REVUE DE STATISTIQUE APPLIQUÉE

J. ROUDIER

Une application de la théorie de la décision statistique bayesienne : dimensionnement d'une expérimentation d'efficacité des « freins anti-bloquants »

Revue de statistique appliquée, tome 22, nº 4 (1974), p. 3-28

<http://www.numdam.org/item?id=RSA_1974__22_4_3_0>

UNE APPLICATION DE LA THÉORIE DE LA DÉCISION STATISTIQUE BAYESIENNE : DIMENSIONNEMENT D'UNE EXPÉRIMENTATION D'EFFICACITÉ DES "FREINS ANTI-BLOQUANTS" (1)

J. ROUDIER
Institut de Recherche des Transports

Des expériences préliminaires tendant à suggérer que l'usage de certains dispositifs de freins anti-bloquants peut réduire le nombre des accidents de la circulation routière, une expérience à grande échelle est envisagée par l'Organisme National de la Sécurité Routière (O.N.S.E.R.). Etant donné le coût élevé de cette expérience et l'importance des conséquences des décisions qui peuvent en résulter, il convient de procéder avec soin à sa conception et à son dimensionnement : c'est l'objet de la recherche menée par l'Institut de Recherche des Transports et dont cet article contient les principaux résultats.

La nécessité de réduire le coût et d'atteindre un optimum économique, la possibilité de disposer de résultats expérimentaux préliminaires ont incité à utiliser les méthodes et les résultats de la théorie bayésienne de la décision statistique et à envisager des procédures séquentielles.

INTRODUCTION

On se propose de dimensionner l'expérience à mener suivante : il existe un dispositif, dit "dispositif de freins antibloquants" qui permet d'améliorer la tenue de route d'un véhicule automobile qui freine violemment et, par là, est susceptible de réduire le nombre et la gravité des accidents. L'expérience a pour but de mettre en évidence l'efficacité de ce dispositif compte tenu du coût d'équipement des véhicules concernés par la décision d'une part, de la gravité et du coût des accidents survenant aux véhicules d'autre part.

Deux raisons incitent à ne pas utiliser un test d'hypothèse classique sur le pourcentage de réduction du taux d'accident dû au dispositif.

Tout d'abord il est toujours difficile de relier les risques à choisir pour déterminer un tel test au bilan économique de la décision. Ensuite le coût d'équipement d'un véhicule expérimental ne peut être négligé car il est notablement plus élevé que celui d'un véhicule normal à cause de l'état de développement de la technique (prototype dans un cas, équipement industriel dans l'autre) et à cause d'un effet d'échelle. C'est pourquoi nous avons préféré une méthode où non seulement la règle de décision, mais aussi la taille de l'échantillon sont à déterminer en optimisant la fonction économique bilan de l'expérience

(1) Article remis le 20/9/73 ; révisé le 29/1/74

qui comprend à la fois son coût et les avantages de ses conséquences pour la collectivité. Nous avons supposé dans la suite que toutes les caractéristiques des accidents pour des véhicules normaux non équipés du dispositif étaient connues et ne faisaient donc pas l'objet d'une détermination par l'expérience.

L'expérience consiste à observer un échantillon de N véhicules. Au vu des résultats, trois décisions sont possibles :

– On décide de rendre l'équipement du dispositif obligatoire pour les véhicules,

– On décide de ne pas équiper les véhicules,

– Eventuellement, on décide de faire une nouvelle expérimentation dimensionnée en partie en fonction des résultats de la première : nous envisagerons l'utilisation d'une telle procédure séquentielle dans la deuxième partie.

Dimensionner l'expérience consiste à déterminer la taille N de l'échantillon que l'on observera et la règle de décision que l'on appliquera en fonction des résultats observés sur cet échantillon.

Nous nous placerons dans le cadre d'un modèle de décision statistique bayésienne que nous allons maintenant élaborer.

1 – MODELE DE DECISION BAYESIENNE NON SEQUENTIELLE :

1.1 – Détermination de la fonction économique

La décision à prendre est celle d'équiper du dispositif mentionné un certain nombre de véhicules : la décision sera appliquée durant plusieurs années et le nombre de véhicules équipés chaque année sera supposé constant et égal à une certaine valeur v_o.

Le coût d'une décision comprend un certain nombre de termes :

– Le coût d'expérimentation : nous noterons γ le coût d'un véhicule de l'échantillon observé.

– Le coût d'équipement des véhicules : nous noterons a le coût d'équipement d'un véhicule et T le coût actualisé d'équipement de tous les véhicules concernés par la décision.

– Le coût des accidents survenant durant leur durée de vie aux véhicules concernés par la décision.

Pour un véhicule étiqueté par l'indice n, la probabilité d'avoir, durant un intervalle de temps donné (nous prendrons par la suite une année) un accident de coût compris entre o et y est donnée par une certaine fonction de répartition $F^{(n)}(y)$: cette fonction attribue une masse importante au point o, ce qui correspond pour un véhicule n à l'absence d'accidents durant la période considérée. Pour tous les véhicules d'une population, le coût global des accidents survenant est une variable aléatoire dont la loi se déduit des lois $F^{(n)}$ des différents individus de la population.

Nous noterons S_i^k le coût des accidents survenant durant leur $i^{ème}$ année d'existence aux v_o véhicules susceptibles d'être équipés durant la $k^{ème}$ année d'application de la mesure. Si ρ désigne le taux d'actualisation choisi, le coût total des accidents, actualisé à l'année d'expérimentation de la mesure est

$$S = \sum_{k=1}^{t} \sum_{i=1}^{10} \frac{1}{(1+\rho)^{i+k-1}} S_i^k$$

Nous avons supposé que la mesure était appliquée durant t années et que la durée de vie d'un véhicule était de dix ans. De même le coût actualisé d'équipement des véhicules est

$$T = \sum_{k=1}^{t} \frac{1}{(1+\rho)^k} .a.\, v_o = a \sum_{k=1}^{t} \frac{v_o}{(1+\rho)^k} = a\, v$$

Nous posons $v = v_o \sum_{k=1}^{t} \frac{1}{(1+\rho)^k}$; v est le "nombre" de véhicules concernés par la décision : c'est un nombre fictif où l'effectif équipé chaque année est pondéré par le facteur nécessaire pour actualiser les coûts correspondant à l'année d'expérimentation. Nous supposerons t assez grand pour pouvoir utiliser la relation $v = \frac{v_o}{\rho}$.

Nous observons sur l'échantillon de taille N une variable aléatoire Z. Comme nous nous interdisons pour le moment la possibilité d'une expérimentation séquentielle, nous cherchons une région d'acceptation H telle que :

– si Z appartient à H, nous prendrons la décision d'équiper v_o véhicules chaque année durant t années.

– si Z n'appartient pas à H, nous déciderons de ne pas équiper.

La fonction de côut pour la collectivité est donc la suivante :

– si Z appartient à H,

$$W(S,Z) = S + T + \gamma N = \gamma N + a v + S$$

– si Z n'appartient pas à H,

$$W(S,Z) = \gamma N + S.$$

Nous noterons

$$W(S,Z) = \gamma N + (S + a v)\, 1_{\{Z \in H\}} + S\, 1_{\{Z \notin H\}}$$

où $1_{\{Z \in H\}}$ est la fonction indicatrice de H qui vaut 1 si Z appartient à l'ensemble H et 0 sinon.

1.2 – Espace des paramètres ; espérance de la fonction de coût

Il convient de préciser maintenant les lois ou les lois conditionnelles des différentes variables aléatoires qui interviennent :

– la distribution des accidents et par suite les variables S_i^k et la variable S dépendent d'un certain paramètre qui prend la valeur p_o lorsque les véhicules ne sont pas équipés du dispositif de freins antibloquants et qui prend une valeur inconnue p lorsque les véhicules sont équipés du dispositif ; le paramètre p est inconnu car il englobe la réduction du taux d'accidents due au dispositif. Nous noterons Ω l'espace des paramètres.

– la variable aléatoire Z observée sur l'échantillon dépend bien sûr du paramètre p, puisque c'est sur p qu'il convient d'obtenir des renseignements ; nous supposerons que Z possède une densité notée $f_N(z,p)$ (l'indice N est là pour rappeler la taille N de l'échantillon).

La méthode bayésienne consiste à supposer une loi de probabilité a priori sur l'espace des paramètres ; on pense, par cette distribution a priori, prendre en compte les informations extérieures dont on dispose, par exemple ici l'avis des spécialistes, sur le pourcentage de réduction que l'on peut "raisonnablement" espérer sur le taux d'accidents des véhicules équipés du dispositif. Nous noterons g(p) la densité de cette distribution a priori.

Les lois précédentes deviennent alors des lois conditionnelles à p fixé.

Nous nous plaçons dans une optique d'analyse économique du coût pour la collectivité : c'est l'espérance de ce coût que nous allons minimiser.

Prenons d'abord l'espérance conditionnelle à p et Z fixés. Si Z appartient à H, l'espérance du coût des accidents est $E(S|p)$; si Z n'appartient pas à H, elle est $E(S|p_o)$;

Par suite

$$E[W|p,Z] = [E(S|p) + av]\, 1_{\{Z\in H\}} + E(S|p_o)\, 1_{\{Z\notin H\}} + \gamma N$$

$$= \gamma N + E(S|p_o) + [E(S|p_o) + [E(S|p) - E(S|p_o) + av]\, 1_{\{Z\in H\}}\,]$$

Nous éliminerons dans la suite la constante $E(S|p_o)$.

Intégrons ensuite par rapport à Z, puis par rapport à p ; nous obtenons

$$E(W) = \Phi\,(N,H) = \gamma N + \int_\Omega g(p)\, dp \int_H f_N\,(z,p)\,[E(S|p) - E(S|p_o) + av]\, dz$$

$\int_H f_N\,(z,p)\, dz$ est la probabilité conditionnelle, à p fixé, que Z appartienne à H, c'est-à-dire la probabilité conditionnelle d'adopter le dispositif ; nous la noterons $P_N\,[Z\in H|p]$.

Nous cherchons donc la taille N de l'échantillon et la zone d'acceptation H, pour la variable Z associée à l'échantillon de taille N, qui minimise la fonction économique :

$$\Phi\,[N,H] = \gamma N + \int_\Omega g(p)\, dp\, [E(S|p) - E(S|p_o) + av]\;\; P_N[Z\in H|p].$$

1.3 – Optimisation de la zone d'acceptation

A N fixé, on peut déterminer simplement une région H (N) qui minimise Φ [N,H] par rapport à H. Supposons pour cela, ce qui ne restreint du reste en rien la généralité de l'exposé, que Z soit un vecteur aléatoire réel et que, par suite, H soit une région de $\mathbf{R}^n$. Pour B, borélien de $\mathbf{R}^n$, considérons la mesure

$$\mu^N \ (B) = \int_\Omega g(p) \ dp \ [E(S \mid p) - E(S \mid p_o) + av] \ P_N \ [Z \in B \mid p]$$

On peut exprimer Φ (N,H) au moyen de μ^N :

$$\Phi \ (N,H) = \gamma \ N + \mu^N (H).$$

H(N) est donc la partie de $\mathbf{R}^n$ qui rende μ^N (B) minimum ; c'est-à-dire le plus grand borélien tel que, pour tout borélien B qu'il contient, μ^N (B) soit néga-tif.

A N fixé, H (N) est défini simplement à partir de μ^N : c'est le support de la partie négative de μ^N.

On trouve

$$\mu^N (B) = \int_B dz \int_\Omega g(p) \ dp \ [E(S \mid p) - E(S \mid p_o) + av] \ f_N \ (z,p).$$

μ^N (B) s'exprimant comme l'intégrale sur B d'une certaine fonction, μ^N (B) est minimum lorsqu'on prend comme ensemble d'intégration B, l'ensemble sur lequel la fonction à intégrer est négative. On en déduit que :

$$H(N) = \left\{ z : \int_\Omega g(p) \ dp \ [E(S \mid p) - E(S \mid p_o) + av] \ f_N \ (z,p) \leqslant 0 \right\}$$

C'est cette dernière expression que nous utilisons par la suite. Nous sommes ramenés à minimiser une fonction d'une seule variable entière :

$$\Psi \ (N) = \Phi \ [N,H(N)].$$

2 – CAS OU LE NOMBRE D'ACCIDENTS SUIT UNE LOI BINOMIALE ET OU CES ACCIDENTS SONT D'UN SEUL TYPE

2.1 – Hypothèses

a) Hypothéses sur la variable Z observée :

La variable Z observée est le nombre d'accidents survenant en une année à l'échantillon de taille N : Z suit une loi binomiale de taille N et de paramètre p.

Le paramètre p représente alors le taux d'accidents survenant aux véhicu-les équipés du dispositif : l'espace Ω des paramètres se réduit à l'ensemble des réels compris entre 0 et 1.

La variable Z est entière, comprise entre O et N : la zone d'acceptation H sera donc une partie de l'ensemble des entiers compris entre O et N. La densité de la loi de cette variable observée est

$$f_N \; (z,p) = C_N^z \; p^z \; (1-p)^{N-z} \text{ pour z entier et } 0 \leqslant z \leqslant N$$

b) Hypothèses sur le coût des accidents :

Nous supposerons qu'il y a un seul type d'accidents de coût unitaire c_o. Alors le coût S_i^k des accidents survenant durant leur $i^{ème}$ année d'existence aux v_o véhicules susceptibles d'être équipés durant la $k^{ème}$ année d'application de la mesure est proportionnel au nombre de ces accidents N_i^k :

$$S_i^k = c_o \; N_i^k .$$

Il convient de noter au passage que l'hypothèse d'un seul type d'accident réduit la variabilité du coût des accidents par rapport à la variabilité du nombre des accidents.

Nous supposerons que les variables N_i^k sont indépendantes et suivent une loi binomiale de taille v_o et de paramètre p ou p_o selon que les véhicules sont ou ne sont pas équipés du dispositif (p_o est ainsi le taux d'accidents des véhicules non équipés).

Alors

$$E(S | p) = \sum_{k=1}^{t} \sum_{i=1}^{10} \frac{1}{(1+\rho)^{i+k-1}} \, v_o \; c_o \; p$$

$$E(S | p) = \left[\sum_{k=1}^{t} \frac{v_o}{(1+\rho)^k} \right] \left[\sum_{i=1}^{10} \frac{c_o}{(1+\rho)^{i-1}} \right] . \, p$$

Posons
$$c = \sum_{i=1}^{10} \frac{c_o}{(1+\rho)^{i-1}}$$

$$E(S | p) = v \; c \; p \text{ et } E(S | p_o) = v \; c \; p_o .$$

c) Loi a priori du paramètre p :

La théorie des distributions conjuguées nous suggère de prendre comme loi a priori du paramètre la loi conjuguée de la loi binomiale de paramètre p : c'est une loi bêta du premier type, de densité :

$$\begin{cases} g(p) = \dfrac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \, \Gamma(\beta)} \, p^{\alpha - 1} \; (1-p)^{\beta - 1} \quad 0 \leqslant p \leqslant 1 \\ g(p) = 0 \text{ sinon.} \end{cases}$$

Rappelons qu'une telle loi a pour moyenne

$$\mu = \int_{o}^{1} p\ g\ (p)\ dp = \frac{\alpha}{\alpha + \beta}$$

et pour variance

$$\sigma^2 = \frac{\alpha \beta}{(\alpha + \beta)^2\ (\alpha + \beta + 1)} = \frac{\mu\ (1-\mu)}{\alpha + \beta + 1}$$

Nous reviendrons plus loin sur le problème de l'ajustement de α et β.

2.2 – Calcul de la fonction économique :

Avec les hypothèses faites, la fonction économique à minimiser peut être simplifiée. Rappelons son expression :

$$\Phi\ [N,H] = \gamma\ N + \int_{\Omega} g(p)\ dp\ [E(S|p) - E(S|p_o) + av]\ P_N\ (Z \in H|p)$$

$$= \gamma\ N + \int_{H} dz \int_{\Omega} g\ (p)\ dp\ [E(S|p) - E(S|p_o) + av]\ f_N(z,p).$$

On peut ici calculer l'intégrale par rapport à p qui vaut

$$\int_{o}^{1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\beta)}\ p^{\alpha-1}\ (1-p)^{\beta-1}\ (vcp - vcp_o + av)\ C_N^z\ p^z\ (1-p)^{N-z}\ dz$$

Nous poserons $\varpi = p_o - \frac{a}{c}$, quantité que nous appellerons "seuil économique": nous verrons plus loin les raisons d'une telle appellation.

$$\int_{o}^{1} p^{\alpha+z-1}\ (1-p)^{N-z+\beta-1}\ (p-\varpi)\ dp =$$

$$\frac{\Gamma(\alpha+z+1)\ \Gamma(N-z+\beta)}{\Gamma(N+\alpha+\beta+1)} - \omega\ \frac{\Gamma(\alpha+z)\ \Gamma(N-z+\beta)}{\Gamma(N+\alpha+\beta)}$$

$$= \frac{\Gamma(\alpha+z)\ \Gamma(N-z+\beta)}{\Gamma(N+\alpha+\beta)} \left[\frac{\alpha+z}{N+\alpha+\beta} - \varpi\right]$$

Ainsi

$$\Phi\ [N,H] = \gamma\ N + vc \sum_{k \in H} C_N^k\ \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\ \Gamma(\beta)}\ \frac{\Gamma(\alpha+k)\ \Gamma(N-k+\beta)}{\Gamma(N+\alpha+\beta)} \left[\frac{\alpha+k}{N+\alpha+\beta} - \varpi\right]$$

2.3 – Détermination de l'optimum

On déduit immédiatement de l'expression précédente la zone d'acceptation H(N), à N fixé :

$$H(N) = \{k : \frac{k+\alpha}{(N+\alpha+\beta)} \leqslant \varpi\}$$

$$H(N) = \{k : k \leqslant (N+\alpha+\beta)\,\varpi - \alpha\}$$

H(N) est donc la forme :

$H(N) = \{k \leqslant y(N)\}$ avec $y(N) = E\,[(N+\alpha+\beta)\,\varpi - \alpha]$ où E désigne la fonction qui à tout nombre réel x associe sa partie entière E(x), c'est-à-dire le plus grand entier relatif inférieur ou égal à x. Nous prendrons donc la mesure d'équipement si le nombre d'accidents observés sur l'échantillon est inférieur à une certaine valeur : ce résultat n'a rien de surprenant.

Nous poserons dans la suite :

$$z(N) = (N+\alpha+\beta)\,\varpi - \alpha - y(N) : 0 \leqslant z(N) < 1 \text{ par définition de } y(N)$$

Il s'agit maintenant de rendre minimum

$$\Psi(N) = \Phi\,[N, H(N)] \text{ avec } H(N) = \{k \text{ entier} : k \leqslant y(N)\}$$

Remarquons au passage que H(N) n'est jamais vide quel que soit N, lorsque $(\alpha+\beta)\,\varpi - \alpha$ est positif.

$(\alpha+\beta)\varpi - \alpha$ est positif si $\varpi \geqslant \mu$, où μ est la valeur moyenne a priori du paramètre : on a alors $p_o - \frac{a}{c} \geqslant \mu$ soit $c(p_o - \mu) \geqslant a$, ce qui revient à dire que la réduction moyenne du coût actualisé des accidents survenant à un véhicule durant sa vie (elle vaut $c(p_o - \mu)$) est supérieure ou égale au coût d'équivalence de ce véhicule.

Au contraire si $(\alpha+\beta)\,\varpi - \alpha$ est négatif, H(N) peut être vide lorsque N est petit.

Nous utiliserons la fonction auxiliaire suivante :

$$\theta(y,N) = \gamma N + \upsilon c \sum_{k=0}^{y} C_N^k \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \int_0^1 (p-\varpi)\, p^{\alpha+k-1}\,(1-p)^{N-k+\beta-1} dp$$

de sorte que $\Psi(N) = \theta\,[y(N), N]$

On peut calculer l'accroissement

$\Delta = \theta\,(y, N+1) - \theta\,(y, N)$ en utilisant la fonction du triangle de Pascal

$$C_{N+1}^k = C_N^k + C_N^{k-1}$$

$$(1-p)\,C_{N+1}^{k} - C_N^k = (1-p)\,(C_N^k + C_N^{k-1}) - C_N^k$$

$$= (1-p)\,C_N^{k-1} - p\,C_N^k$$

ce qui donne

$$\Delta = \gamma + \upsilon c\,.\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \sum_{k=0}^{y} \int_0^1 (p-\omega)\,[C_N^{k-1}\,p^{\alpha+k-1}\,(1-p)^{N-k+\beta} - C_N^k\,p^{\alpha+k}$$

$$(1-p)^{N-k+\beta-1}]\,dp$$

Les termes disparaissent deux à deux et il ne reste que

$$\Delta = \gamma - \upsilon c\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,C_N^y \int_0^1 (p-\varpi)\,p^{\alpha+y}\,(1-p)^{N-y+\beta-1}\,dp$$

$$\Delta = \gamma - \upsilon c\,C_N^y\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\frac{\Gamma(\alpha+y+1)\,\Gamma(N-y+\beta)}{\Gamma(\alpha+\beta+N+1)}\left[\frac{\alpha+y+1}{N+\alpha+\beta+1} - \varpi\right]$$

$$\Delta = \theta\,[y, N+1] - \theta\,(y, N)$$

$$= \gamma - \upsilon c\,C_N^y\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\frac{\Gamma(y+\alpha+1)\,\Gamma(N-y+\beta)}{\Gamma(N+\alpha+\beta+2)}\,[y+\alpha+1-\varpi(N+\alpha+\beta+1)]$$

Deux cas sont alors possibles : $y(N+1)$ est égal à $y(N)$ ou à $y(N)+1$ selon que $\varpi + Z(N)$ est inférieur ou supérieur à 1.

1er cas :

$$y(N+1) = y(N) \text{ si } \varpi + Z(N) < 1$$

Alors

$$\Psi(N+1) - \Psi(N) = \gamma - \upsilon c\,C_N^y\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\frac{\Gamma[y(N)+\alpha+1]\,\Gamma[N-y(N)+\beta)}{\Gamma(N+\alpha+\beta+2)}\,[1-\varpi-z]$$

2ème cas :

$$y(N+1) = y(N) + 1 \text{ si } \varpi + Z(N) \geqslant 1$$

$$\Psi(N+1) - \Psi(N) = \theta\,[y(N)+1, N+1] - \theta\,[y(N), N+1] - \theta\,[y(N), N] + \theta\,[y(N), N+1]$$

Il nous faut calculer la première différence en évaluant

$$\Delta' = \theta\,(y+1, N+1) - \theta\,(y, N+1)$$

$$= \upsilon c \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} C_{N+1}^{y+1} \int_0^1 (p-\varpi)\, p^{\alpha+y}\,(1-p)^{N-y+\beta-1}\, dp$$

$$= \upsilon c \frac{\Gamma(\alpha+.\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} C_{N+1}^{y+1} \frac{\Gamma(\alpha+y+1)\,\Gamma(N-y+\beta)}{\Gamma(N+\alpha+\beta+1} \left[\frac{y+\alpha+1}{N+\alpha+\beta+1} - \varpi\right]$$

$$= \upsilon c\, C_{N+1}^{y+1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\;\mathrm{I}(\beta)} \frac{\Gamma(\alpha+y+1)\,\Gamma(N-y+\beta)}{\Gamma(N+\alpha+\beta+2} \left[y+\alpha+1-\varpi(N+\alpha+\beta+1)\right]$$

$$\Psi(N+1) - \Psi(N) = \gamma + \upsilon c \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \frac{\Gamma[y(N)+\alpha+1]\,\Gamma[N-y(N)+\beta]}{\Gamma(N+\alpha+\beta+2}$$

$$[1-\omega-z(N)]\,[C_{N+1}^{y(N)+1} - C_N^{y(N)}]$$

On trouve finalement

$$\Psi(N+1) - \Psi(N) = \gamma + \upsilon c\, C_N^{y(N)+1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma[y(N)+\alpha+1]\,\Gamma[N-y(N)+\beta]}{\Gamma(N+\alpha+\beta+2)} |1-\varpi-z(N)]$$

On peut synthétiser les deux cas dans la formule suivante :

$$\Psi(N+1) - \Psi(N) = \gamma - \upsilon c\, C_N^{y(N+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma[y(N)+\alpha+1]\,\Gamma[N-y(N)+\beta]}{\Gamma(N+\alpha+\beta+2)} |1-\omega-z(N)|$$

Il nous faut chercher le minimum de $\Psi(N)$, ce qui est assez facile avec la formule donnant l'accroissement $\Psi(N+1) - \Psi(N)$ et celle donnant $y(N$

2.4 – Majoration de l'effectif optimal :

La formule de Stirling, qui permet d'approcher avec une excellente précision la fonction Γ, fournit une majoration de N

Posons $y = y(N)$

$$D = C_N^y \frac{\Gamma(N-y+\beta)\,\Gamma(y+\alpha+1)}{\Gamma(N+\alpha+\beta+2)}$$

Rappelons que $\Gamma(x) \simeq (x-1)^{x-1}\, e^{-x+1} \sqrt{2\pi(x-1)} \sim x^{x-1}\, e^{-x} \sqrt{2\pi x}$ par la formule de Stirling. Il vient, tout calculs faits,

$$D \simeq \varpi^{\alpha}\,(1-\varpi)^{\beta-1}\, N^{-2}$$

et

$$\Psi(N+1) - \Psi(N) \geqslant \gamma - \upsilon c \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \frac{\varpi^{\alpha}\,(1-\varpi)^{\beta-1}}{N^2} \sup_N |1-\varpi-z(N)|\, \mathrm{Sup}_N\left(1, \frac{N-1}{y+1}\right)$$

Le dernier terme prend en compte le fait que c'est $C_N^{y(N+1)}$ qui intervient dans l'accroissement $\Psi(N+1) - \Psi(N)$

$$\Psi(N+1) - \Psi(N) \geqslant \gamma - vc \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \frac{\varpi^{\alpha-1}\,(1-\varpi)^{\beta-1}}{N^2} [\mathrm{Sup}\,(\varpi,\ 1-\varpi)]^2$$

Ainsi Ψ (N) croit, pour N assez grand car son accroissement devient positif. Ceci donne la condition de majoration : l'effectif optimal est inférieur à N_{sup} avec

$$N^2_{sup} = \frac{vc}{\gamma} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \varpi^{\alpha-1}\,(1-\varpi)^{\beta-1}\,[\sup\,(\varpi,\ 1-\varpi)]^2$$

2.5 : Probabilité à priori d'adopter le dispositif :

La probabilité à priori d'adopter le dispositif est la probabilité a priori que Z soit inférieur à y(N) : notons Q(N) cette probabilité.

$$Q(N) = P_N\ [z \leqslant y(N)] = \sum_{k=0}^{y(N)} C_N^k \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} p^{k+\alpha-1}\,(1-p)^{N-k+\beta-1}\,dp$$

$$Q(N) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \sum_{k=0}^{y(N)} C_N^k \frac{\Gamma(\alpha+k)\,\Gamma(\beta+N-k)}{\Gamma(\alpha+\beta+N)}$$

Un calcul analogue à celui que nous avons fait pour la fonction $\Psi(N)$ conduit à la formule d'accroissement :

$$Q(N+1) - Q(N) = \epsilon\, C_N^{y(N+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \frac{\Gamma[\alpha+y(N)+1]\,\Gamma[N-y(N)+\beta]}{\Gamma(\alpha+\beta+N+1)}$$

où ϵ vaut + 1 ou − 1 selon que $y(N+1) = y(N) + 1$ ou $y(N+1) = y(N)$: ϵ est le signe de $\varpi + z(N) - 1$.

Les formules de récurrence que nous avons démontrées et la majoration de l'effectif optimal permettent de construire un algorithme simple pour résoudre notre problème de dimensionnement : on calcule, par accroissements successifs, la fonction économique $\Psi(N)$ pour N inférieur à N_{sup}.

On retient la valeur N qui donne le minimum ; la mesure est acceptée si le nombre d'accidents observés est inférieur ou égal au seuil y(N). On a en outre la probabilité a priori d'accepter le dispositif, qui vaut Q(N) et que l'on calcule aussi par accroissements successifs.

3 – MODELE DE DECISION SEQUENTIELLE BAYESIENNE ETALEE SUR DEUX ANS (HYPOTHESES BINOMIALES AVEC UN SEUL TYPE D'ACCIDENTS) :

3.1 – Détermination de la fonction économique :

Nous allons reconstruire le modèle dans le cadre d'une décision séquentielle étalée sur deux ans. La procédure est la suivante. On considère un premier échantillon de taille N sur lequel on observe une variable aléatoire Z. Plusieurs décisions sont possibles :

– si Z appartient à une certaine région A dite " zone d'acceptation" on décide d'équiper les véhicules mis en cause par le problème.

– si Z appartient à une certaine région R, dite "zone de refus", on décide de ne pas équiper les véhicules.

– si Z appartient au complémentaire E de $A \cup R$ on décide de poursuivre l'expérience. Pour cela on prend un nouvel échantillon dont la taille N' dépend de l'observation précédente Z. Sur cet échantillon on observe une variable aléatoire Z'. Au vu de Z et Z' on prend la décision :

– si (Z, Z') appartient à une certaine région A' on décide d'équiper les véhicules.

– si (Z, Z') appartient à la région R' telle que

$$\{(Z, Z') \in R'\} \cup \{(Z, Z') \in A'\} = \{Z \in E\},$$

on décide de ne pas équiper les véhicules.

La condition imposée sur R' résulte du fait que l'on veut une décision séquentielle en deux étapes au plus.

Il s'agit de déterminer N, A, R, E et, pour Z appartenant à E, les fonctions N'(Z) et A'(Z).

La fonction de coût W (S, Z, Z') est la suivante :

– si Z appartient à A,

$$W(S, Z, Z') = \gamma N + S + a v$$

où l'on prend en compte le coût de l'expérience plus celui des accidents.

– si Z appartient à R,

$$W(S, Z, Z') = \gamma N + S.$$

– si Z appartient à E et (Z, Z') à A',

$$W(S, Z, Z') = \gamma N + \frac{\gamma}{1+\rho} N' + S + a\left(v - \frac{v_o}{1+\rho}\right)$$

$$= \gamma N + \frac{\gamma}{1+\rho} N' + S + \frac{av}{1+\rho}$$

(nous avons actualisé le coût d'expérimentation pour la deuxième année et seuls les véhicules à partir de la deuxième année sont équipés du dispositif).

– si enfin Z appartient à E et (Z, Z') à R',

$$W(S, Z, Z') = \gamma N + \frac{\gamma}{1+\rho} N' + S.$$

Ainsi

$$W(S, Z, Z') = \gamma N + (S + a v) 1_{\{Z \in A\}} + S 1_{\{Z \in R\}}$$

$$+ \frac{\gamma}{1+\rho} N' 1_{\{Z \in E\}}$$

$$+ (S + \frac{a v}{1+\rho}) 1_{\{(Z,Z') \in A'\}} + S 1_{\{(Z,Z') \in R'\}}$$

3.2 – **Lois des variables aléatoires intervenant :**

Il nous reste maintenant à préciser les lois ou les lois conditionnelles des différentes variables aléatoires qui interviennent.

Nous supposerons dans ce qui suit que tous les accidents sont du même type de coût unitaire c_o. Nous avons encore la formule

$$S = \sum_{k=1}^{t} \sum_{i=1}^{10} \frac{1}{(1+\rho)^{i+k-1}} S_i^k \quad :$$

S_i^k est le coût d'accidents survenant durant leur $i^{ème}$ année de vie aux véhicules concernés durant la $k^{ème}$ année d'application ; $S_i^k = c_o N_i^k$ où N_i^k est le nombre d'accidents correspondant.

Nous supposerons que toutes les variables aléatoires N^k sont indépendantes, de même loi quels soient les indices i et k : nous prendrons des lois binomiales de taille v_o et respectivement de paramètre p_o (taux des accidents des véhicules non équipés du dispositif) si les véhicules ne sont pas équipés Z appartenant à R ou (Z,Z') appartenant à R'), de paramètre p si les véhicules ont été équipés après une expérimentation sur une année (Z appartenant à A) de paramètre p' si les véhicules ont été équipés au bout de deux ans d'expérimentation (Z,Z') appartenant à A').

Nous prendrons comme variables observées Z et Z' le nombre d'accidents survenant aux N et N' véhicules des deux échantillons : Z suit une loi binomiale de taille N et de paramètre p ; Z' suit une loi binomiale de taille N' et de paramètre p'. Nous supposerons de plus Z et Z' indépendantes entre elles et indépendantes des variables N_i^k. Les régions A, R, E forment une partition de l'ensemble des entiers N ; A' et R' une partition de l'ensemble $E \times N$, sous ensemble de $N \times N$.

Comme nous raisonnons dans un cadre "bayésien" nous devons de plus attribuer des lois de probabilité aux paramètres p et p' pour traduire notre connaissance a priori de l'efficacité du dispositif. Nous prendrons encore pour p une loi bêta de paramètres (α, β). Par contre pour p' nous allons tenir compte de l'information apportée par le premier l'échantillon. Nous prendrons comme loi a priori de p' la loi a postériori du paramètre p conditionnée par la valeur de l'observation Z : un calcul simple montre que cette loi est encore une loi bêta de paramètres $(\alpha+k, \beta+N-k)$ si k est la valeur de la variable Z observée sur le premier échantillon de taille N.

Nous avons finalement une cascade de lois conditionnelles sur tous les éléments aléatoires qui interviennent (Z, Z', p, p', S) :

– loi de p : bêta de paramètres (α, β)

– loi de Z conditionnée par p : binomiale de taille N, de paramètre p

– loi de p' conditionnée par Z et p : bêta de paramètres $(\alpha + Z, \beta + N - Z)$

– loi de Z' conditionnée par Z, p, p' : binomiale de taille N', de paramètre p'

– loi de N_i^k conditionnée par Z, Z', p, p' : binomiale de taille v_0,

paramètre p si Z appartient à A

paramètre p_o si Z appartient à R ou (Z, Z') à R'

paramètre p_o si (Z, Z') appartient à A' et k est égal à 1

paramètre p' si (Z,Z') appartient à A' et k est supérieur à 1

Nous pouvons calculer la fonction économique en intégrant par rapport aux différentes variables :

$$\Theta(N, A, E, N', A') = E\,[W(S, Z, Z')]$$

Il nous faut d'abord calculer l'espérance conditionnelle de S :

* si Z appartient à A

$$E(S) = \sum_{k=1}^{t} \frac{1}{(1+p)^k} \sum_{i=1}^{10} \frac{1}{(1+\rho)^{i-1}} E(N_i^k) = p \sum_{k=1}^{t} \frac{v_o}{(1+\rho)^k} \sum_{i=1}^{10} \frac{c_o}{(1+\rho)^{i-1}}$$

$= vc\,p$ (nous reprenons la notation précédente $c = \sum_{i=1}^{10} \frac{c_o}{(1+\rho)^{i-1}}$ coût moyen actualisé des accidents survenant durant la durée de vie d'un véhicule)

* si Z appartient à R ou (Z, Z') à R'

$$E(S) = \sum_{k=1}^{t} \frac{v_o}{(1+\rho)^k} \sum_{i=1}^{10} \frac{c_o}{(1+\rho)^{i-1}} \cdot p_o$$

$$= v\,c\,p_o$$

* si z Z appartient à E et (Z, Z') à A'

$$E(S) = \sum_{k=1}^{t} \frac{1}{(1+\rho)^k} \sum_{i=1}^{10} \frac{c_o}{(1+\rho)^{i-1}} E(N_i^k)$$

$$= \frac{1}{1+\rho} \sum_{i=1}^{10} \frac{c_o}{(1+\rho)^{i-1}} \, \upsilon_o \, p_o + \sum_{k=2}^{t} \frac{1}{(1+\rho)^k} \sum_{i=1}^{10} \frac{c_o}{(1+\rho)^{i-1}} \, \upsilon_o \, p'$$

$$= \frac{\upsilon_o}{1+\rho} \, c \, p_o + \frac{\upsilon}{1+\rho} \, c \, p' = \frac{\upsilon}{1+\rho} \, c \, p' + \frac{\upsilon \, \rho}{1+\rho} \, c \, p_o$$

par suite

$$E(W \mid Z', p', Z, p) = \upsilon \, c \, p_o \, 1_{\{Z \in R\}} + (\upsilon \, c \, p + a \upsilon) \, 1_{\{Z \in A\}}$$

$$+ \upsilon \, c \, p_o \, 1_{\{(Z,Z') \in R'\}}$$

$$+ \left(\frac{\upsilon \, c \, p'}{1+\rho} + \frac{\upsilon \, c \, p_o}{1+\rho} + \frac{a \upsilon}{1+\rho} \right) 1_{\{(Z,Z') \in A'\}}$$

$$+ \gamma \, N + \frac{\gamma \, N'}{1+\rho} \, 1_{\{Z \in E\}}$$

Introduisons la notation $\varpi = p_o - \frac{a}{c}$ et faisons disparaître le terme constant $\upsilon \, c \, p_o$:

$$E(W \mid Z', p', Z, p) = \upsilon \, c \, (p - \varpi) \, 1_{\{Z \in A\}} + \upsilon c \left(\frac{p'}{1+\rho} + \frac{p_o}{1+\rho} - \frac{a \, p_o}{c(1+\rho)} \right) 1_{\{(Z,Z') \in A'\}}$$

$$+ \gamma \, N + \frac{\gamma N'}{1+\rho} \, 1_{\{Z \in E\}}$$

$$E(W \mid Z', p', Z, p) = \gamma \, N + \upsilon \, c \, (p - \varpi) \, 1_{\{Z \in A\}}$$

$$+ \frac{1}{1+\rho} \, 1_{\{Z \in E\}} [\gamma N' + \upsilon \, c \, (p' - \varpi) \, 1_{\{(Z,Z') \in A'\}}]$$

Avant d'intégrer par rapport à Z', nous allons introduire une notation : A' (Z) désignera l'ensemble des valeurs de Z' pour lesquelles (Z,Z') appartient à A' : A' est la réunion, pour z variant dans E, des ensembles $\{z\} \times A'(z)$.

$$E(W \mid p', Z, p) = \gamma N + \upsilon c \, (p - \varpi) \, 1_{\{Z \in A\}}$$

$$+ \frac{1}{1+\rho} \, 1_{\{Z \in E\}} [[\gamma N' + \upsilon c \, (p' - \varpi) \, P \, [Z' \in A'(Z) \mid p', Z, p]]$$

$P[Z' \in A'(Z) \mid p', Z, p]$ est la probabilité conditionnelle que Z' appartienne à A'(Z) : elle vaut

$$\sum_{k' \in A'(Z)} C_{N'}^{k'} \, (p')^{k'} \, (1 - p')^{N' - k'}$$

Nous allons maintenant intégrer par rapport à p'. Rappelons en la généralisant la notation du paragraphe précédent : si A est la région d'acceptation pour une expérimentation durant une seule année avec un échantillon de taille N et une loi a priori de paramètres α et β, la fonction économique est

$$\Phi(N, A ; \alpha,\beta) = \gamma N + vc \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \sum_{k\in A} C_N^k \frac{\Gamma(\alpha+k)\Gamma(\beta+N-k)}{\Gamma(\alpha+\beta+N)} \left[\frac{\alpha+\beta}{\alpha+\beta+N} - \varpi\right]$$

Si nous intégrons maintenant par rapport à p', le crochet vaut en utilisation la notation précédente Φ [N', A'(Z) ; α + Z, β + N - Z]

Donc

$$E(W \mid Z,p) = \gamma N + vc(p-\varpi) 1_{\{Z\in A\}} + \frac{1}{1+\rho} 1_{\{Z\in E\}} \Phi[N'(Z), A'(Z) ; \alpha+Z, \beta+N-Z]$$

(Il ne faut pas oublier que N' est aussi une fonction de Z).

Il ne reste plus qu'à intégrer par rapport à Z et p :

$$\Theta(N, A, E, N', A') = \Phi(N, A ; \alpha, \beta) +$$

$$\frac{1}{1+\rho} \sum_{k\in E} C_N^k \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+k)\Gamma(\beta+N-k)}{\Gamma(\alpha+\beta+N)} \Phi[N'(k), A'(k) ; \alpha+k, \beta+N-k]$$

3.3 – Détermination de l'optimum :

Il s'agit de trouver N, A, E, la fonction N' (K) et l'ensemble A' (K) pour K variant dans E qui rendent Φ minimum.

Nous allons essayer de nous ramener à une fonction d'une seule variable. Pour N et K fixés, nous savons calculer N'(N,K) et A'(N,K) qui rendent Φ (N',A' ; α + K, β + N -K) minimum : c'est l'objet de la méthode mise au point pour une expérimentation sur une seule année. On trouve en particulier que A'(N,K) est l'ensemble des entiers inférieurs ou égaux à Y' avec :

$$Y' = E[(N'+N+\alpha+\beta)\varpi - \alpha - K]$$

soit

$$A'(K) = \{K' \in \mathbf{N} : (K+K') \leqslant E[(N+N'(N,K)+\alpha+\beta)\varpi - \alpha]\}$$

Nous commenterons plus loin ce premier résultat.

Notons θ (N,K) le minimum de

$$\Phi[N'(N,K), A'(N,K) ; \alpha+K, \beta+N-K]$$

Alors $\hat{\Theta}(N, A, E) = \min_{N', A'} \Theta[N, A, E, N', A']$ vaut :

$$\hat{\Theta}(N, A, E) = \Phi(N, A ; \alpha,\beta) + \frac{1}{1+\rho} \sum_{k\in E} C_N^k \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+k)\Gamma(\beta+N-K)}{\Gamma(\alpha+\beta+N)} \theta(N, K)$$

$$\hat{\Theta}(N, A, E) = \gamma N + \sum_{k \in A} C_N^k \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \frac{\Gamma(\alpha+k)\,\Gamma(\beta+N-k)}{\Gamma(\alpha+\beta+N)} \upsilon c \left[\frac{\alpha+k}{\alpha+\beta+N} - \varpi\right]$$

$$+ \sum_{k \in E} C_N^k \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \frac{\Gamma(\alpha+k)\,\Gamma(\beta+N-k)}{\Gamma(\alpha+\beta+N)} \frac{\theta(N,k)}{1+\rho}$$

Cela va nous permettre de déterminer, à N fixé, A et E qui rendent $\hat{\Theta}$ (N, A,E) minimum. A(N) est l'ensemble des entiers K inférieurs ou égaux à N tels que $\upsilon c \left[\frac{\alpha+K}{\alpha+\beta+N} - \varpi\right]$ soit inférieur ou égal à $\frac{\theta(N,K)}{1+\rho}$ et soit négatif ou nul;de même E(N) est l'ensemble des entiers K tels que $\frac{\theta(N,K)}{1+\rho}$ soit inférieur à υc $\left[\frac{\alpha+K}{\alpha+\beta+N} - \varpi\right]$ et soit négatif. Nous sommes donc ramener à optimiser par rapport à la seule variable N, A(N) et E(N) étant déterminés comme indiqué.

Pour préciser un peu mieux A(N) et E(N), nous allons comparer Θ (N,A,E) à $\hat{\Phi}(N) = \min_y \Phi(N, y ; \alpha ; \beta)$

On sait que :

$$\hat{\Phi}(N) = \gamma N + \upsilon c \sum_{k=0}^{\bar{y}(N)} C_N^k \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \frac{\Gamma(\alpha+k)\,\Gamma(\beta+N-k)}{\Gamma(\alpha+\beta+N)} \left[\frac{\alpha+k}{\alpha+\beta+N} - \varpi\right]$$

avec $\bar{y}(N) = E[(N+\alpha+\beta)\varpi - \alpha]$

$$\hat{\Theta}(N,A,E) - \hat{\Phi}(N) = \upsilon c \sum_{\substack{k \in E \\ k > y(N)}} C_N^k \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+k)\,\Gamma(\beta+N-k)}{\Gamma(\alpha+\beta+N)} \left[\frac{\theta(N,k)}{1+\rho} - \upsilon c \left(\frac{\alpha+k}{\alpha+\beta+N} - \varpi\right)\right]$$

$$+ \upsilon c \sum_{\substack{k > \bar{y}(N) \\ k \in E}} C_N^k \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} \frac{\Gamma(\alpha+k)\,\Gamma(\beta+N-k)}{\Gamma(\alpha+\beta+N)} \frac{\theta(N,k)}{1+\rho}$$

Comme on veut rendre $\hat{\Phi}$ (N,A,E) minimum, on en déduit que E(N) est constitué de deux parties :

$$E(N) = \left\{ k : k \leqslant \bar{y}(N) \text{ et } \frac{\theta(N,k)}{1+\rho} < \upsilon c \left(\frac{\alpha+k}{\alpha+\beta+N} - \varpi\right) \right\}$$

$$\cup \{k > \bar{y}(N) \text{ et } \theta(N, K) < 0\}$$

Cela montre que A(N) est contenu dans l'ensemble $\{k \leqslant \overline{y}(N)\}$, qui serait l'ensemble d'acceptation pour un effectif N, si l'expérience durait une seule année. On peut montreur que E(N) a, dans tous les cas, la forme d'un intervalle entourant $\overline{y}(N)$.

Nous sommes ainsi ramener, au terme de ces calculs, à optimiser une fonction de la seule variable N, sachant que l'on peut déterminer pour chaque valeur de N les ensembles A(N) et E(N) et, pour chaque valeur de k dans E(N) l'ensemble A'(N,k) et la fonction N' (N,k).

Rappelons rapidement la procédure de prise de décision, une fois déterminé N optimum. Cn observe Z sur l'échantillon de taille N :

– si Z appartient à A(N) on décide d'équiper v_0 véhicules par an durant la durée de l'application t de l'expérience.

– si Z appartient à R(N) on décide de ne pas équiper

– si Z appartient à E(N) : alors, si Z vaut k, on observe à nouveau N'(k) véhicules durant une année. La décision dépend du nombre total Z+Z' d'accidents observés avec le même seuil que si on avait fait une expérience sur un échantillon total de (N + N') véhicules : on décide d'équiper v_0 véhicules par an à compter de l'année suivante si Z + Z' est inférieur ou égal à $(\alpha+\beta+N+N')\varpi-\alpha$.

On voit qu'il est possible de construire un algorithme de la façon suivante : pour chaque valeur de N on calcule la fonction $\hat{\Phi}$ (N) qui correspond à une expérimentation durant une année ; on modifie ensuite $\hat{\Phi}$(N) pour les valeurs de k autour de y(N) par la formule écrite au dessus ; il reste alors à chercher le minimum de la fonction de la variable entière N ainsi obtenue.

4 – RESULTATS

4.1 – Valeurs numériques des paramètres :

Nous avons utilisé des données fournies par l'O.N.S.E.R. sur le nombre et le coût des accidents survenus durant l'année 1970 et faisant intervenir une voiture particulière ou une camionette :

– Coût total des accidents : $8\ 10^9$ F.

– Nombre total d'accidents : $2{,}8\ 10^6$ F.

– Coût moyen d'un accident : $2{,}87\ .\ 10^3$ F.

$$C_o = 2870 \text{ F}$$

Ce dernier chiffre fournit une estimation de l'espérance mathématique du coût des accidents annuels pour une voiture particulière ou une cammionnette, en supposant que la probabilité d'accident est la même pour tous les véhicules. Nous l'avons supposée constante pendant les dix années de la vie moyenne d'un véhicule (lorsque le véhicule vieillit, l'augmentation du taux d'accident par kilomètre parcouru est compensée par la diminution du parcours moyen annuel)

Nous avons pris un taux d'actualisation

$$\rho = 10\ \%$$

Cela nous conduit à c = 17.500 F

Nous prendrons un taux d'accident, c'est-à-dire probabilité d'être impliquée dans un accident, des véhicules non équipés de $p_0 = 0,333$ (0,35 selon DEPOID. Applications de la statistique). Ce chiffre est supérieur au quotient du nombre d'accidents par le nombre de véhicules pour tenir compte du fait qu 'un accident implique souvent plusieurs véhicules.

Nous avons fait les calculs pour les valeurs suivantes du prix de l'équipement et de l'expérience :

$$a = 500, 600, 700, 800 \text{ F},$$

$$\gamma = 3000, 4000, 5000 \text{ F, prix hors taxes.}$$

Les premières expériences et les premiers résultats connus sur le dispositif de freins antibloquants permettent de prévoir un pourcentage de réduction du taux d'accidents de 15 % environ en rase campagne, c'est-à-dire essentiellement sur les accidents corporels, et de 5 % environ sur les accidents matériels ; ceci donne un pourcentage moyen de réduction de 9 % environ. Cela nous conduit à essayer les distributions a priori suivantes :

Taux de réduction	6%	7	8	9	10	11	5	12	13
α	3,6	3,1	3,4	3,7	3,3	3,7	3,8	3,6	3,7
β	7,9	6,9	7,7	8,8	8,2	8,8	8,2	8,7	9,1

Un paramètre donc la détermination est particulièrement importante sur le dimensionnement de l'expérience est l'effectif de véhicules sur lesquels porte la décision d'équipement. Nous avons considéré deux valeurs de υ :
* υ = 10.000.000 ce qui correspond à une production annuelle de 1 000 000 de véhicules neufs que l'on équipe et un taux d'actualisation de 10 % pour ramener tous les effets économiques à l'année de réalisation de l'expérience.

* υ = 1.000.000 ce qui correspond à l'équipement du dixième de la production annuelle actualisé commpe précédemment.

4.2 – Résultats pour une expérience durant une seule année :

Dans le premier tableau :

$\upsilon = 1.000.000$

$\alpha = 3,7$

$\beta = 8,5$

ce qui correspond à un pourcentage de réduction du taux d'accidents de 9 %. Nous avons fait varier le prix de l'équipement a et celui de l'expérience γ ; dans chaque case figure l'effectif optimal N et le seuil d'acceptation y du nombre d'accidents.

a en F / γ en F	500	600	700	800
Pourcentage de réduction économique	TE = 8,58	TE = 10,30	TE = 12,01	TE = 13,73
3000	N = 1329 y = 404	N = 1331 y = 397	N = 1330 y = 389	N = 1329 y = 381
4000	N = 1158 y = 352	N = 1157 y = 345	N = 1149 y = 336	N = 1148 y = 329
5000	N = 1030 y = 313	N = 1033 y = 308	N = 1033 y = 302	N = 1026 y = 294

Dans le deuxième tableau v = 1.000.000
α = 700 F
γ = 4000 F,

ce qui donne un pourcentage de réduction économique de 12,01 %. Nous avons fiat varier la loi a priori du paramètre, c'est-à-dire α etβ,de façon à faire varier le pourcentage de réduction du taux d'accident tout en gardant un coefficient de variation à peu près constant. ;

Moyenne de la loi à priori	0,317	0,313	0,310	0,306	0,303	0,300	0,396	0,293	0,280
Pourcentage de réduction	4,9	6,0	6,9	8,0	8,9	9,9	11,1	12,1	13,2
α	3,8	3,6	3,1	3,4	3,7	3,3	3,7	3,6	3,7
β	8,2	7,9	6,9	7,7	8,5	7,7	8,8	8,7	19,1
$\alpha + \beta$	12,0	11,5	10,0	11,1	12,2	11,0	12,5	12,3	12,8
Effectif optimal N	1146	1139	1098	1125	1149	1125	1159	1152	1162
Seuil d'acceptation y	335	333 ,	321	329 ,	336	329	339	337	340

On peut faire sur tous ces résultats les remarques suivantes :

– une augmentation de coût de l'expérience se traduit évidemment par une diminution de l'effectif de l'expérience.

– la variation en fonction du coût de l'équipement, tout au moins dans les limites que nous avons considérées, est peu significative

– l'influence de la distribution a priori attribuée au paramètre ne semble pas très grande sur la taille de l'échantillon : elle n'excède pas 10 %. L'influence sur le seuil d'acceptation est beaucoup plus sensible. De toute façon la détermination du seuil optimal par la méthode bayésienne que nous avons employée n'empêchera pas de se reposer le problème de la fixation du seuil d'acceptation de la décision. Il faut remarquer que la variation de la moyenne $\frac{\alpha}{\alpha+\beta}$ de la loi à priori ne suffit pas à tout expliquer. En particulier des quantités (α',β') proportionnelles à (α, β) ne donnent surement pas en raison de la modification de la variance de la loi à priori, le même effectif optimal ni le même seuil d'acceptation : l'effet de la quantité ($\alpha + \beta$) paraît sensible.

On peut dresser les mêmes tableaux que précédemment pour un effectif à équiper v de 10.000.000 de véhicules

Dans le tableau suivant : $v = 10.000.000$
$\alpha = 3,7$
$\beta = 8,5$

a en F / γ en F	500	600	700	800
Pourcentage de réduction économique	TE = 8,58	TE = 10,30	TE = 12,01	TE = 13,73
3000	N = 3992 y = 1214	N = 3970 y = 1185	N = 3934 y = 1152	N = 3927 N = 1127
4000	N = 3522 y = 1071	N = 3491 y = 1042	N = 3470 y = 1016	N = 3464 y = 994
5000	N = 3147 y = 957	N = 3173 y = 347	N = 3149 y = 922	N = 3130 y = 898

α	β	Taux de réduction	Seuil économique	Prix de l'équipement	Prix de l'expérience	Effectif à équiper v = 1.000.00			Effectif à équiper : v = 10.000.000		
						Effectif Optimal	Seuil Optimal	Probabilité à priori	Effectif Optimal	Seuil Optimal	Probabilité à priori
3,7	8.5	8,93	8,58	500 F	3000 F	1329	404	0,536	3992	1214	0,535
3,7	8,5	8,93	10,30	600	3000	1331	397	0,519	3970	1185	0,518
3,7	8,5	8,93	12,01	700	3000	1330	389	0,502	3934	1152	0,501
3,7	8,5	8,93	13,73	800	3000	1329	381	0,484	3927	1127	0,484
3,7	8,5	8,93	8,58	500	4000	1158	352	0,536	3522	1071	0,535
3,7	8,5	8,93	10,30	600	4000	1157	345	0,519	3491	1042	0,518
3,7	8,5	8,93	12,01	700	4000	1149	336	0,502	3470	1016	0,501
3.7	8,5	8,93	13,73	800	4000	1148	329	0,484	3400	994	0,484
3,7	8,5	8,93	8,58	500	5000	1030	313	0,536	3147	957	0,535
3,7	8,5	8,93	10,30	600	5000	1033	308	0,519	3173	947	0,518
3,7	8,5	8,93	12,01	700	5000	1033	302	0,502	3149	922	0,502
3,7	8,5	8,93	13,73	800	5000	1026	294	0,484	3130	898	0,484
3,8	8,2	4,9	14,01	700	4000	1146	335	0,460	3433	1005	0,460
3,6	7,9	5,0	12,01	700	4000	1139	333	0,474	3443	1008	0,474
3,1	6,9	6,9	12,01	700	4000	1098	321	0,489	3279	960	0,489
3,4	8,0	8,0	12,01	700	4000	1125	329	0,496	3412	999	0,496
3,7	8,9	8,9	12,01	700	4000	1149	336	0,502	3470	1016	0,501
3,3	7,7	9,9	12,01	700	4000	1125	329	0,515	3395	994	0,515
3,7	8,5	11,1	12,01	700	4000	1159	339	0,525	3456	1012	0,524
3,6	8.7	12,1	12,01	700	4000	1152	337	0,536	3466	1015	0,535
3,7	9,1	13,2	12,01	700	4000	1162	340	0,547	3483	1020	0,546

Dans ce dernier tableau v = 10.000.000
a = 700
γ = 4000

Moyenne de la loi à priori	0,317	0,313	0,310	0,306	0,303	0,300	0,296	0,293	0,283
Pourcentage de réduction	4,9	6,0	6,9	8,0	8,9	9,9	11,1	12,1	13,2
α	3,8	3,6	3,1	3,4	3,7	3,3	3,7	3,6	3,7
β	8.2	7,9	6,9	7,7	8.5	7,7	8,8	8,7	9,2
$\alpha + \beta$	12,0	11,5	10,0	11,1	12,2	11,0	12,5	12,3	12,8
Effectif optimal N	3433	3443	3279	3412	3470	3395	3456	3466	3483
Seuil d'acceptation y	1005	1008	960	999	1016	994	1012	1015	1020

On voit sur ces chiffres que la taille optimale de l'échantillon est à peu près proportionnelle à la racine carrée de l'effectif concerné par la décision d'équipement. Les autres conclusions sur la sensibilité à la loi à priori, au prix d'équipement et au prix d'expériences restent les mêmes. Le tableau précédent regroupe l'ensemble des résultats précédents ainsi que dans chaque cas, la probabilité à priori d'accepter le dispositif. Cette probabilité est voisine de 0,5 et varie assez peu suivant le jeu des hypothèses : remarquons simplement qu'elle augmente lorsque le pourcentage de réduction du taux d'accidents augmente.

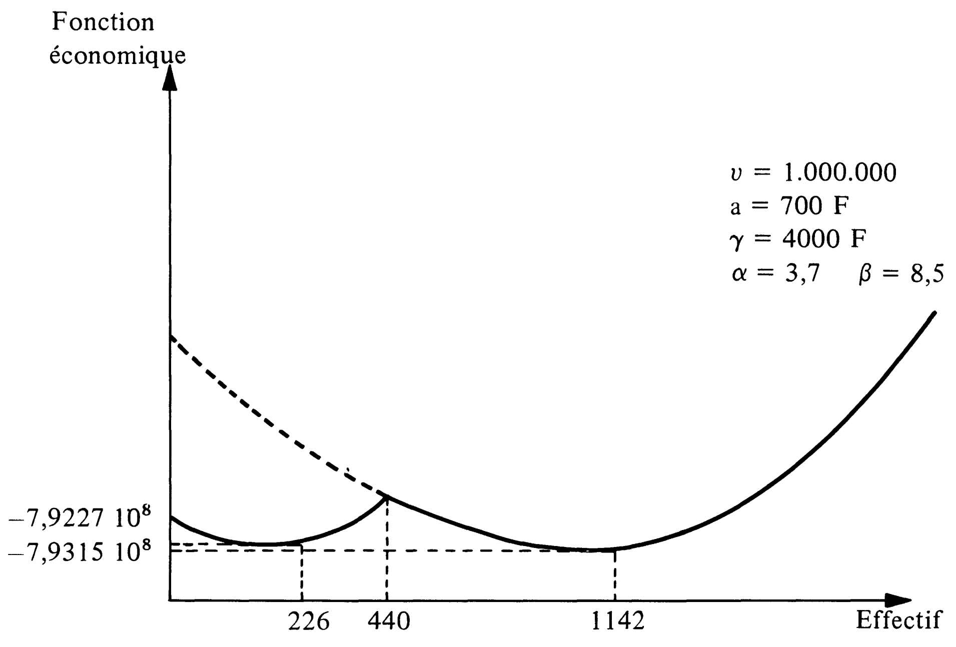
Il convient ici de faire une remarque : il existe une large zone d'effectif de l'échantillon autour de l'optimum dans laquelle la valeur de la fonction économique reste très près du minimum ; une diminution de l'ordre de 10 % de l'effectif se traduit par une perte inférieure à 1 % sur la fonction économique. Il semble donc que l'on puisse dimensionner l'expérience à mener au-dessous de l'optimum sans trop d'inconvénients.

4.3 - Dimensionnement d'une expérience séquentielle sur deux ans :

Les vingt jeux d'hypothèses différentes précédemment retenus ont été utilisés pour chaque valeur de l'effectif concerné par la décision.

Les résultats sont extrêmement décevants. On trouve que, dans tous les cas, l'expérimentation optimale ne doit durer qu'un an et que l'effectif de l'échantillon est celui précédemment obtenu lorsque l'expérience ne durait qu'un an.

Le diagramme suivant schématise la variation de la fonction économique à minimiser en fonction de l'effectif de l'échantillon expérimenté durant la première année :



Pour les valeurs faibles de l'effectif, la portion de courbe en pointillé correspond à la valeur de la fonction économique pour une expérience durant un an seulement ; à partir d'une valeur de l'effectif inférieure à 440 dans notre exemple, une expérience séquentielle n'apporte plus rien par rapport à une expérience d'un an. L'optimum est obtenu avec un effectif de 1142 véhicules pour une expérience durant un an.

Il apparaît un optimum relatif pour la valeur N = 226 de l'effectif expérimenté la première année. Il m'a paru interessant de regarder les caractéristiques de l'expérience à tenter durant la deuxième année dans ce cas là. Pour cet effectif, le seuil d'acceptation de la mesure sans autre expérimentation est de 60 accidents survenus la première année aux 226 véhicules témoins ; le seuil de refus est de 79 accidents.

La probabilité à priori de devoir effectuer une expérimentation durant la deuxième année est de 0,23 ; la moyenne de l'effectif supplémentaire est de 415,5 véhicules ; mais la moyenne conditionnelle de cet effectif si on est dans l'obligation d'expérimenter la deuxième année est de 1810. Le tableau des effectifs supplémentaires de la deuxième année en fonction du nombre d'accidents survenus la premiére année est le suivant :

Nombre d'accidents	Effectif supplémentaire
61	1888
62	2004
63	2096
64	2178
65	2270
66	2294
67	2287
68	2253
69	2195
70	2113
71	2004
72	1878
74	1581
75	1383
76	895
77	575
78	371

La situation est alors la suivante :

– ou on réalise l'expérience optimale avec un effectif de 1142 véhicules en une seule année,

– ou on réalise l'expérience sous optimale étalée sur deux ans que nous venons de décrire. Cette expérience est, en moyenne, nettement moins coûteuse pour des différences faibles sur le bilan économique global de l'opération mais elle présente le risque de devoir équiper un nombre important de véhicules pour l'échantillon expérimenté la deuxième année.

Ce phénomène est lié au fait que la rentabilité marginale d'un véhicule supplémentaire pour l'échantillon équipé la première année est positive jusque pour des valeurs élevées de l'effectif de cet échantillon (1142 dans notre exemple) : mais si elle est très forte pour les premiers véhicules, elle devient ensuite rapidement très faible.

BIBLIOGRAPHIE

MORRIS H. DEGROOT – Optimal Statistical decisions
McGraw – Hill Book Company (1970)

G. HANSEL ET D. GROUCHKO – Prévision séquentielle par la Méthode de Bayes – *Revue de statistique Appliquée* (1965)
Vol. XIII n° 3, p. 67 – 81

R. RAIFFA and SCHLAIFER – Applied statistical decision theory. Harvard University Press Boston (1961)

J. ULMO – La décision statistique dans le cadre bayésien *Revue de Statistique Appliquée* (1971) Vol. XIX n° 3, p. 27 – 66.

A. WALD – Statistical Decision Functions – John Wiley & Sons. Inc., New-York (1950).