivante :

- Même processus de regroupement en classe consécutif aux faibles valeurs de certains des effectifs théoriques calculés après estimation de λ par $856/1334 = 0.641$.
- Là aussi, grande valeur de χ^2 qui se passe évidemment de commentaires.

Nombre de sinistres	Nombre observé de véhicules	Nombre de sinistres observés	Nombre théorique de véhicules	Contribution au χ^2
x_i	n_i	$x_i n_i$	n'_i	$(n_i - n'_i)^2/n'_i$
0	774	0	702.2	7.34
1	375	375	450.6	12.68
2	120	240	144.6	4.19
3	40	120	30.9	2.68
4	15	60	5.0	20.00
5	5	25	0.6	123.56
6	2	12	0.1	
7	1	7	0.0	
8	1	8	0.0	
9	1	9	0.0	
TOTAUX	1334	856	1334.0	170.45

EXEMPLE 3 : Sur 648 femmes travaillant pendant 5 semaines ouvrables, on a observé la répartition suivante d'accidents :

- $\lambda = \frac{303}{648} = 0.467$
- 63.25 est évidemment beaucoup plus grand que le χ^2 limite tolérable dans le cas d'homogénéité et pour un seuil de 1 ‰ soit 16.27 (3 degrés de liberté).

Nombre d'accidents	Nombre observé d'ouvrières accidentées	Nombre d'accidents observés	Nombre théorique d'ouvrières accidentées	Contribution au χ^2
x_i	n_i	$n_i x_i$	n'_i	$(n_i - n'_i)^2/n'_i$
0	447	0	406	4.14
1	132	132	189	17.19
2	43	86	45	0.09
3	21	63	7	28.00
4	3	12	1	13.83
≥ 5	2	10	0.1	
TOTAUX	648	303	648.1	63.25

EXEMPLE 4 : Le relevé des accidents survenant à des machinistes observés pendant 3 mois a donné le tableau suivant :

- $\lambda = 0.483$
- χ^2 tolérable : 16.27.

Nombre d'accidents	Nombre observé de machinistes accidentés	Nombre d'accidents observés	Nombre théorique de machinistes accidentés	Contribution au χ^2
x_i	n_i	$n_i x_i$	n'_i	$(n_i - n'_i)^2 / n'_i$
0	296	0	256	6.25
1	74	74	122	18.88
2	26	52	30	0.53
3	8	24	5	1.80
4	4	16	1	72.00
≥ 5	6	34	0.1	
TOTAUX	414	200	414.1	99.46

III – UN PALLIATIF A L'HETEROGENEITE DE LA POPULATION : LA DECOMPOSITION EN SOUS-POPULATIONS HOMOGENES

Devant l'impossibilité d'admettre un nombre moyen d'accidents constant pour tous les individus considérés, on se pose le problème de savoir s'il ne serait pas possible de décomposer la population globale étudiée en un ensemble de sous-populations dont les membres puissent être considérés comme étant homogènes entre eux.

Partons d'un nombre n d'individus dont m_0 ont eu 0 accident, m_1 1 accident, etc.

Il a été vu précédemment que le λ estimé par

$$\frac{\sum i m_i}{\sum m_i}$$

ne permettrait pas après calculs des effectifs théoriques d'approcher suffisamment les résultats observés.

Il s'agirait donc maintenant de décomposer le groupe étudié en k sous-groupes d'effectifs n_1, n_2, \dots, n_k et de moyennes $\lambda_1, \lambda_2, \dots, \lambda_k$ tels que :

- chaque sous-groupe obéisse à la loi de Poisson
- et que la somme des résultats partiels reproduise le plus fidèlement possible la loi expérimentale.

D'où la recherche d'un système d'équations à $2k$ inconnus :

$$(n_j, \lambda_j) \quad j = 1, 2, \dots, k$$

On peut démontrer que les λ_j sont solutions de l'équation :

$$\begin{vmatrix} 1 & \lambda & \lambda^2 & \dots & \lambda^k \\ n & M(1) & M(2) & \dots & M(k) \\ M(1) & M(2) & \dots & \dots & M(k+1) \\ \dots & \dots & \dots & \dots & \dots \\ M_{(k-1)} & M(k) & \dots & \dots & M(2k-1) \end{vmatrix} = 0$$

et que les n_j en sont déduits à l'aide du système :

$$\sum_{j=1}^k n_j \lambda_j^i = M(i) \quad i = 0, 1, \dots$$

où $M(i)$ est le moment factoriel d'ordre "i".

Mais pour que la décomposition ait un sens il faudra :

- que l'équation en λ ait toutes ses racines réelles et positives
- que les valeurs n_j soient également positives.

A ces conditions seulement, on pourra admettre que la population hétérogène d'individus accidentés, est équivalente à la somme de k sous-populations homogènes d'indices 1, 2, ... k , le sous groupe i étant caractérisé par n_j individus ayant chacun un nombre d'accidents poissonien de moyenne constante, c'est-à-dire indépendante de la personne considérée.

Du point de vue pratique, il s'agira évidemment de pousser la décomposition en augmentant successivement le nombre k de sous-groupes jusqu'à aboutir à une adéquation valable avec les observations expérimentales (1).

EXEMPLE : Pour illustrer cette méthode reprenons le tableau 4 relatif aux accidents survenus à des machinistes, en y développant la case : "plus de 5 accidents".

Nombre d'accidents	Observations	Nombre d'accidents observés
0	296	0
1	74	74
2	26	52
3	8	24
4	4	16
5	4	20
6	1	6
7	0	0
8	1	8
TOTAUX	414	200

(1) A la limite ($k = 1$), il est intéressant de remarquer que le déterminant se réduit à :

$$\begin{vmatrix} 1 & \lambda \\ n & M(1) \end{vmatrix} = 0 \text{ soit } \lambda = \frac{M(1)}{n} = \frac{\sum i n_i}{n}$$

même valeur que celle que l'on obtient dans le test d'homogénéité de la population ; ceci constitue en quelque sorte une vérification de la méthode proposée.

et supposons qu'il s'agisse de décomposer en 2 sous-groupes homogènes ($k = 2$)

- a - les 2 moyennes λ_1 et λ_2 des nombres individuels d'accidents pour chaque sous-groupe seront obtenus grâce à la résolution ou l'équation :

$$\begin{vmatrix} 1 & \lambda & \lambda^2 \\ n & M(1) & M(2) \\ M(1) & M(2) & M(3) \end{vmatrix} = 0$$

avec

$$n = 414 ; M(1) = 200 ; M(2) = 314 ; M(3) = 840$$

Cette équation s'écrit :

$$89996 \lambda^2 - 284960 \lambda + 69404 = 0 \text{ et admet pour solutions les valeurs } 0.266 \text{ et } 2.900.$$

Les effectifs n_1 et n_2 de chaque sous-population se calculent en prenant les 2 premières équations du système précédent à savoir :

$$n_1 + n_2 = 414$$

$$0.266 n_1 + 2.9 n_2 = 200 \text{ soit}$$

$$n_1 = 380$$

$$n_2 = 34$$

environ

D'où la décomposition en 2 sous-populations :

I

x_i	n_i	$n_i x_i$	n'_i	$(n_i - n'_i)^2 / n'_i$
0	296	0	296.78	0.002
1	74	74	73.30	0.006
2	10	20	9.83 ⁽²⁾	0.003
TOTAUX	380	94	379.91	0.011

$$\lambda_1 = \frac{94}{380} = \underline{\underline{0.247^{(1)}}}$$

(1) Du fait du remplacement des n_1 et n_2 fractionnaires obtenus après résolution des équations précédentes, par les chiffres arrondis 380 et 34, un nouveau calcul des λ_1 s'impose.

On obtient d'ailleurs des valeurs sensiblement différentes de celles trouvées après résolution initiale de l'équation consistant à annuler le déterminant (0.266 et 2.900).

II

x_i	n_i	$n_i x_i$	n'_i	$(n_i - n'_i)^2 / n'_i$
2	16	32	13.44 ⁽²⁾	0.48
3	8	24	7.59	0.02
4	4	16	5.92	0.62
5	4	20	3.69	0.02
6	1	6	1.92	
7	0	0	0.85	0.39
8	1	1	0.33	
TOTAUX	34	106	33.74	1.53

$$\lambda_2 = \frac{106}{34} = \underline{\underline{3.12^{(1)}}}$$

d'effectifs 380 et 34 dont les tableaux ci-joints donnent la répartition empirique et théorique ainsi que les très faibles χ^2 correspondant, indiquant ainsi une excellente homogénéité interne à chaque sous-population.

b - De même, si l'on voulait décomposer en 3 sous-groupes homogènes, on devrait résoudre l'équation.

$$\begin{vmatrix} 1 & \lambda & \lambda^2 & \lambda^3 \\ n & M(1) & M(2) & M(3) \\ M(1) & M(2) & M(3) & M(4) \\ M(2) & M(3) & M(4) & M(5) \end{vmatrix} = 0$$

avec $M(4) = 2616$ et $M(5) = 7920$.

Cette équation s'écrit :

$$\Delta_0 - \lambda \Delta_1 + \lambda^2 \Delta_2 - \lambda^3 \Delta_3 = 0$$

avec

$$\Delta_0 = \begin{vmatrix} 200 & 314 & 840 \\ 314 & 840 & 2616 \\ 840 & 2616 & 7920 \end{vmatrix} = -0.265 \times 10^8$$

$$\Delta_1 = \begin{vmatrix} 414 & 314 & 840 \\ 200 & 840 & 2616 \\ 314 & 2616 & 7920 \end{vmatrix} = -0.86 \times 10^8$$

(2) Pour éviter de manipuler des lois de Poisson tronquées supérieurement (tableau I) ou inférieurement (tableau II) à la valeur 2, ces chiffres théoriques représentent les effectifs des individus ayant :

- au moins 2 accidents (tableau I)
- au plus 2 accidents (tableau II).

$$\Delta_2 = \begin{vmatrix} 414 & 200 & 840 \\ 200 & 314 & 2616 \\ 314 & 840 & 7920 \end{vmatrix} = 0.27 \times 10^8$$

$$\Delta_3 = \begin{vmatrix} 414 & 200 & 314 \\ 200 & 314 & 840 \\ 314 & 840 & 2616 \end{vmatrix} = 1.8 \times 10^8$$

soit $-0.265 + 0.86 \lambda + 0.27 \lambda^2 - 1.8 \lambda^3 = 0$ ou encore

$$\boxed{\lambda^3 - 0.15 \lambda^2 - 0.47 \lambda + 0.14 = 0}$$

En l'écrivant sous sa forme réduite :

$\mu^3 - 0.48 \mu - 0.024 = 0$ avec $\lambda = \mu + h$ où h est choisi de façon à annuler le coefficient de μ^2 ($h = 0,05$), on trouve :

- une première racine approchée (obtenue par approximations successives) :

$$\mu_1 = 0.725 \quad \text{soit} : \lambda_1 = 0.775$$

- les 2 autres racines résultent de la résolution de :

$P_2(\mu) = 0$ où $P_2(\mu)$ est le le polynôme du second ordre quotient de la division de $(\mu^3 - 0.48 \mu - 0.024)$ par $(\mu - 0.725)$.

Approximativement $P_2(\mu)$ s'écrit $\mu^2 + 0.725 \mu + 0.046$ et ses racines μ_2 et μ_3 sont négatives ; cela donne après calcul :

$$\lambda_2 > 0 \quad \text{et} \quad \lambda_3 < 0.$$

Du fait de l'existence d'une racine négative (λ_3), il n'est donc pas possible de décomposer la population initiale en 3 sous-populations homogènes.

On pourrait ici poursuivre l'étude pratique en reprenant la décomposition de chacun des tableaux précédents ; mais il semble plus important de tirer les conclusions d'une telle méthode en insistant sur les faits suivants :

- complexité des calculs (nécessité en particulier de programmes sur ordinateur pour ce qui est de la recherche des zéros du polynôme de degré k en λ soit $P_k(\lambda)$ dès qu'on cherche une décomposition en plus de 3 sous-groupes ; or ce cas peut se présenter assez fréquemment, en particulier lorsque la population est étudiée sur une longue durée d'exposition, ce qui donne ainsi à l'aléa "nombre d'accidents" un champ de variation assez large).
- l'existence éventuelle de zéros négatifs pour le polynôme $P_k(\lambda)$ constitue un obstacle aux décompositions à k fixé a priori ; de telles subdivisions peuvent être inspirées par un choix de politique envers le personnel, ou plus simplement par une simple intuition consécutive à un premier examen des statistiques d'accidents (industriels ou autres...).

— le cas heureux où la décomposition est possible entraîne presque toujours le classement en des sous-groupes différents (mais limitrophes) de certaines personnes ayant eu le même nombre d'accidents pendant la période d'exposition (cf. exemple 4 où la classe : "2 accidents" se trouve scindée en deux).

Cette lacune dans le domaine de l'individualisation du risque sera comblée par le paragraphe qui suit.

IV — A LA LIMITE, INDIVIDUALISATION DE LA MOYENNE D'ACCIDENTS ; LA LOI BINOMIALE NEGATIVE (L.B.N.)

Si à la limite on considère chaque individu comme formant un sous-groupe à lui tout seul ($k = n$), λ devient alors une variable aléatoire prenant toutes les valeurs réelles (non négatives).

Si l'on suppose par exemple que λ suit une loi normale tronquée aux valeurs négatives (c'est-à-dire que l'on considère comme négligeable la probabilité que λ soit inférieur à zéro) on trouve pour un individu quelconque la probabilité d'avoir x accidents :

$$p_x = \text{prob. (x accidents)} = \frac{e^{-1/2(\sigma^2 - 2m)} I_x}{\sigma x! \sqrt{2\pi}} \quad \text{avec}$$

$$I_x = (m - \sigma^2) I_{x-1} + \sigma^2 (x-1) I_{x-2}, \text{ équation}$$

récurrente permettant de calculer les valeurs successives I_2, I_3, \dots à partir des deux données initiales :

$$I_0 = \sigma \sqrt{2\pi} \quad \text{et} \quad I_1 = \sigma \sqrt{2\pi} (m - \sigma^2).$$

Notons-en néanmoins les quelques résultats partiels :

$$p_0 = e^{-1/2(\sigma^2 - 2m)}$$

$$p_1 = (m - \sigma^2) e^{-1/2(\sigma^2 - 2m)}$$

$$p_2 = \frac{((m - \sigma^2)^2 + \sigma^2) e^{-1/2(\sigma^2 - 2m)}}{2}$$

etc.

Mais cette hypothèse de normalité est contredite par le fait que λ est supposé varier asymétriquement. Il existe évidemment plusieurs courbes de probabilité asymétriques dont le choix a priori est tout à fait arbitraire. Mais la loi de Pearson du type III semble particulièrement intéressante du fait qu'elle ne conduit pas à des calculs inextricables. On verra d'ailleurs plus loin que ce choix est justifié.

Prenons donc :

$$f(\lambda) = \frac{e^{-c\lambda} \lambda^{p-1} c^p}{(p-1)!}$$

La probabilité précédente p_x s'écrit alors :

$$p_x = \frac{(x+p-1)!}{x!(p-1)!} \frac{c^p}{(1+c)^{x+p}}$$

expression de la loi binomiale négative dont il est utile de rappeler qu'elle peut être matérialisée par le tirage avec remise de p boules blanches en exactement $(x+p)$ coups, les proportions de boules blanches et noires étant

$$\text{respectivement } \frac{c}{c+1} \quad \text{et} \quad \frac{1}{c+1}.$$

Les caractéristiques de cette loi (espérance mathématique, moment du 2ème ordre et variance) sont classiques :

$$E(X) = p/c$$

$$E(X^2) = \frac{p(c+p+1)}{c^2}$$

$$V(X) = \mu_2(X) = \frac{p(c+1)}{c^2}$$

et il est maintenant utile de voir dans quelle mesure, cette nouvelle loi approche les résultats expérimentaux donnés dans les tableaux précédents. Il faut donc, à partir de l'estimation des paramètres p et c de cette loi, recalculer les effectifs théoriques et réappliquer le test du χ^2 .

Estimation des paramètres par la méthode des moments

Pour la simplification des calculs, on a fait un changement

de paramètre sous la forme $\frac{p}{c} = m$ ce qui donne :

$$p_x = \frac{\left[\frac{p}{p+m} \right]^p}{\left[\frac{p}{p+m} \right]^p} \frac{(x+p-1)!}{(p-1)! x!} \left[\frac{m}{p+m} \right]^x$$

Bien que non indispensable, ce changement de paramètres est cependant utile du fait que l'on a ainsi

$E(X) = m$; la moyenne \bar{x} des accidents est ainsi un estimateur sans biais de m .

En appelant m et p les estimateurs cherchés et en égalant les 2 premiers moments théoriques aux moments empiriques correspondants, on trouve :

$$E(X) = \frac{p}{c} = m = \sum_{i=1}^n \frac{x_i}{n} = \bar{x} \text{ d'où : } m = \bar{x}$$

$$V(X) = \mu_2(X) = \frac{p(c+1)}{c^2} = \mu_2'' \text{ où } \mu_2'' = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

n est le moment centré empirique du 2ème ordre qu'au facteur multiplicatif $\frac{n-1}{n}$ près, mais présente cependant l'avantage d'être un estimateur sans distorsion de la variance.

En remplaçant c par $\frac{p}{\bar{m}} = \frac{p}{\bar{x}}$, on tire de cette équation :

$$\bar{p} = \frac{\bar{x}^2}{\mu_2'' - \bar{x}}$$

On peut alors dans les exemples 3 et 4 précédents recalculer les effectifs théoriques et la valeur des nouveaux χ^2 .

EXEMPLE 3 :

Nombre d'accidents	effectifs observés	effectifs calculés	Contribution au χ^2
0	447	442	0.05
1	132	140	0.45
2	43	45	0.09
3	21	14	3.50
4	3	5	0.80
≥ 5	2	2	0.00
	648	648	4.89

$$\bar{m} = 0.467 \quad \bar{p} = 0.960$$

EXEMPLE 4 :

Nombre d'accidents	Effectifs observés	Effectifs calculés	Contribution au χ^2
0	296	299	0.03
1	74	69	0.36
2	26	26	0.00
3	8	11	0.81
4	4	5	0.20
5	4	2	2.00
≥ 6	2	2	0.00
	414	414	3.40

$$\bar{m} = 0.483 \quad ; \quad \bar{p} = 0.442$$

Le résultat est une bonne adéquation de la loi binomiale négative puisque les χ^2 obtenus ne sont pas significativement grands.

V – GENERALISATION DE LA LOI BINOMIALE NEGATIVE A PARTIR DE LA LOI DE PEARSON A 3 PARAMETRES

La Loi de Pearson de type III à 2 paramètres a été choisie comme densité de la probabilité de λ grâce :

- à sa forme asymétrique
- au fait que les observations empiriques ont montré une distribution caractérisée par plusieurs individus ayant une faible prédisposition et progressivement de moins en moins en ayant de plus grandes.

On peut néanmoins décider d'introduire un troisième paramètre λ_0 représentant un risque minimum : (1)

$$f(\lambda) \text{ devient alors : } \frac{e^{-c(\lambda - \lambda_0)} c^p (\lambda - \lambda_0)^{p-1}}{(p-1)!}$$

(1) Il est utile de rappeler ici que la propension dont on parle est la résultante :

- de la prédisposition (facteur individuel d'accidents)
- et des risques extérieurs d'accidents relatifs à l'activité effectuée

λ_0 correspondrait ainsi au cas idéal de prédisposition inhérente nulle ce qui n'empêche évidemment en rien l'existence pratiquement inévitable des risques extérieurs.

Les nouvelles probabilités p_x ont été calculées par M. DELAPORTE :

$$p_x = \frac{e^{-\lambda_0} c^p}{x! (p-1)! (1+c)^{x+p}} \sum_{i=0}^x e_x^i [\lambda_0(1+c)]^i$$

$$(x+p-i-1)!$$

Par exemple : $p_0 = e^{-\lambda_0} \left(\frac{c}{1+c}\right)^p$; $p_i = e^{-\lambda_0} \left(\frac{c}{1+c}\right)^p \frac{(p+\lambda_0(1+c))}{(1+c)}$

De même que précédemment et en vue cette fois de transformer les tableaux des exemples 1 et 2, on estime les 3 paramètres λ_0 , p et c par la "Méthode des Moments". Il faut pour cela être en possession des 3 moments théoriques m_1 , μ_2 , μ_3 qu'il s'agit d'égaliser aux 3 moments empiriques : m'_1 , μ'_2 , μ'_3 .

On trouve après calcul : (DELAPORTE : Un problème de tarification de l'assurance "accidents d'automobiles" "examiné par la statistique mathématique") :

$$m_1 = \lambda_0 + p/c$$

$$\mu_2 = \lambda_0 + \frac{p(1+c)}{c^2}$$

$$\mu_3 = \lambda_0 + \frac{p}{c} \left(1 + \frac{3}{c} + \frac{2}{c^2}\right)$$

EXEMPLE 1

Nombre d'accidents	Nombre observé	Nombre calculé	Contribution χ^2
0	1316	1315.6	0.000
1	323	324.7	0.008
2	81	78.7	0.067
3	18	19.0	0.052
4	6	6.0	0.000
TOTAUX	1744	1744.0	0.127

$$m'_1 = 0.324 \quad \bar{\lambda}_0 = 0.124$$

$$\mu'_2 = 0.423 \quad \bar{p} = 0.4$$

$$\mu'_3 = 0.720 \quad \bar{c} = 2.0$$

ce qui permet de déduire du système :

$$m_1 = m'_1$$

$$\mu_2 = \mu'_2$$

$$\mu_3 = \mu'_3$$

les paramètres λ_0 , p et c en fonction des moments empiriques.

$$D'où les résultats : \bar{\lambda}_0 = m'_1 - \frac{-2(\mu'_2 - m'_1)^2}{(\mu'_3 - 3\mu'_2 + 2m'_1)}$$

$$\bar{p} = \frac{4(\mu'_2 - m'_1)^3}{(\mu'_3 - 3\mu'_2 + 2m'_1)^2}$$

$$\bar{c} = \frac{2(\mu'_2 - m'_1)}{(\mu'_3 - 3\mu'_2 + 2m'_1)}$$

Le calcul des nouveaux effectifs théoriques dans les tableaux (1) et (2) par la formule :

nombre observé X $p_x(\lambda_0, p, c)$, fournit alors les nouveaux χ^2 .

EXEMPLE 2

Nombre d'accidents	Nombre observé	Nombre calculé	Contribution au χ^2
0	774	783.4	0.112
1	374	349.8	1.815
2	120	131.3	0.972
3	40	46.2	0.832
4	15	15.7	0.031
5	5	5.2	0.007
6	5	3.36	0.800
TOTAUX	1334	1334.96	4.569

$$m'_1 = 0.641 \quad \bar{\lambda}_0 = 0.35$$

$$\mu'_2 = 1.01 \quad \bar{p} = 0.23$$

$$\mu'_3 = 2.68 \quad \bar{c} = 0.79$$

La bonne adéquation obtenue (χ^2 à 1 degré de liberté dans l'exemple 1 et à 3 degrés dans l'exemple 2, car il y a 4 relations : 3 pour l'estimation des paramètres et 1 pour l'effectif global) infirme le fait qu'il faut bien parler d'une prédisposition aux accidents et que la propension d'accidents correspondante est bien représentée par une loi de Pearson du Type III à 2 ou 3 paramètres.

IV – LES INSUFFISANCES DE LA METHODE D'ESTIMATION PRECEDENTE LA METHODE DU MAXIMUM DE VRAISEMBLANCE (M.V.)

Quand on parle d'estimation, on pense assez généralement à la méthode du Maximum de Vraisemblance qui présente le grand intérêt de fournir des estimateurs asymptotiquement efficaces, c'est-à-dire à variance minimum (on sait en effet que cette méthode permet parfois d'atteindre la borne inférieure de la variance d'un estimateur, fournie par l'inégalité de Fréchet-Darmonis-Cramer-Rao et l'on démontre en plus que si à partir du M.V. on n'atteint pas la borne inférieure de la variance, il est illusoire de chercher à "faire mieux" à partir de toute autre méthode).

Le problème ne s'était pas toujours posé de rechercher type d'estimateurs car l'on traitait surtout des lois normales où les propriétés d'efficacité ("efficiency") se trouvaient vérifiées par les estimateurs trouvés à partir de la "Méthode des Moments".

Dans le cas présent par contre, l'utilisation du test du χ^2 à partir d'estimateurs non efficaces risque d'entraîner le rejet d'une hypothèse vraie (cf. plus loin) et il est par conséquent très utile de trouver les estimateurs du M.V.

Pour la loi binomiale négative, de tels estimateurs m et p , ont été calculés par FISHER et SICHEL qui trouvent :

$$\hat{m} = \bar{x}$$

$$\hat{p} \text{ solution de } \frac{1}{n} \sum_i [\text{Dig}(\hat{p} + x_i - 1) - \text{Dig}(\hat{p} - 1)] - \text{Log}\left(1 + \frac{\bar{x}}{\hat{p}}\right) = 0$$

équation où la fonction Digamma (Dig) symbolise la dérivée logarithmique de la fonction gamma d'Euler et qui permet de calculer p à l'aide d'un processus d'approximations successives utilisant les valeurs numériques tabulées de la fonction :

$$\beta(x, \hat{p}) = \text{Dig}(\hat{p} + x - 1) - \text{Dig}(\hat{p} - 1)$$

Il est maintenant intéressant de voir pratiquement comment les 2 méthodes d'estimation peuvent entraîner des résultats opposés dans un test du χ^2 . On considère à cet effet, des observations d'absentéisme résumées dans le tableau suivant :

EXEMPLE 5

Nombre d'absences	Observations	Effectifs calculés par la méthode		Contribution au χ^2	
		des moments	du M.V.	Méthode des moments	M.V.
0	217	206.7	214.9	0.513	0.020
1	44	60.6	53.4	4.547	1.655
2	29	25.9	23.4	0.371	1.340
3	11	12.3	11.8	0.137	0.054
4	11	6.1	6.3		
5	2	3.1	3.5	1.620	0.431
6	4	1.6	2.0		
7	0	1.7	2.7		
TOTAUX	318	318.0	318.0	7.188	3.500

Pour 5 classes et 2 paramètres estimés, on a 2 degrés de liberté et le χ^2 limite correspondant à un seuil de 5 % est de 5.99 ce qui montre bien que dans le cas d'une estimation par le "Méthode des Moments", on est conduit à rejeter une hypothèse que l'estimation du M.V. fait apparaître comme exacte ($7.188 > 5.990 > 3.500$).

VII – CONCLUSION

Il a été ainsi fait état de tentatives mathématiques successives, donnant lieu à des méthodes de plus en plus fines, destinées à mettre en évidence la prédisposition individuelle aux accidents. On sait combien le problème est important tant en ce qui concerne l'analyse des causes d'accidents que la mise en forme d'un contrôle préventif éventuel. C'est d'ailleurs en vue de résoudre ce problème par la détection et l'élimination des individus fortement prédisposés, qu'il semble que les chercheurs aient négligé des questions de fond telle celle d'une définition précise du terme "prédisposition" en particulier en ce qui concerne sa dépendance à l'égard des risques extérieurs. Les interprétations "optimistes" des phénomènes ont ainsi assez souvent caché les difficultés fondamentales et il n'est pour cela que de rappeler les conclusions hâtives d'existence de la prédisposition, consécutives à un simple examen de la concentration des accidents alors que l'on connaît pertinemment bien l'inévitable "concentration rémanente", relative au cas de distribution des sinistres au hasard ("pure chance model").

A titre d'exemple, pour 500 personnes ayant 300 accidents, la répartition au hasard (loi de Poisson) serait de 275, 165, 49, 10, 1 personnes ayant respectivement 0, 1, 2, 3, 4 accidents ; d'où une concentration de 44 % des accidents pour 12 % de la population.

Il est d'ailleurs intéressant de noter, que dans le cadre des exemples 1, 2, 3, 4, les pourcentages 6%, 13.8 %, 10.6 % et 10.6 % de la population correspondent respectivement aux proportions 42 %, 56.2 %, 56.4 %, 63 % des accidents ; pour les mêmes exemples, les calculs se rapportant aux effectifs théoriques donnent : 27.7 % des accidents concernant 4.2 % de la population 47.4 %, 13.6 %, 37.9 %, 8.2 % et 39.5 %, 8.7 %.

La mise en regard pour chaque exemple des points "pratique" et "théorique" appartenant à la courbe de concentration ne fournit donc pas d'incompatibilité, ce qui prouve bien qu'on ne peut ainsi se suffire de ces chiffres pour conclure à l'existence de la prédisposition. Il faudrait pour cela, construire dans chaque cas la courbe de concentration rémanente, calculer les points de la courbe de concentration observée et juger si ces points sont "suffisamment" à l'extérieur du domaine limite par la courbe rémanente et l'axe des abscisses. Il reste évidemment à définir exactement, en l'occurrence par un test à bâtir, ce qu'on entend par "courbe significativement plus concentrée qu'un autre".

L'adéquation d'une loi théorique de prédisposition n'est elle-même pas suffisante puisque la précision de l'ajustement réalisé est étroitement liée à l'étendue de la

période d'exposition ; en outre, la loi binomiale négative ou plus généralement une distribution théorique quelconque des accidents, peut résulter d'un ensemble d'assertions assez différentes entre elles, concernant les hypothèses initiales sur la prédisposition aux accidents (1). D'où, préalablement à l'acceptation de l'hypothèse de prédisposition, la nécessité impérieuse de rejeter toute autre hypothèse pouvant éventuellement entraîner la même forme de distribution théorique des sinistres.

Le problème est donc bien plus compliqué qu'on ne pense et l'idéal pour les chercheurs serait évidemment d'étudier une population non sélectionnée de futurs agents (conducteurs, pilotes, etc.) qui seraient ainsi tous testés et obligés d'enregistrer les accidents même les plus bénins.

Il reste à se demander, si en dehors d'un état fortement policé, un tel "paradis de statisticiens", serait un rêve souhaitable ou un cauchemar.

TROISIEME PARTIE

LE MODELE DE L'INTERVALLE DE TEMPS MOYEN

I – INTRODUCTION LES INSUFFISANCES ANTECEDENTES

La conclusion précédente a évoqué rapidement les problèmes les plus épineux, problèmes auxquels les chercheurs depuis 1920 jusqu'à nos jours, n'ont cessé de se mesurer. La gravité des contradictions soulevées justifié pleinement un temps d'arrêt, d'abord nécessaire à une récapitulation concise, ensuite indispensable à l'évaluation de l'intérêt attaché aux nouveaux modèles qui seront proposés dans la suite.

1 – Non seulement, une bonne adéquation de la L.B.N. n'est pas un indicateur irréfutable de la prédisposition aux accidents, mais encore, la correspondance des observations avec une loi poissonnienne à paramètre

(1) Si l'on considère par exemple les 2 assertions suivantes :

- les individus ont initialement des propensions différentes qu'ils conservent tout au long de leur exposition
- les individus ont la même propension à l'origine mais acquièrent progressivement des différences mutuelles consécutives à l'expérience de l'activité exercée,

on se rend compte que toutes deux conduisent à la même distribution théorique des accidents et qu'il est impossible de savoir laquelle des 2 interprétations est meilleure.

constant, n'exclut pas forcément l'existence éventuelle de différences individuelles dans les propensions aux sinistres. Certains auteurs ont donc proposé de substituer aux distributions univariates de fréquences souvent jugées imprécises et par trop relatives, des distributions bivariates consistant à relever des effectifs d'accidentés n_{12} ayant été impliqués dans x_1 et x_2 sinistres, correspondant à deux périodes consécutives d'exposition au risque (le plus souvent de même amplitude).

1.1. — Si l'on appelle (1) et (2) les périodes d'exposition, X_1 et X_2 les variables aléatoires "nombres d'accidents", x_1 et x_2 les réalisations correspondantes, $p_1(x_1)$ et $p_2(x_2)$ les probabilités qui y sont attachées et $p(x_1, x_2)$ la probabilité d'implication dans x_1 et x_2 sinistres pendant les deux périodes consécutives, on peut trouver des lois bivariates dont les distributions marginales sont poissonniennes :

$p(x_1, x_2) = f(\lambda_1, \lambda_2, x_1, x_2)$ avec :

$$p_1(x_1) = \sum_{x_2} p(x_1, x_2) = \frac{e^{-\lambda_1} \lambda_1^{x_1}}{x_1 !}$$

$$p_2(x_2) = \sum_{x_1} p(x_1, x_2) = \frac{e^{-\lambda_2} \lambda_2^{x_2}}{x_2 !}$$

L'ajustement de telles lois à des distributions n_{12} d'accidents ne différant pour aucune des deux périodes du "pure chance model" (loi de POISSON) (1), a permis d'obtenir de très forts coefficients de corrélation pour les populations bivariates considérées ; ceci est évidemment en contradiction flagrante avec la conclusion d'absence de prédisposition consécutive à la bonne adéquation du "pure chance model".

Même dans le cas favorable où la distribution univariate sur la période (1) + (2) permettrait, de par l'inégalité de la moyenne à la variance, d'accepter l'existence de prédisposition, l'approche corrélationnelle du phénomène conserve tout son intérêt, la conclusion relative à l'existence ou non de différences dans les propensions aux accidents, risquant de varier avec la longueur de la période d'exposition dont l'optimum est a priori, très difficile à choisir.

(1) Le test pratique consiste à vérifier l'égalité de la moyenne à la variance.

1.2. — Venons-en maintenant à un bon ajustement de la L.B.N. sur des observations expérimentales d'accidents. Va-t-on en déduire immédiatement l'existence de la prédisposition ? Rien n'est moins sûr du moment que l'on a pu construire des distributions hypothétiques de fréquences, qui malgré leur bonne représentation par une L.B.N. donnaient, une fois leur séparation en deux périodes d'exposition effectuée, de très faibles coefficients de corrélation de la population bivariate.

L'explication théorique du phénomène n'est d'ailleurs pas difficile à trouver : la construction de la distribution binomiale négative bivariate et non corrélée à partir de la L.B.N. classique

$$p(x) = C_{x+p-1}^x \left(\frac{p}{p+m}\right)^p \left(\frac{m}{p+m}\right)^x$$

$$\text{s'écrit : } p_i(x_i) = C_{x_i+p-1}^{x_i} \left(\frac{p}{p+m}\right)^{p/2} \left(\frac{m}{p+m}\right)^{x_i} \quad i=1,2$$

et il n'y a a priori aucune raison pour que le paramètre p représentant l'ordre de la loi de PEARSON du type III change, selon qu'il corresponde à une exposition initiale, ou à son fractionnement en deux.

2 — De là à en conclure que l'approche par corrélation du phénomène est "the most direct method of establishing proneness in a group of people, all of whom to be exposed to the same environmental risk", il n'y a qu'un pas. . . qu'il faut se garder de franchir allègrement.

2.1. — La méthode par corrélation présente un biais systématique ("skew nature") dû au fait que la simple inclusion d'un faible groupe de "sinistres inguérissables" (1) peut faire rapidement varier le coefficient de corrélation de la population bivariate et le rendre ainsi significatif ; la justification pratique et intuitive des choses en est naturellement difficile à trouver car on voit mal comment cette inclusion peut raisonnablement influencer sur la susceptibilité réelle de la population concernée.

2.2. — Quoi qu'on en pense, l'originalité de cette méthode ne résiste pas à un examen profond : il est en effet relativement simple de montrer, aussi

(1) ou "accident repeaters" : individus ayant beaucoup d'accidents au cours des deux périodes consécutives d'exposition au risque.

bien mathématiquement que pratiquement, que ce moyen d'approche se réduit à une méthode univariante, consistant à chiffrer le "degré d'existence du phénomène de prédisposition" (other than chance factor) dans un ensemble d'observations par la racine carrée du rapport $\frac{\text{Variance-Moyenne}}{\text{Variance}}$, qui évidemment, dans le

cas du "pure chance model", devrait être nul.

2.2.1 – **Mathématiquement** : avec les mêmes notations utilisées en 1.1, on peut écrire

$$p(x_1, x_2) = p(x_1 + x_2) C_{x_1 + x_2}^{x_1} (1/2)^{x_1 + x_2}$$

avec $p(x_1 + x_2) = P(X = x_1 + x_2)$

On en déduit :

$$E(X_1) = E(X_2) = \frac{E(X)}{2}$$

$$E(X_1^2) = E(X_2^2) = \frac{E(X^2) + E(X)}{4}$$

$$E(X_1 X_2) = \frac{E(X^2) - E(X)}{4}$$

D'où

$$V(X_1) = V(X_2) = \frac{V(X) + E(X)}{4}$$

$$r = \frac{E(X_1 X_2) - E(X_1) E(X_2)}{V(X_1)} = \frac{V(X) - E(X)}{V(X) + E(X)}$$

r' (correction de SPERMAN-BROWN) =

$$\frac{2r}{1+r} = \frac{V(X) - E(X)}{V(X)}$$

qui n'est rien d'autre que le rapport introduit dans 2.2.

2.2.2 – **Pratiquement** : les calculs de ces deux coefficients au vu d'observations expérimentales ont fourni des résultats très proches et le problème du choix entre ces deux méthodes, n'est plus alors qu'une question de commodités, commodité de calcul ou même de préférences personnelles.

2.3. – Quant aux contre-exemples bâtis pour réfuter les équivalences absolues : POISSON \leq absence de prédisposition et L.B.N. \leq prédisposition, s'ils sont mathématiquement possibles, il n'en reste pas moins que pratiquement, ils se présentent comme hautement improbables.

2.3.1 – Dans le premier cas (POISSON \leq absence de prédisposition), si les distributions poissonniennes de sinistres résultent de susceptibilité individuelle constante, elles ne devraient pas, théoriquement, donner lieu à une corrélation significative ; il est alors naturellement très difficile de voir par quel concours fortuit de circonstances, on aboutit à une corrélation valable pour la population bivariate.

2.3.2 – Dans le deuxième cas (L.B.N. \leq prédisposition) l'équivalence n'a pu être rejetée qu'en imaginant certaines personnes ayant un très grand, puis un très petit nombre d'accidents pour les périodes successives considérées ; on admettra donc volontiers l'improbabilité de telles hypothèses.

2.4. – Du point de vue opérationnel, il faut maintenant noter que la méthode par corrélation peut parfois se révéler difficile à appliquer.

Tout en étant un problème statistique important, la prédisposition aux accidents est avant tout un problème pratique de sécurité et il serait regrettable d'oublier qu'il s'agit bien moins d'attendre des périodes successives en invoquant la nécessité de conclusions statistiquement plus sûres (ce qui d'ailleurs reste à démontrer), que de réduire effectivement le nombre des accidents.

Le cas fréquent est d'ailleurs celui où cette attente induit une distorsion systématique dans l'analyse statistique, l'observation des mêmes individus pendant deux périodes différentes n'allant pas de soi, puisque les plus accidentés ont tendance pour les raisons les plus diverses (licenciement, hospitalisation, mort), à ne plus se retrouver dans l'échantillon correspondant à la deuxième période d'exposition.

2.5. – Il semblerait donc, que malgré sa légitimité et l'intérêt nouveau qu'elle suscite, cette méthode, n'étant pas à tout considérer plus sûre que les approches univariates du phénomène, ne soit pas vraiment indispensable.

3 – En fait, ne serait-ce que la nécessité d'exposition au risque identique pour tous les individus, en vue d'en tirer des déductions statistiques "sûres", il apparaît, suite aux inconvénients soulevés dans 2.4 qui sont valables aussi bien pour l'approche univariante que bivariate, qu'aucune de ces deux méthodes n'est vraiment concluante. D'où la nécessité de moyens nouveaux utilisant des méthodes statistiques centrées sur l'individu, exhaustives, (c'est-à-dire opérant à partir de toutes les données disponibles), et permettant aussi bien l'obtention d'estimateurs sûrs de propensions, que des tests d'hypothèses de constance de susceptibilité aux accidents.

C'est dans cette optique que se place le modèle de l'intervalle de temps moyen.

II – FREQUENCES D'ACCIDENTS OU INTERVALLES DE TEMPS ENTRE LES SINISTRES CONSECUTIFS

1 – La question de savoir lequel des deux indicateurs : “fréquences ou intervalles de temps” sera le support de l'analyse statistique mérite d'être posée.

Si l'étude des fréquences a jusqu'alors été la plus courante (cas des tarifications d'assurance par exemple), il ne faut pas oublier que la “matière première” statistique, c'est-à-dire les données de base disponibles, est constituée par une suite ordonnée d'intervalles temporels séparant les accidents consécutifs.

Il peut d'ailleurs sembler futile et vain d'argumenter sur ces deux moyens d'étude, puisqu'ils sont en fait équivalents dès que le taux moyen d'accident peut être considéré comme constant dans le temps. Mais il n'empêche que l'on peut parler d'une “meilleure méthode” lorsqu'il s'agit de certains problèmes précis.

2 – Pour ce qui est de notre étude particulière, il nous faut sans réserve conclure à la supériorité du critère “intervalle de temps”.

2.1. – Le fait que ce soit une variable continue permet d'envisager assez simplement des ajustements de diverses distributions statistiques aux observations effectuées ; le problème est plus complexe lorsqu'il s'agit de la variable discrète qu'est la fréquence d'accidents par période référentielle de temps. Dans le dernier cas, il faut en outre, attendre l'échéance de cette période d'exposition pour tirer des conclusions statistiques valables. (Il est vrai que l'on pourrait fractionner la durée d'observation en plusieurs périodes et extrapoler les résultats obtenus en premier lieu à l'ensemble des intervalles successifs d'étude. Mais le nombre d'observations s'en trouverait considérablement réduit et les méthodes statistiques perdraient ainsi de leur acuité).

2.2. – D'un autre côté, les intervalles temporels successifs sont plus aptes à juger d'améliorations dans les survenances de sinistres occasionnées par l'expérience de l'agent soumis au risque ou inversement de détériorations résultant de causes très diverses ; ce jugement est d'ailleurs très important à effectuer dans des délais rapides, d'abord pour des raisons internes de gestion dans le cas d'entreprises industrielles ou de transport par exemple, ensuite pour l'appréciation de l'efficacité de campagnes ou de mesures éventuelles de sécurité.

III – LE MODELE DE L'INTERVALLE DE TEMPS MOYEN

Les données réclamées par le modèle de l'intervalle de temps moyen, qui sera explicité par la suite, sont donc pour chaque individu :

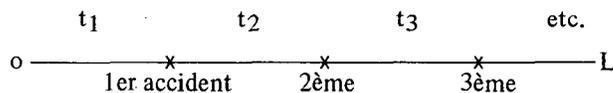
- la durée L de l'exposition au risque (mesurée en unités de temps, ou éventuellement en “kilomètres parcourus” dans le cas des conducteurs par exemple).
- le nombre r d'accidents (on supposera que L est assez grand pour qu'il s'en produise suffisamment).
- la date de survenance de chacun de ces accidents.

On pourra donc en déduire :

- les intervalles de temps entre les accidents successifs (le dernier de ces intervalles, c'est-à-dire celui qui va jusqu'à l'échéance de L n'étant pas pris en compte).

– l'intervalle de temps moyen $\bar{T} = \frac{t_1 + t_2 + \dots + t_r}{r}$ où t_1

est mesuré à partir du début d'exposition.



Si la personne considérée est indéfiniment exposée, on sait que la loi de probabilité de chaque intervalle est du type “gamma 1” :

$\frac{1}{\theta} e^{-t/\theta}$, où le rapamètre θ est la moyenne d'une

infinité d'intervalles successifs. C'est ce θ , inversement proportionnel à la propension aux accidents, qui caractérisera cette dernière, une grande valeur de θ indiquant une faible propension et inversement.

Les différents problèmes qui se posent sont donc : la validité de la loi exponentielle précédente, l'étude probabiliste et statistique du modèle.

A – TESTS D'ADEQUATION DE LA LOI $\Gamma_1(\theta)$

Cette loi est couramment admise dans ce genre de calculs ; elle est d'ailleurs conséquente à la loi poissonienne qui régit comme on le sait, le nombre d'événements sur un intervalle fini. Ses applications sont

en fait très variées et elle s'est vue ainsi ajuster avec un plus ou moins grand bonheur à des distributions d'intervalles de temps entre des événements très divers tels que : accidents industriels, maladies, explosions souterraines, etc.

Pour tester son adéquation, on propose de se ramener à un test classique du χ^2 après transformation de la variable T en $X = \log T$. Cette transformation a pour but d'aboutir à une variable dont la variance est indépendante du paramètre, ce qui n'est évidemment pas le cas de la variable T. L'artifice est d'ailleurs relativement classique, les distributions doublement exponentielles auxquelles on aboutit étant des lois du type FISHER, TIPPETT, GUMBEL, WEIBULL.

On peut démontrer les relations :

$$E(X) = \text{Log } \theta - C \text{ ou } C = 0.577216 \text{ est la constante d'EULER}$$

$$V(X) = \pi^2/6 \quad (\sigma(X) = 1.282)$$

On en arrive ainsi à la conclusion importante que la dispersion de X autour de sa moyenne est la même pour tous les individus ; E(X) dépend par contre de θ et caractérise ainsi au même titre que θ la prédisposition individuelle aux accidents.

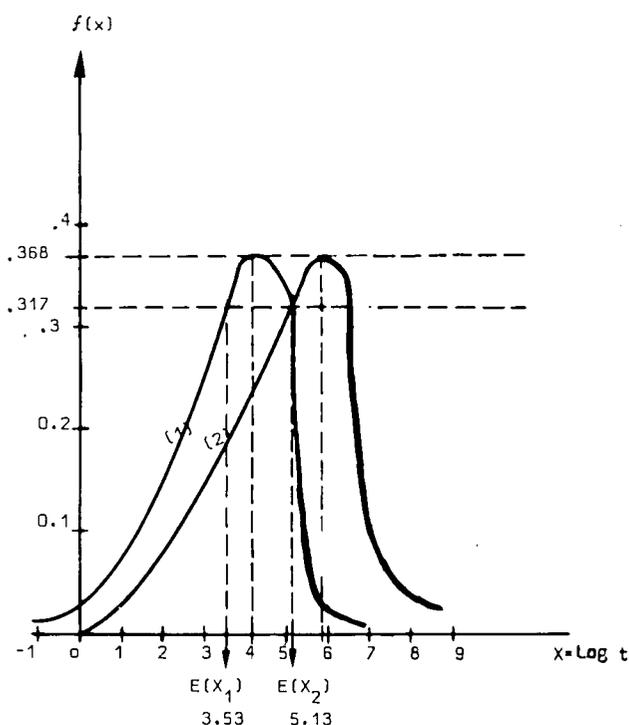
Exploitation graphique :

La densité de probabilité de X soit $f(x) = \frac{1}{\theta} e^{-x/\theta}$

est toujours positive et tend vers zéro pour les valeurs infinies de x ; sa dérivée s'annule pour $\text{Log } \theta$ où la fonction prend la valeur constante 0.368 . Quant à sa valeur à l'origine $\frac{e^{-1/\theta}}{\theta}$, elle apparaît comme une

fonction décroissante de θ dès que la paramètre dépasse la valeur "1 jour", c'est-à-dire pratiquement toujours. Ces résultats permettent d'aboutir à la configuration suivante pour deux conducteurs très différents dont les θ sont respectivement estimés par 60 et 300 jours par exemple. On remarquera que le graphique ci-joint résume le fait que les deux distributions sont également dispersées et que c'est seulement la moyenne $\text{Log } \theta - C$ qui caractérise la prédisposition individuelle aux accidents.

CONDUCTEUR		
	(1)	(2)
$\hat{\theta}$	60 jours	300 jours
$\text{Log } \hat{\theta}$	4.1	5.7
E(X)	3.53	5.13
f(o)	0.0164	0.0033
	f(6) = 0.008 f(3.53) = 0.317	f(7) = 0.095 f(5.13) = 0.318



En dilatant maintenant l'échelle des ordonnées en $r_1 f(x)$ et $r_2 f(x)$ ou r_1 et r_2 sont les nombres d'accidents des conducteurs (1) et (2) relatifs à une certaine période, les maxima des 2 courbes n'auront évidemment plus la même ordonnée (0.368) et il s'agira de savoir si les histogrammes observés peuvent être correctement représentés par les courbes théoriques précédentes. Tel est l'objet du test du χ^2 .

Test du χ^2 :

Les fréquences théoriques se calculent, une fois trouvé un estimateur de θ . La relation $E(X) = \text{Log } \theta - C$

permet d'opter pour l'estimateur $\theta^* = e^{c + \bar{x}_r}$ avec $\bar{x}_r = \frac{1}{r} \sum_{i=1}^r \text{Log } t_i$; on lui préfère néanmoins l'esti-

mateur du maximum de vraisemblance : $\hat{\theta} = \bar{t}_r = \frac{\sum t_i}{r}$.

Pour un mauvais conducteur d'autobus par exemple, exposé au risque pendant 1680 jours, on a relevé 36 accidents, le dernier s'étant passé au 1656ème jour, ce qui donne $\hat{\theta} = \bar{t}_{36} = 46$ jours/accident.

Le tableau ci-dessous résume l'ensemble des observations et calculs effectués ; N_{ij} y représente le nombre d'accidents tels que les logarithmes des intervalles temporels appartiennent à (x_i, x_j) et $F(x)$ la fonction

de répartition de la variable X :

$$F(x) = 1 - e^{-x/\theta}$$

La table du χ^2 à 3 degrés de liberté (5 classes et 1 paramètre estimé) donne pour un seuil de 5 % un χ^2 limite de 7.815 ; ceci permet ainsi d'admettre la validité de la loi exponentielle négative.

Un tel résultat néanmoins n'est valable que pour un individu déterminé caractérisé par son θ . L'étude d'un ensemble de conducteurs doit donc être précédée par

une vérification préalable d'adéquation de la loi $\frac{1}{\theta} e^{-t/\theta}$

pour chaque individu, cette vérification étant faite à partir du dossier correspondant (1) (enregistrement des accidents passés).

(1) X variant théoriquement de $-\infty$ à $+\infty$ il est rare d'avoir, et même pour les "meilleurs conducteurs" des observations x plus grande que 7. Si l'on écrit en effet $\text{Log } t = 7$, on trouve $t = 1000$ jours ce qui constitue en quelque sorte une performance.

La remarque reste valable pour la limite inférieure : il faudrait vraiment de la mauvaise volonté pour aller au-delà d'un certain $\text{Log } t$ considéré comme limite inférieure pratique.

$] x_i, x_j]$	N_{ij} observé	$F(x_j) - F(x_i)$	N_{ij} théorique	$(N_{ij\text{obs}} - N_{ij\text{théo}})^2$	$(N_{ij\text{obs}} - N_{ij\text{théo}})^2$ $N_{ij\text{théo}}$
$] -\infty, 1.8]$	6	0.123	4.43	2.46	0.55
$] 1.8 ; 3.3]$	7	0.323	11,62	21,34	1.83
$] 3.3 ; 3,75]$	6	0.185	6.66	0.43	0.06
$] 3.75 ; 4.20]$	9	0.135	4.86	17.14	3.52
$] 4.20 ; + \infty]$	8	0.234	8.42	0.17	0.02
TOTAL	36	0.991	35,99		5.98

Test à partir des cartes de contrôle :

Il peut être beaucoup plus intéressant que le test χ^2 précédent incapable de discerner des changements dans le temps du paramètre θ .

Certains tests appliqués au dessin des cartes de contrôle, permettent par contre de déceler une amélioration dans les conditions de conduite (augmentation de θ) due à l'expérience, ainsi qu'inversement une détérioration et un accroissement conséquent du risque individuel.

Les limites de la carte de contrôle sont obtenus à partir de la fonction de répartition $F(x) = 1 - e^{-x/\theta}$

Pour des probabilités $F = 0.999, 0.975, 0.500, 0.025, 0.001$, elles s'écrivent :

$$x_F = \text{Log}(r\bar{t}_F) - \text{Dig}(r-1) + \text{Log} \left[\text{Log} \frac{1}{1-F} \right]$$

Dans le cas de l'exemple précédent on trouve :

$$x_F = 3,836 + \text{Log} \left[\text{Log} \frac{1}{1-F} \right]$$

ce qui permet de dresser le tableau suivant :

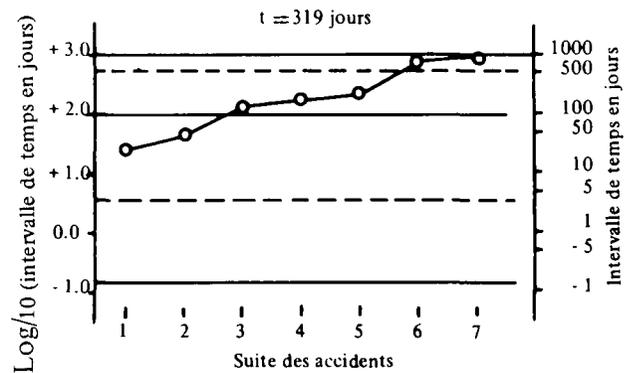
F	0.999	0.975	0.500	0.025	0.001
X_F	5.766	5.138	3.470	0.136	3.174

donnant les limites \hat{x} relatives à un individu observé pendant une assez longue période et dont on possède une estimation de la susceptibilité aux accidents. Que cette propension vienne à changer et les cartes de contrôle dressées, permettront immédiatement d'en déceler la modification.

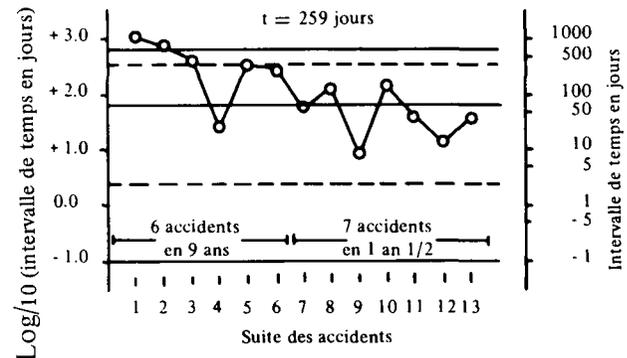
Si l'on se donne en outre une référence θ_0 , il en découle de nouvelles frontières $\hat{x}_F(\theta_0)$ et donc la distinction importante entre "bon" et "mauvais" conducteur selon que l'individu étudié possède une concentration de points représentatifs au-dessus ou au-dessous de la ligne médiane $\hat{x}_{0.5}(\theta_0)$.

SICHEL a fourni des exemples de conducteurs d'auto-bus dont la propension a varié au cours du temps :

Conducteur s'étant progressivement amélioré



Conducteur ayant progressivement empiré



B - ETUDE PROBABILISTE ET STATISTIQUE DU MODELE

Les calculs attachés à l'estimateur $\bar{T} = \frac{\sum_{i=1}^R T_i}{R}$ moyenne

d'ordre aléatoire poissonien de variables exponentielles, sont longs et difficiles ; on se bornera ici à en rappeler les principaux résultats :

$$\text{Prob} \left(\bar{T} = \frac{t}{L} < \bar{t} \leq \frac{L}{l} \right) = \frac{1}{i+1} \sum_{i=1}^R \frac{r^i}{\theta(e^\lambda - 1)^r} \frac{t^{r-1}}{(r-1)!} dt$$

probabilité élémentaire présentant des discontinuités aux points L/i ($i = 2, 3, \dots, \infty$).

$$E(\bar{T}) = \theta \left(1 - \frac{1}{e^{\lambda} - 1}\right)$$

$$V(\bar{T}) = \frac{\theta^2}{2(e^{\lambda} - 1)} \left[\lambda^2(S(\lambda) + 1/2) - e^{\lambda}(1 + \lambda) + 4\lambda + 1 - \frac{2\lambda^2}{e^{\lambda} - 1} \right]$$

$$\lim_{\lambda \rightarrow \infty} \sigma(\bar{T}) = \theta / \sqrt{\lambda}$$

avec : $\lambda = L/\theta$

$$S(\lambda) = \sum_{i=1}^{\infty} \frac{\lambda^i}{i!} = \bar{E}_i(\lambda) - C - \text{Log } \lambda$$

où $C = 0.577216$ est la constante d'Euler et

$\bar{E}_i(\lambda) = \int_{\infty}^{\lambda} \frac{e^{-x}}{x} dx$ est l'intégrale exponentielle classique dont les valeurs sont tabulées.

En outre les bornes θ_{α} et $\theta_{1-\alpha}$ de l'intervalle de confiance attaché à θ sont solutions de l'équation :

$$\frac{1}{(e^{L/\theta^{\alpha}} - 1)} \sum_{i=1}^{\infty} \left[1 - \left(\frac{\bar{t}}{L}\right)^i \right] \frac{(L/\theta^{\alpha})^i}{i!} = \alpha$$

écrite pour α et $(1-\alpha)$ respectivement et résoluble à l'aide de procédés de calculs numériques.

On voit ainsi que \bar{T} est un estimateur biaisé de θ et que l'estimateur sans distorsion correspondant ne peut être exprimé analytiquement du fait de la forme non linéaire de l'équation donnant $E(\bar{T})$.

On notera cependant que le biais relatif B.R. (\bar{T})

$$= \frac{E(\bar{T}) - \theta}{\theta} = \frac{\lambda}{e^{\lambda} - 1}$$

qui est toujours négatif, décroît au fur et à mesure que λ augmente. On peut aussi remarquer que la distorsion peut être négligée dès que l'exposition au risque L dépasse 5 fois la réticence θ .

L'estimateur a été calculé par SICHEL pour les conducteurs d'autobus d'une compagnie sud-africaine de transports publics et un classement satisfaisant en a résulté. Le tableau correspondant sera donné à l'occasion de l'étude d'un second estimateur.

Il faut pour le moment remarquer que l'estimateur T ne tient pas compte de la période de temps écoulée entre le dernier accident et l'extrémité de l'intervalle d'exposition. Cette méthode sous-estime donc la réticence individuelle aux accidents et apparaît ainsi comme "injuste" pour les conducteurs qui n'auraient eu qu'un faible nombre d'accidents au début de la période d'exposition. Dans ce qui suit, on essaye de voir si l'on ne peut trouver un estimateur mieux adapté à ce problème.

Pour cela on maximise la vraisemblance du paramètre θ pour les observations t_1, t_2, \dots, t_r ($r > 1$) et l'on trouve que l'estimateur $\hat{\lambda}$ correspondant est solution de l'équation

$$\frac{2r}{\lambda} - \frac{\sum t_i - e^{\lambda}}{L e^{\lambda} - 1} = 0$$

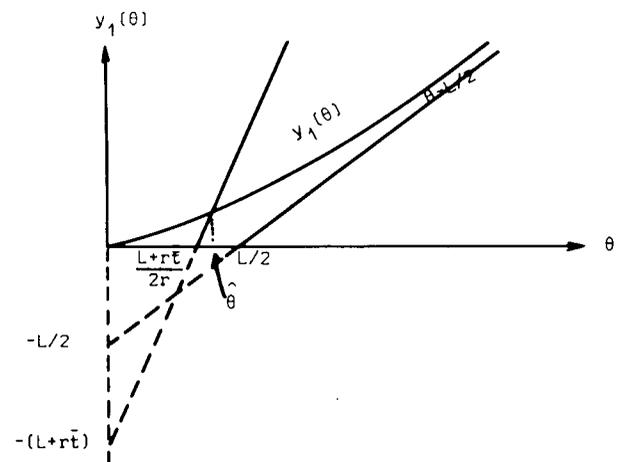
équation dont la solution approchée lorsque λ est de l'ordre de 5, est

$$\hat{\theta}_{app} = \frac{L}{\lambda_{app}} = \frac{L + r\bar{t}}{2r}$$

Quand à la solution exacte $\hat{\theta}$, on peut remarquer qu'elle est donnée graphiquement par l'abscisse du point d'intersection de la droite $2r\theta - (L + r\bar{t})$ avec la courbe

$$y_1(\theta) = \frac{L}{e^{L/\theta} - 1}$$

Il est facile de voir que $y_1(\theta)$ est croissante, qu'elle part de l'origine avec une pente nulle, qu'elle admet pour asymptote la droite $\theta = L/2$ et qu'elle ne possède pas enfin de point d'inflexion ; de telle sorte que θ se déduit aisément de la représentation graphique suivante :



Il est intéressant de noter que les 3 estimateurs \bar{t} , $\hat{\theta}_{app}$ et $\hat{\theta}$ vérifient entre eux les inégalités suivantes :

$\bar{t} \leq \hat{\theta}_{app} \leq \hat{\theta}$, les égalités limites correspondantes ayant lieu :

- la première, quand le r^e accident survient à l'extrémité droite de l'intervalle d'exposition ($L = \bar{r}t$) ;
- la deuxième, lorsque λ est infini, ce qui n'a évidemment pas de sens pratique.

Dans une certaine mesure $\hat{\theta}_{app}$ répond donc à l'objection soulevée au sujet de \bar{t} qui surestimait la propension aux accidents.

Pour les valeurs élevées de λ , l'inégalité de Cramer permet de voir que la borne inférieure de $V(\theta)$ tend vers $\theta^2/2\lambda$; on a donc :

$$\lim_{\lambda \rightarrow \infty} \text{Inf } \sigma(\hat{\theta}) = 0,7 \frac{\theta^{3/2}}{\sqrt{L}} = 0,7 \theta / \sqrt{\lambda}$$

On se souvient que dans le cas de l'estimateur \bar{t} , l'écart type $\sigma(\bar{t})$ tendait carrément vers $\theta/\sqrt{\lambda}$.

Le tableau suivant compare, pour $\theta = 100$ jours et pour diverses valeurs de λ (5, 10, 15), les limites des écarts-types $\sigma(\bar{t})$ et $\text{Inf } \sigma(\hat{\theta})$.

$\theta = 100$ jours		
λ	Limite pour $\lambda \rightarrow \infty$ de	
	$\sigma(\bar{t})$ (jours)	$\text{Inf } \sigma(\hat{\theta})$ (jours)
5	56	31.8
10	39	22.3
15	15	10

Si l'on en revient à l'estimateur $\hat{\theta}_{app}$, on peut écrire son biais relatif sous la forme

$$\text{B.B.}(\hat{\theta}_{app}) = \frac{\lambda}{2(e^{\lambda}-1)} [\text{Ei}(\lambda) - C - \text{Log } \lambda - \frac{(e^{\lambda}-1)}{\lambda}]$$

et le tableau suivant donne pour quelques valeurs de λ , l'expression des biais relatifs B.R. (\bar{t}) et B.R. ($\hat{\theta}_{app}$) exprimés en %.

λ	B.R. (\bar{t})	B.R. ($\hat{\theta}_{app}$)
0.5	- 77.1	- 28.0
1	- 58.2	- 12.0
2	- 31.3	8.0
3	- 15.7	14.8
4	- 7.5	16.0
5	- 3.4	14.0
10	- 0.04	6.5
15	- 0.00	4.0

Il montre en particulier que malgré sa convergence moins rapide vers zéro, l'estimateur $\hat{\theta}_{app}$ est plus adapté que \bar{t} pour les faibles valeurs de λ ; bien entendu on peut démontrer que :

$$\lim_{\lambda \rightarrow \infty} \text{B.R.}(\hat{\theta}_{app}) = 0$$

Quant à la variance de $\hat{\theta}_{app}$, elle s'écrit :

$$V(\hat{\theta}_{app}) = \frac{\theta^2}{4(e^{\lambda}-1)} \left[\lambda^2 \Sigma(\lambda) + \frac{1}{2} [\lambda^2(S(\lambda) + 1/2) - e^{\lambda+1} - \lambda e^{\lambda}] - \frac{\lambda^2 S^2(\lambda)}{e^{\lambda}-1} \right]$$

$$\text{avec } \Sigma(\lambda) = \sum_{r=1}^{\infty} \frac{\lambda^r}{r^2 \cdot r!} = \int_0^{\lambda} S(t) \frac{dt}{t}$$

Pour $\theta = 100$ jours et $\lambda = 5$, on trouve $\sigma(\hat{\theta}_{app}) = 50$ jours, valeur qui est à rapprocher de $\lim \text{Inf } \sigma(\hat{\theta}) = 31.8$ jours et de $\lim \sigma(\bar{t}) = 56$ jours.

L'intervalle de confiance où se situe la vraie valeur du paramètre θ en regard des observations est donné par la relation :

$$\text{Prob} [\hat{\theta} - 1.29 \text{inf } \sigma(\hat{\theta}) < \theta < \hat{\theta} + 1.29 \text{inf } \sigma(\hat{\theta})] = 80\%$$

qui tient compte du fait que $\frac{(\hat{\theta} - \theta)}{\text{Inf } \sigma(\hat{\theta})}$ suit asymptotiquement une loi normale centrée réduite.

On peut alors confronter les résultats relatifs aux intervalles de confiance, suivant que l'on utilise l'estimateur \bar{t} ou $\hat{\theta}$ (1).

Ce tableau dont les chiffres relatifs à l'estimateur t sont dus aux observations faites par SICHEL sur une compagnie de transport par autobus, permet de remarquer en particulier :

– la corrélation positive des rangs entre l'étendue de l'intervalle de confiance correspondant à chacun des 2 estimateurs et l'estimation de la réticence aux accidents ; la raison en est la proportionnalité de

(1) C'est $\hat{\theta}_{app}$ qui a été effectivement calculé ; les valeurs élevées de r justifient évidemment cette approximation.

N° d'ordre du conducteur	Nombre d'accidents	Exposition (jours)	Intervalle moyen (jours)	Estimation M.V. approchées		Limite inférieure (jours) $\theta_{0.1}$		Limite supérieure (jours) $\theta_{0.9}$		Longueur de l'intervalle à 80 %	
				θ_{app}	λ_{app}	Esti \bar{t}	Esti θ_{app}	Esti \bar{t}	Esti θ_{app}	Esti \bar{t}	Esti θ_{app}
1	14	192	34	34.5	14.2	26	26.4	54	42.6	28	16.2
2	15	809	42	48.0	16.8	33	37.4	62	58.6	29	21.2
3	11	723	54	60.0	12.0	41	44.2	89	75.7	48	31.5
4	32	2.041	59	61.3	33.2	50	51.6	74	70.0	24	18.4
5	20	1.943	85	91.0	21.3	68	73.2	119	108.8	51	35.6
6	16	1.843	108	111.6	16.5	84	86.6	164	136.6	80	50.0
7	12	2.260	188	188.2	12.0	141	138.7	328	237.7	187	99.0
8	8	2.004	227	238.7	8.4	164	163.6	505	313.8	341	150.2

lim $\sigma(\bar{t})$ et de lim inf $\sigma(\hat{\theta})$ à $\theta^{3/2} \sqrt{L}$ dont une estimation est donnée par $\bar{t}^{3/2} \sqrt{L}$. Il y a donc accroissement rapide avec \bar{t} et diminution lente avec L . L'étendue de l'intervalle de confiance étant directement liée à l'écart type cela explique :

- la corrélation positive des rangs indiquée ci-dessus ;
- l'exception pour la conducteur 4 dont l'intervalle de temps moyen est proche de celui du conducteur 3 mais dont la durée d'exposition est bien plus grande.

– les meilleurs résultats donnés par $\hat{\theta}_{app}$, l'intervalle de confiance correspondant attaché à θ étant moins étendu que dans le cas de l'estimateur \bar{t} .

On notera d'ailleurs que $\hat{\theta}_{app}$ agit surtout en abaissant la borne supérieure $\theta_{0.9}$ ($\theta_{0.9}(\hat{\theta}_{app}) < \theta_{0.9}(\bar{t})$) et l'on pourra chercher à expliquer les variations de la réduction relative de $\theta_{0.9}$ soit : $\frac{\theta_{0.9}(\bar{t}) - \theta_{0.9}(\hat{\theta}_{app})}{\theta_{0.9}(\bar{t})}$

qui peuvent a priori sembler erratiques ; avec le même classement de conducteurs que précédemment, cette réduction prend en effet les valeurs suivantes :

Conducteurs	Nombre d'accidents	$\frac{\theta_{0.9}(\bar{t}) - \theta_{0.9}(\hat{\theta}_{app})}{\theta_{0.9}(\bar{t})}$ en %
1	14	21.1
2	15	5.4
3	11	14.9
4	32	5.4
5	20	8.5
6	16	16.6
7	12	27.5
8	8	37.8

L'explication en est facile à trouver dès que l'on sait que la réduction relative précédente, qui varie comme $\theta_{0.9}(\bar{t})$ est par conséquent une fonction croissante de r tant que L reste approximativement constant. Le tableau donné ci-dessus est évidemment conforme à cette constatation : pour les conducteurs 4, 5, 6, 7, 8 ainsi que pour les conducteurs 2, 3 dont la valeur approchée de l'exposition au risque est de 2.000 et 800 jours respectivement, la réduction relative de $\theta_{0.9}(\bar{t})$ varie bien en sens contraire du nombre des accidents.

La quantité \bar{t} étant un estimateur non biaisé de $f(\theta)$, elle devrait plutôt être confrontée avec $f(\hat{\theta}_{app})$ et $f(\theta)$.

Pour un conducteur ayant été impliqué dans 3 accidents par exemple sur une période de 3 ans et dont l'intervalle de temps moyen de 9 mois résulte des observations

$t_1 = 6$ mois, $t_2 = 12$ mois, $t_3 = 9$ mois, on trouve :

$$\hat{\theta}_{app} = 10.5 \text{ mois} \quad : \quad \hat{\theta} \text{ O } 10.8 \text{ mois}$$

$$f(\hat{\theta}_{app}) = 9.3 \text{ mois} \quad ; \quad f(\hat{\theta}) = 9.47 \text{ mois}$$

de telle sorte que l'estimation de $f(\theta)$ est, selon le choix : $\hat{\theta}_{app}$ ou $\hat{\theta}$, de 5.2 ou 3.3 % respectivement plus grande que \hat{t} .

Néanmoins, de même que précédemment, l'intervalle de confiance attaché à θ reste démesurément grand en particulier dans les cas où λ est inférieur à 3. En appliquant les résultats de STEVENS, SICHEL a montré que l'on réduisait sensiblement l'étendue de l'intervalle de confiance en choisissant pour estimateur :

$$\hat{t}_r = L / (r + z + 3/8) \text{ où } z \text{ est un aléa uniforme sur } [0,1]$$

Pour les valeurs 1, 2, 3 de r par exemple, on réduit la largeur de l'intervalle de 47,3 %, 30,9 % et 23,8 % respectivement lorsqu'on substitue \hat{t}_r à $\hat{\theta}_2$.

CONCLUSION

LES GENERALISATIONS POSSIBLES

Elles portent sur différents domaines.

1) Amélioration des estimateurs proposés

Aucun des deux estimateurs précédemment étudiés ne peut être appliqué quand le nombre d'accidents au cours de la période d'exposition est nul ; ce cas peut cependant être courant dans la pratique :

A cet effet SICHEL a récemment proposé ("Estimation of individual accident liability". Communication libre à la 37ème session de l'Institut International de Statistique) de substituer à l'estimateur "classique" $\hat{\theta}_1 = L/r$, l'estimateur $\hat{\theta}_2 = L/(r+1)$ habituellement utilisé dans les études de fiabilité. Quand r est nul, on peut ainsi admettre que la réticence du conducteur est L .

On a les relations :

$$E(\hat{\theta}_2) = \theta (1 - e^{-\lambda})$$

$$V(\hat{\theta}_2) = \theta^2 e^{-\lambda} \{ \lambda [Ei(\lambda) - \log \lambda - C] - 2(Ch \lambda - 1) \}$$

et le biais relatif de $\hat{\theta}_2$ devient inférieur à 5 % dès que λ dépasse la valeur 3.

2) Introduction de variables nouvelles

Il a souvent été dit que l'accident devait être évité en soi, que le nombre de sinistres, indépendamment des conséquences directes ou indirectes, matérielles ou corporelles, avait sa valeur propre.

En dehors de quelques cas très nets, la responsabilité du conducteur impliqué est toujours délicate à déterminer ; outre son caractère aléatoire, elle n'est d'ailleurs fixée qu'après un certain délai. Il importerait pourtant qu'elle soit prise en compte. Il en est de même de la gravité des sinistres censée donner une idée des conséquences financières.

Les généralisations possibles porteront donc sur ces deux éléments. Il faut pour cela disposer des lois de probabilité suivies par ces 2 variables. Dans ce domaine, on ne peut se livrer qu'à des conjonctures plus ou moins vérifiées, puisque l'imprécision demeure grande malgré les récentes recherches actuarielles.

On rechercherait ainsi à estimer non plus un paramètre de réticence aux accidents mais le quotient d'un tel paramètre par un degré de responsabilité ou par une variable "coût de sinistre" ; cela en rapportant chaque intervalle de temps observé à la responsabilité du conducteur étudié ou au coût du sinistre occasionné. Il semblerait qu'on puisse enfin ainsi aboutir à un indicateur valable de propension aux accidents.

Les résultats des recherches menées sont encore trop superficiels pour mériter un compte rendu dans le cadre de cette étude ; il est néanmoins à espérer que ces recherches deviennent consistantes dans un proche délai.

1. Rappel sur l'utilisation des fichiers "Assurances"

Le principe d'une prime proportionnelle au risque, se traduit pratiquement par un classement homogène par rapport à certains critères tels que : puissance et usage du véhicule, lieu de garage, etc. En admettant alors que tous les membres d'un même sous-ensemble ont des risques équivalents, on est conduit à astreindre chaque personne au paiement de la prime annuelle suivante :

$$\pi = \frac{\bar{c} \cdot p}{(1-g)}$$

Dans cette expression \bar{c} est le coût moyen d'un sinistre de la classe et p le rapport du nombre annuel total de sinistres de la classe au nombre de voitures-années (1) de cette même classe. Autrement dit, p est l'estimation statistique de la probabilité annuelle d'accidents relative à chaque membre de la classe considérée. La prime pure $c \cdot p$ est transformée en une prime commerciale π grâce au facteur g tenant compte des frais généraux, de la réserve de sécurité et du bénéfice de la compagnie.

L'application pratique de ce principe est relativement facile ; mais ce formalisme est assez loin de la réalité, puisque les critères choisis expliquent seulement 10 % de la variance totale. Les indicateurs d'hétérogénéité interne à la classe sont d'ailleurs élevés : dans le cas où l'on admet que les membres d'une même classe n'ont pas le même risque, le coefficient de variation, rapport de l'écart-type des risques à leur moyenne est très élevé (de l'ordre de 1), ce qui semble bien confirmer cette hypothèse. Il en est de même de la dispersion relative de ces risques par rapport au risque dominant (2) : les valeurs atteintes peuvent être de 3 ou 4.

L'utilisation récente de calculs de primes grâce à des formules tenant compte aussi bien des caractéristiques du véhicule que de celles du conducteur, a permis de porter à 20 ou 30 % la variance expliquée ; la partie de la variance gérée en mutuelle reste cependant très élevée. L'intérêt d'une connaissance plus complète des dommages et des erreurs imputables aux conducteurs est donc évident.

Le fichier national des conducteurs qui va être créé dans ce but devrait conduire à une estimation progressive du risque individuel, basée non pas sur les caractéristiques "instantanées" du couple conducteur-véhicule, mais sur les observations d'infractions et d'accidents, cumulées sur une période pouvant être très longue.

(1) Une voiture année est un véhicule assuré pour un an.

(2) Ce risque dominant ou "mode" qui est le plus probable de la classe, caractérise ainsi la voiture "normale" de la classe.

2. Utilisation du fichier national des conducteurs

On voit qu'il s'agira essentiellement de "suivre" chaque conducteur en notant certaines caractéristiques de ses infractions et accidents et en s'en servant pour imposer des mesures tarifaires et pénales. C'est en effet, l'aspect "prévention" du fichier qui nous intéresse ici et non l'aspect pratique de centralisation et d'homogénéisation des renseignements.

a) Incidence tarifaire

Dans quelle mesure les observations effectuées sur un conducteur impliqué dans divers sinistres, peuvent-elles aider à l'élaboration des primes futures ? C'est une réponse statistique qui est recherchée ici : il faut en effet noter que la pratique actuelle du bonus-malus relève surtout d'impératifs commerciaux où l'on recherche une récompense ou une punition du client selon qu'il semble être un bon ou un mauvais risque.

Or, il faut noter que les avis sont partagés en ce qui concerne la valeur du critère "nombre d'accidents" en tant que base de prévision du risque d'un conducteur. Peu d'études ont par contre été faites sur un indicateur aussi complet que celui qui est proposé ici : intervalle de temps entre deux sinistres consécutifs rapporté à la responsabilité et au coût. Il semble donc intéressant, connaissant la valeur individuelle moyenne de cette variable relative à une époque passée, d'essayer d'en prévoir la valeur future conformément au modèle proposé et de moduler les primes en conséquence. Ce sont vers ces préoccupations que s'orientent actuellement nos recherches.

Pour rattacher le modèle à la réalité, des enquêtes en vue de fournir les mêmes renseignements sont indispensables : elles peuvent être conçues à partir de sondages dans les fichiers de Sociétés d'assurances.

b) Incidence pénale

On voudrait ici proposer un rattachement, fondé des mesures judiciaires (amendes, retraits de permis, etc.) aux infractions observées. Bien entendu, ces infractions seront considérées différemment suivant leur gravité ; de même il semble que l'on doive attacher plus d'importance aux "infractions dynamiques" ("moving violations") qu'à l'irrespect formel des lois (absence de contrôle technique par exemple...). Ces remarques sont suggérées par des études américaines ayant dénoté en particulier une forte corrélation entre le nombre d'infractions et la propension aux accidents. C'est d'ailleurs bien pour cela, que l'on envisage de définir une "sinistralité" de chaque infraction, c'est-à-dire la faculté qu'a l'infraction

tion d'entraîner un accident dont les conséquences sont plus ou moins graves, censée rendre compte de la gravité de l'infraction en tant que cause virtuelle d'accidents, la "sinistralité" peut être définie comme le produit de :

Fréquence en % des accidents auxquels l'infraction donne lieu. \times Taux moyen des invalidités permanentes résultant de chaque accident.

Un ensemble de probabilités conditionnelles est à déterminer. Puisque la sinistralité requiert la détermination des probabilités que chaque type d'infraction entraîne effectivement un accident. Les réponses exigent des calculs complexes ainsi qu'un recueil de données très lourd. Mais elles auront l'avantage d'étayer statistiquement les mesures tarifaires et pénales à mettre en œuvre.

Les problèmes d'application, concernant a) et b) seront les suivants :

- choix de ou des indicateur (s) retenu (s) en tenant compte de critères tels que : simplicité de calcul et d'actualisation, significativité statistique, etc. .
- mise en place du système de recueil et de traitement de l'information ; réduction des délais en vue de garder son sens à l'actualisation des caractéristiques "accidentéisme" des conducteurs ;
- élaboration de critères statistiquement valables faisant correspondre l'échelle de variation de l'indicateur à l'ensemble des mesures tarifaires ou pénales à imposer.

SUMMARY

Very numerous studies have treated the subject of "accident proneness" but they have often been confined to a psychological or psychometrical aspect, which consisted in selecting drivers according to their responses to some tests. The choice of these tests was made by taking into account all human characteristics related to accident occurrence : biographical data, physiological measures, intellectual and motor capacities, features of character, etc.

However, the prevention aspect has generally been neglected : the proportion of bad drivers detected by the tests was too small and the number of accidents did not really decrease.

Mathematical attempts have been different : the question was not to "predict" bad drivers, but only to try to know, considering the involvement of some individuals in accidents, if one could reasonably plead for any proneness factor. Thus one needed first a differentiation criterium of individuals (number of accidents during a certain period, degree of responsibility...), secondly a reference value for comparing prone individuals to normal ones.

At the beginning goals were precise but the statistician-observer was too quickly confined into his art ; the results were only of the following type : "two individuals each of them involved in one accident during one year's study are similar ; perhaps are they slightly less prone than a person to whom two accidents occurred during the same period with a similar exposition to risk".

It was recently shown that statistical methods permitted not only a classification of individuals, but also an individualization of their own risk. Naturally a continuous differentiation criterium had to be used (time interval between two consecutive accidents, for exemple) instead of the previous discrete one (number of accidents during a certain period) ; the scientific process remained nevertheless the same, i.e. essentially experimental : observing and trying to infer were the statistician's goals. No supposition was made as to the predictive value of the results but this of course does not mean that such a position is not to be considered.

Methods recommended in this study have been put into practice recently. However their applications have not been very numerous up to the present day. But for the first time (ref. 29) psychometric and statistic methods are giving similar results. This makes quite clear the necessity of extending research in the two fields.

REFERENCES BIBLIOGRAPHIQUES

OUVRAGES GENERAUX

1. BASS
Cours de Mathématiques
2. GIRAULT
a) Calcul de Probabilités
b) Processus Aléatoires
3. INDJOUJIAN
a) Echantillonnage et nombres au hasard (Cours I.S.U.P.).
b) Théorie de l'Estimation (Cours I.S.U.P.).
4. KENDALL
The advanced Theory of Statistics.
5. KUNTZMANN
Mathématiques de la physique et de la technique
6. LUNDBERG
On random processes and their applications to sickness and accident statistics.
7. PARODI
Mathématiques appliquées à l'art de l'ingénieur
8. WILKS
Mathematical Statistics

ARTICLES PARTICULIERS

1. ADELSTEIN
Accident proneness : a criticism of the concept based upon an analyses of shunter's accidents
Journal of the Royal Statistical Society 115, 3, 354-410, 1952.
2. ARBOUS - KERRICH
a) Accident statistics and the concept of accident proneness
Biométries 1957, 7, 340-432.
b) Le phénomène de la prédisposition aux accidents
Industrial medicine and Surgery : April 53.

3. ARIBAUD
Quelques remarques sur l'assurance automobile
Bulletin de l'Institut des Actuaire n° 241 - Décembre 1962.
4. BLUM - MINTZ
Correlation versus curve fitting in research in accident proneness
Psycho. Bull. 48 (5), 1951, 413-8.
5. BRODY
Personal characteristics of chronic violators
Highway Research Board Bulletin 152.
6. BURKE
a) Notes concerning the, Webb-Jones article
Psycho. Bull. 50 (2), 1953, 137-9.
b) Chi-square test for proneness in accident data
Psycho. Bull. 48 (6), 1951, 496-504.
7. CHAMBERS
Psychological tests for accident proneness and industrial proficiency
Medical Research Council Memorandum n° 31, 1955.
8. DELAPORTE
a) Un problème de tarification de l'assurance accidents d'automobiles examiné par la Statistique Mathématique (Compte rendu du XVI^e Congrès International d'Actuaire).
Bruxelles 1960, Vol. 2, 121-135.
b) Tarification du risque individuel d'accidents d'automobiles par la prime modelée sur le risque.
Colloque de l'A.S.T.I.N. : Trieste 1963.
c) Estimation statistique progressive du risque individuel et tarification de l'assurance automobile.
Bull. of the International Statistical Institute, Ottawa 1963, 40 p. 275-284.
d) Quelques problèmes de Statistique Mathématique posés par l'assurance automobile et le bonus pour non sinistre.
Bull. Trim. Inst. Actuaire Français n° 227 Juin 1959, p. 87-102.

- e) Sur l'efficacité des critères de tarification de l'assurance contre les accidents d'automobiles
The **ASTIN Bulletin**, Vol. II Part. I, Janv. 1962 p. 84-95.
9. FISHER
The negative binomial distribution
Annals of Eugenics London 1941, 11, 182
10. FORBES
Analyses of "near-accident" reports
Highway Research Board, Bulletin 152.
11. GOLDSTEIN
Human variables in Traffic accidents
Traffic Safety Research Review : March 64.
12. GREENWOOD – WOODS
The incidence of industrial accidents upon individuals with special reference to multiple accidents
Industrial Fatigue Research Board – Report n° 4, 1919.
13. GREENWOOD – YULE
a) An inquiry into the nature of frequency distributions representative of multiple happenings
Journal of the Royal Statistical Society 1920, 83, 255.
b) Accident Proneness
Biometrika, Vol. 37, 1950.
14. HAKKINEN
Psychological tests and traffic accident frequency
Institute of Occupational health, Juillet 57, p. 156.
15. HALDANE
The fitting of binomial distributions
Annals of Eugenics, London 1941, 11, 179.
16. HORN
A study of Pilots with repeated aircraft accidents
J. Aviation medicine, 18 (5), 1947, 440-9.
17. JONES
The extent to which proneness operates in air force aircraft accidents
An. Psychologist, 6 (7), 391, 1951.
18. LEYGUE
Le problème des accidents de la route : Etude générale de l'exposition au risque
O.N.S.E.R., Rapport sur l'enquête : "Pompes à essence".
19. MAGUIRE – PEARSON – WYNN
The time intervals between industrial accidents
Biometrika, 39 (1 et 2), 1952, 168-180.
20. MARITZ
On the validity of inferences drawn from the fitting of Poisson and negative binomial distributions to observed accident data
Psycho. Bull. 47 (5), 1950, 434-43.
21. McFARLAND
a) Human variables in motor vehicle accidents
Harvard school of public health 1955.
b) Opinions courantes américaines sur le concept de la prédisposition aux accidents.
22. MINTZ – BLUM
a) A re-examination of the accident proneness concept
Journal of Applied Psychology 33 (3), 195-211, 1949.
b) Time intervals between accidents
Journal of Applied Psychology 38 (6), 401-406, 1954.
23. NEWBOLD
Practical applications to the statistics of repeated accidents particularly of industrial accidents
Journal of the Royal Statistical Society, 1927, 90, 487.
24. PRESTON – HARRIS
Psychology of drivers in traffic accidents
Journal of Applied Psychology, 1964, n° 4.
25. R.A.T.P.
Implication des machinistes dans les accidents d'autobus.
26. SCHUSTER – GUILFORD
The psychometric prediction of problem drivers
Traffic Safety Research Review : 6 (4), 1962.
27. SHAW – LYNETTE – SICHEL
The reduction of accidents in a transport company by the determination of the accident liability of individual drivers
Traffic Safety Research Review : December 1961.

28. SICHEL
- a) Estimation of the parameters of a negative binomial distribution
Psychometrika 16, 1, 1951.
- b) The statistical estimation of individual accident liability
Traffic Safety Research Review : March, 1965.
29. SPANGENBERG
- The use of projective tests in the selection of bus drivers
Traffic Safety Research Review : December 1968.
30. SROUR
- Méthodes mathématiques et statistiques appliquées à l'étude de la prédisposition aux accidents
 Thèse de 3ème Cycle ; Faculté des Sciences de l'Université de Paris.
31. TEEL – DUBOIS
- Psychological research on accidents : some methodological considerations
Journal of Applied Psychology, 38 (6), 1954, 397-400.
32. THORNDIKE
- The human factor in accidents with special reference to aircraft accidents
 Project 21, 30, 001, Report 1, 1951.

33. VAN LENNEP
- Psychological factors in driving
Traffic Quartely 6, 4, 1952.
34. WEBB – JONES
- Some relations between two statistical approaches to accident proneness
Psycho. Bull. 50 (2), 133-6.

TABLES

1. BOUVART – RATINET
- Nouvelles tables de logarithmes
2. I.S.U.P.
- Tables statistiques
3. LABORDE
- Tables numériques de fonctions élémentaires
4. WHITTAKER – WATSON
- A course in Modern Analyses
 Tables des Ei , Γ , Γ'/Γ , etc.

