SUR LA FONCTION DE DISTRIBUTION DU SINISTRE LE PLUS ELEVE

ED. FRANCKX

Bruxelles, Belgique

A. POSITION DU PROBLEME

La question posée au colloque constitue un cas particulier du problème de la chance et de la malchance. Il appartient à une catégorie plus large de problèmes qui touche aussi bien à la statistique qu'à la recherche opérationnelle.

Pour une catégorie de risque déterminée, il faut définir *deux éléments essentiels.*

a) pour *une période de temps de référence* conventionnelle — la semaine par exemple, il faut disposer de la *loi de survenance des sinistres* c'est-à-dire que pour la période de temps considérée, il faut disposer de la loi qui constate qu'il se présente

o sinistre   I sinistre - - - - - - n sinistres - - - - -

avec les probabilités

q_0 q_1 - - - - - - q_n - - - - - - -

avec

$$\sum_{i=0}^{\infty} q_i = 1$$

nous désignons pour $Q(s) = \sum_{i=0}^{\infty} s^i q_i$

la *fonction génératrice* de cette loi de survenance, toujours définie pour $0 \leqslant s \leqslant 1$.

b) pour la catégorie de risques, nous considérons d'autre part *la fonction de distribution des sinistres survenus.*

$$F(m) = prob \{S \leqslant m\}$$

S désignant la variable aléatoire du montant du sinistre, m la somme payée.

https://doi.org/10.1017/S0515036100001938 Published online by Cambridge University Press

Ces deux éléments essentiels peuvent être donnés soit sur leur forme mathématique (fonctions spécifiées) sont sur leur forme expérimentale (image statistique d'un échantillon).

B. THEOREME GENERAL A.

Nous désignons: par M, la variable aléatoire donnant le sinistre le plus élevé survenant pendant la période conventionnelle pour une catégorie de risque déterminée,

Notons par

$$\Phi(m) = prob\,\{M \leqslant m\}$$

sa fonction de répartition.

On peut énoncer le théorème suivant:

Quelle que soit la loi de survenance des sinistres;
Quelle que soit la fonction de distribution des sinistres survenus;
on a:

$$\boxed{\Phi\,(m) = Q\,[F(m)]} \tag{1}$$

le second membre indiquant que dans la fonction génératrice $Q(s)$, on remplace s par $F(m)$.

Démonstration du théorème.

Désignons par E l'événement $\{M \leqslant m\}$

La loi de survenance définit une suite infinie d'événements distincts qui détermine une partition de la certitude, dès lors on a la formule de décomposition

$$p_E = \sum_{i=0}^{\infty} q_i\, p_{E/i}$$

$_{E/i}$ désignant le fait que le plus grand sinistre est inférieur ou égal à m, si le nombre de sinistre survenus est *exactement* i.

or si $i = 0$ $\quad p_{E/i} = 1,\quad$ car le montant payé est nul, donc inférieur à m.

si $i \neq 0$ $\quad p_{E/i} = (F(m)\,)^i$, en effet il faut que l'on ait pour les i sinistres survenus que simultanément chaque montant soit inférieur ou égal à la somme m. Comme nous admettons *l'hypothèse que, les sinistres survenus sont indépendants a posteriori, l'expression de* p_E/i est justifiée.

Dès lors nous obtenons:

$$\Phi(m) = p_E = \sum_{i=0}^{\infty} q_i \, [F(m)]^i$$

ce qui prouve le théorème.

Remarque.

Ce théorème est très général puisqu'il peut s'appliquer indistinctement à n'importe quelle catégorie de risques homogènes ou hétérogènes; la seule hypothèse faite étant l'indépendance a posteriori des sinistres survenus.

C. APPLICATIONS THEORIQUES

Pour la théorie de la fonction génératrice, on se réfère utilement à Feller: *Probability Theory and its applications* Chapitres 11 et suivants.

1.- *Si la loi de survenance est binomiale* (schéma de Bernouilli).

$$Q_i = C_n^i p^i (1 - p)^{n-i} \qquad i = 0, 1 \ldots\ldots$$

on a:

$$Q(s) = [(1 - p) + ps]^r$$

et par suite

$$\Phi(m) = [1 - p\,[1 - F(m)]\,]^r$$

Si la loi de survenance est la loi de Poisson.

$$q_i = e^{-h} \frac{h^i}{i!} \qquad i = 0, 1 \ldots\ldots$$

$$Q(s) = e^{-h + hS}$$

et par suite

$$\Phi(m) = e^{-h[1 - F(m)]}$$

Ces lois correspondant au cas d'une classe de risque homogène.

2. - *Si la loi de survenance est la loi de Pascal.*

$$q_i = C_{i+k-1}^{i} \, p^i (1 - p)^k$$

on a

$$Q(0) = \left(\frac{p}{1 - (1 - p)s}\right)^k$$

et par suite

$$\Phi(m) = \left[\frac{p}{(1 - F(m)) + p\,F(m)}\right]^k$$

pour $k = 1$, on a en particulier le cas de la distribution géométrique. On sait en particulier que si les risques sont hétérogènes, et la distribution de l'hétérogénéité est définie par la distribution Γ et que la loi de survenance dans chaque classe est de Poisson, on obtient la loi binominale négative, qui est un cas particulier de Pascal.

3. - *Classes hétérogènes. Cas général. Théorème B.*

On pourra toujours supposer au point de vue pratique que les risques peuvent être classés en un nombre au plus dénombrable de m classes homogènes.

$$K_0\, K_1 \ldots\ldots\ldots \qquad K_j \ldots\ldots\ldots$$

et y appartiennent avec la loi de spécification $k_j \geqslant 0 \quad \sum\limits_{j=1} k_j = 1$

D'autre part on peut supposer que dans chaque classe, la loi de survenance est particulière et que dans la classe j, la probabilité d'avoir i sinistres est définie par la loi de survenance

$$l_j^i \geqslant 0 \qquad \sum_i l_i^j = 1$$

Théorème B.

Dans le cas de risques hétérogènes, la fonction de distribution du sinistre le plus élevé est une moyenne pondérée, suivant la loi de spécification de l'hétérogénéité, des fonctions correspondantes de chaque classe homogène.

En vertu des propriétés de la fonction génératrice on a

$$Q(0) = \sum_{j=0}^{\infty} k_j \sum_{i=0}^{\infty} l_j^i\, s^i$$

Si on applique le théorème général, on voit que si $\Psi^j(m)$ désigne

https://doi.org/10.1017/S0515036100001938 Published online by Cambridge University Press

la fonction de répartition du sinistre le plus élevé dans la classe spécifiée j, on a

$$\Psi^{\bullet}(m) = \sum_{j=0}^{\infty} k_j \Psi^{j}(m) \qquad (2)$$

En particulier ce résultat exprime que pour avoir la fonction de distribution du sinistre le plus élevé d'une compagnie, il suffit de pondérer suivant le nombre de risques chacune des fonctions de distribution du sinistre le plus élevé des branches qu'elle exploite.

La formule (2), contient le théorème (1) comme cas particulier. Remarquons qu'en introduisant une intégrale de Stieltjes, le théorème B s'étend au cas général d'une loi de spécification continue ou non.

D. - APPLICATIONS PRATIQUES

Le praticien et en particulier l'actuaire peut concevoir deux types particuliers de problèmes.

Type 1. Calculer numériquement la fonction $\Psi^{\bullet}(m)$.

Type 2. Calculer statistiquement la fonction $\Psi^{\bullet}(m)$.

Problème du type 1.

On peut concevoir deux méthodes pratiques utilisables sur calculatrices électroniques.

Cas A.-

La fonction $\Phi(m)$ possède une expression mathématique connue.

Dans ce cas, le tout revient à calculer une fonction connue, problème classique.

Cas B.

La fonction $\Phi(m)$ n'est pas exprimable sous une forme simple.

Dans ces conditions on peut concevoir, en application du théorème général la méthode suivante (voir fig. 1 en annexe).

a) la loi de survenance q_i permet (fig. 1a) *le calcul numérique* ou *graphique de la fonction génératrice* dans son domaine utile.

$$0 \leqslant s \leqslant 1$$

b) la fonction de distribution des sinistres permet d'établir la table ou le graphique en $F(m)$ fig. 1 b.

Problème Type 1.

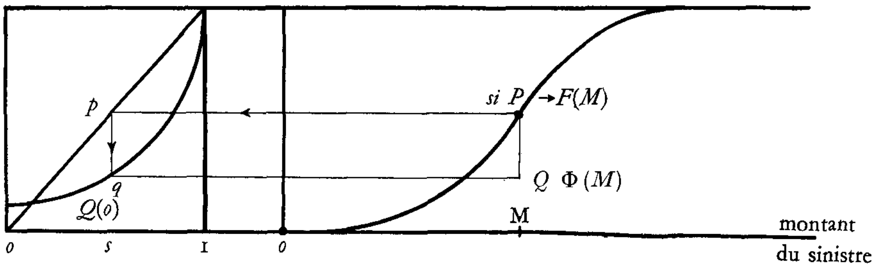


fig. 1a fig. 1b

Dès lors pour trouver un point de Ψ (m), il suffit, si les deux diagrammes ont des ordonnées identiques

1°) partir de P appartenant $F(x)$
2°) chercher le point p appartenant à la bissectrice de fig. 1 a
3°) passer au point correspondant q de la fonction $Q(s)$ se trouvant sur la verticale de p.
4°) reporter le point q par une horizontale au point Q appartenant à la verticale de P.

Théorème C. Le point Q appartient à Ψ (m).
car si $F(m) = u$

Le point r à la verticale de p a comme abscisse u, puisque p appartient à la bissectrice.

Donc $Q(u) = Q(F(m))$; ce qui, en vertu du théorème général justifie que Q appartient à Ψ (m).

La méthode indiquée donne en quelque sorte, la traduction graphique du théorème fondamental. Elle met en lumière, que si la loi de survenance du risque est la même, la même fonction $Q(i)$ reste applicable, quelle que soit la fonction de répartition des sinistres survenus.

Ici joue en particulier le *cas de la réassurance, qui est fondamentalement un opérateur modifiant la fonction de distribution des sinistres, sans modfier la loi de survenance des sinistres.*

Problème du type 2.

La méthode consiste à obtenir, par la méthode de Monte Carlo, un échantillon statistique de la variable M, représentant le plus grand sinistre dans une suite d'épreuves fictives. Pour cela nous

Problème Type 2.

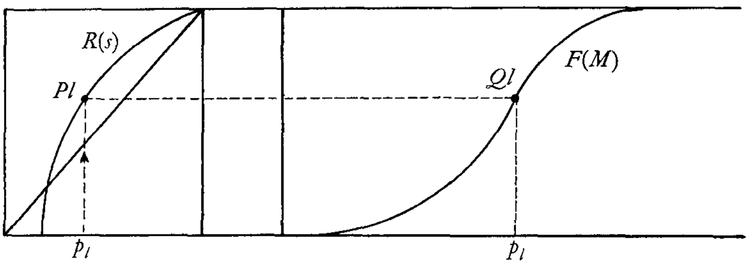


fig. 2ª fig. 2ᵇ

travaillons sur le diagramme dual de la fig. 1, donnée par la fig. 2 (voir annexe).

a) on considère la fonction $R(s)$, symétrique de la fonction génératrice par rapport à la bissectrice (fig. 3 en annexe).

b) on prend un nombre quelconque de points p_l au hasard entre 0 et 1 (valeur de s choisies au hasard — random numbers).

c) si p_l est un de ces points on effectue la construction suivante:

1⁰) prendre le point φ_l, appartenant à $R(s)$ sur la verticale de p_l

2⁰) rejoindre horizontalement le point Q_l appartenant à la fonction de distribution des sinistres.

3⁰) prendre le point P_l sur l'axe des sinistres.

Théorème D. Le point P appartient à l'échantillon de la variable du sinistre maximum.

a) En effet soit m_o une valeur quelconque de l'axe des sinistres. Calculons la probabilité de l'événement E, qui consiste dans le fait que P_l tombe à gauche.

En vertu de la construction effectuée, cette probabilité est égale à la probabilité pour que les points p_l au hasard au départ tombent à gauche de s_0

Comme les points p_l sont pris au hasard, on a donc

$$p_E = s_0$$

avec $F(m_0) = R(s_0)$

donc $F(m_0) = R(p_E)$. (1)

b) Si nous tenons compte de la relation qui existe entre les fonctions $Q(s)$ et $A(s)$

si P est un point quelconque appartenant à la fonction génératrice,

https://doi.org/10.1017/S0515036100001938 Published online by Cambridge University Press

son symétrique par rapport à bissectrice appartient à $(R(s)$.

Il en résulte que:

$$Q[R(s)] = s$$

car on a $Q(R(s)) = s$ si et uniquent si (5) appartient à la droite ① ②

- donc si et uniquement si [(2) (3) (4) (5)] définit un *carré*
- donc il faut et il suffit que $R(s)$ soit le symétrique de $Q(s)$ par rapport à la diagonale principale.

c) Dès lors d'après 1, $Q(s)$ étant une fonction monotone

$$Q[F(m_0)] = Q[R(p_E)]$$

mais d'après 2

$$Q(R(p_E)) = p_E$$

donc

$$p_E = Q(F(m_0))$$

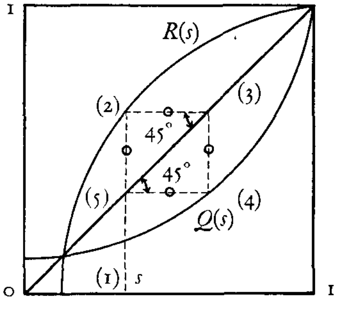
ce qui en vertu du théorème de base démontre le théorème D.

GENERALISATIONS ET CONCLUSIONS

a) Comme il est indiqué dans l'introduction, le problème traité dépasse de loin le cadre des assurances.

Il se présente toute généralité en recherche opérationnelle dans les problèmes de „input — output"
Situons la question:

fig. 3 Relation entre $R(s)$ et $Q(s)$



1) Pendant une période de référence, se présentent en nombre aléatoire des événements E_i.

2) Lorsque l'événement se présente, il déclenche l'arrivée d'une variable aléatoire quantitative X.

Citons des exemples.

Evénements E_i.	*Variable X.*
appels téléphoniques	durée des conversations
panne de machines	durée de réparation
arrivées des malades	durée de traitements prix de l'hospitalisation
arrivées des transports	durée de déchargement prix de la manutention
arrivées de litiges	durée de liquidation frais de litige
déclarations de malades	journée de maladie — frais de maladie
arrivée de sinistres	montant de l'indemnisation
arrivée d'invalidité	valeur actuelle des prestations futures
demande de prêts	montant du prêt
arrivée des malfaçons	indemnisation financière des malfaçons
arrivée des tempêtes	hauteur de la marée
arrivée d'atomes sur un réseau cristallin	quantité de produit catalysé

etc.....

Tous ces problèmes ont la même structure

b) Pour chacun de ces problèmes on peut se poser la question de définir statistiquement la variable donnant le „maximum observable" de la variable quantitative X.

https://doi.org/10.1017/S0515036100001938 Published online by Cambridge University Press

Tous ces problèmes sont isomorphes et la solution donnée pour le „sinistre maximum'' reste valable dans les autres cas. C'est là une première conclusion de cette note.

Sur le plan de la solution trouvée, ce qui est important à notre avis, c'est la possibilité de la disjonction du problème en deux phases.

1) la première phase, qui ne dépend que de la loi de survenance des événements E_i et qui se traduit par le calcul de la fonction génératrice. Cette phase est exhaustive en ce qui concerne les événements E_i, qui déclenchent le phénomène aléatoire.

2) la deuxième phase, elle ne dépend que de la variable X et plus spécifiquement de sa fonction de distribution. Peu importe le problème il suffira d'avoir une „estimation'' de cette fonction de distribution et c'est ici que le théorème de Cantelli Glivenko'' d'une part — le test non paramétrique de Kolmogoroff Smirnoff d'autre part jouent un rôle.

c) D'une manière générale, nous constatons que si la variable quantitative X arrive au hasard entre 0 et 1, la fonction génératrice donne la fonction de répartition du maximum observable pour $0 \leqslant s \leqslant 1$.

C'est là une interprétation nouvelle de la fonction génératrice.