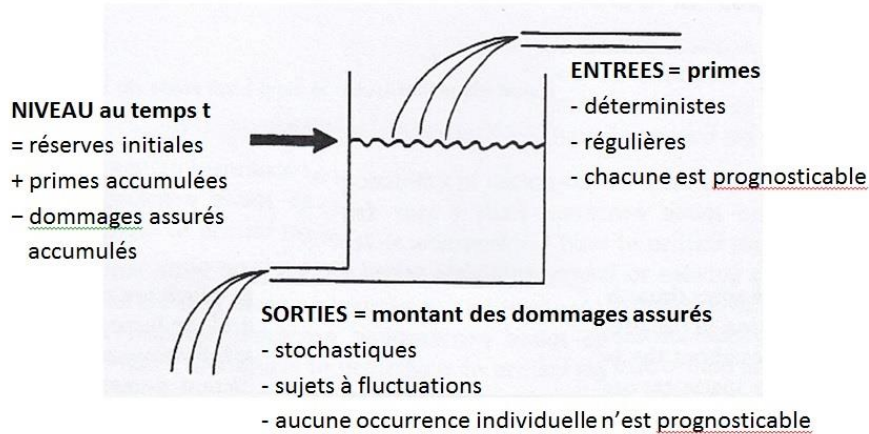
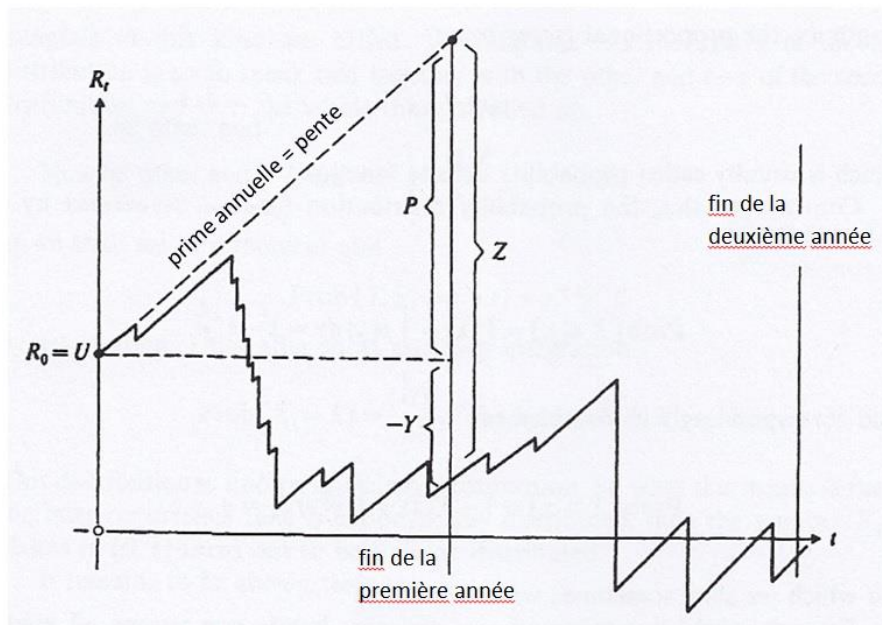


Mathématiques de l'assurance non-vie

Le modèle de base de l'assurance : le réservoir d'eau



L'évolution du niveau d'eau en fonction du temps est un processus stochastique qui se laisse représenter de façon exemplaire par le graphe suivant :



La compagnie d'assurance entame la première année avec des réserves $R_0 = U$ substantielles et ces réserves augmentent en raison de l'afflux de primes. A la fin janvier, il faut payer un premier dommage, ce qui cause une petite chute subite du niveau. Suivent quatre mois sans le moindre sinistre puis une cascade de petits et moyens dommages (causés peut-être par un tremblement de terre et provoquant un grand nombre de dégâts auprès des assurés de la compagnie). Durant le reste de l'année, le volume des primes est plus ou moins égal à celui des dommages à payer. Le résultat de la première année : les réserves ont diminué de 75%. La première moitié de la deuxième année semble prometteuse et pourtant, le résultat à la fin de la deuxième année est désastreux, principalement en raison d'un sinistre exceptionnellement élevé en juillet. A cette occasion, le niveau

tombe en-dessous de zéro, ce qui signifie que la compagnie, n'étant momentanément plus capable de s'acquitter des paiements dus grâce aux seuls moyens financiers de la ligne d'assurance en considération, est forcée de mobiliser d'autres moyens financiers. Dans ce cas de figure, on parle de ruine actuarielle. Il ne s'agit néanmoins que d'une ruine technique et non pas d'une banqueroute de la compagnie d'assurance. Durant le reste de l'année, trois autres sinistres sont à payer si bien qu'à la fin de la deuxième année, toutes les réserves ont fondu, le processus stochastique finissant légèrement dans la négative.

Il est clair que si les sinistres avaient eu lieu en d'autres moments, la courbe en zigzags aurait eu une autre forme. En fait, celle-ci dépend des éléments suivants, dont deux sont de nature déterministe et deux de nature stochastique :

- des réserves initiales $R_0 = U$, qui définissent le point de départ du processus stochastique,
- de la prime annuelle P , qui détermine la pente des parties diagonales de la ligne en zigzags,
- de la suite des durées entre deux occurrences successives T_1, T_2, T_3, \dots , où T_1 dénote la durée entre le début du processus $t = 0$ et l'occurrence du premier sinistre, et ainsi de suite,
- de la suite des montants des dommages individuels X_1, X_2, X_3, \dots , ceux-ci déterminant les chutes verticales du niveau.

Si R dénote le niveau d'argent dans le réservoir immédiatement suite à l'occurrence du $n^{\text{ième}}$ sinistre, alors

$$R = U + PT_1 - X_1 + PT_2 - X_2 + PT_3 - X_3 + \dots + PT_n - X_n.$$

Il est commode d'admettre que les variables aléatoires T_i et X_i sont indépendantes deux à deux (dans toutes les combinaisons possibles impliquant soit un T et un X , soit deux T ou soit deux X) et identiquement distribuées – on note cela i.i.d. – et de loi exponentielle. On pose ainsi que $X := X_i$, $i = 1, 2, 3, \dots$ est de loi exponentielle de paramètre $1/\mu$ et $T := T_i$, $i = 1, 2, 3, \dots$ est de loi exponentielle de paramètre λ . Il s'agit du modèle dit d'Erlang¹.

Il est à noter que lorsque le temps entre deux occurrences suit une distribution exponentielle, le nombre d'occurrence suit quant à lui une loi de Poisson, avec laquelle il est facile de calculer. Le modèle d'Erlang est donc un modèle privilégié dans les assurances².

Il est intellectuellement satisfaisant de savoir que ce modèle du réservoir s'applique à d'autres domaines que celui de l'assurance. La gestion des stocks de la Migros par exemple se laisse modéliser de la même façon : Pour chacun des centaines de produits en vente, le stock de marchandise doit être évalué au plus proche de l'espérance de vente, sans quoi on court le risque de rupture de stock ou au contraire de frais de stockage inutilement élevés. Le stock (grandeur à déterminer) correspond à la prime, l'achat d'un client correspond à un sinistre. Enfin, on peut remarquer qu'en politique énergétique, la gestion du niveau d'eau d'un barrage est un problème dual à celui de l'assurance. Dual dans le sens où ce sont les entrées d'eau dans le réservoir qui sont de nature aléatoire et c'est la

¹ Erlang est un mathématicien danois qui a développé un modèle stochastique prenant en compte à la fois la durée d'un appel téléphonique et le temps entre deux appels successifs. Le modèle d'Erlang sert entre autres à la description de processus comme celui d'une file d'attente à un guichet.

² Et lorsque la loi de Poisson décrit le nombre de sinistres de façon insatisfaisante, on a recours à la loi *binomiale négative*.

sortie d'eau qui est déterministe, les turbines tournant à régime constant (selon les saisons) pour une production électrique régulière.

Quelques connaissances utiles³ en calcul des probabilités

Que la variable aléatoire X soit de loi exponentielle de paramètre $1/\mu$ signifie que sa densité est

$$P(x < X \leq x + dx) = e^{-\frac{x}{\mu}} \frac{dx}{\mu}.$$

Pour s'en convaincre, il faut faire le petit calcul de l'intégrale de 0 à ∞ , devant valoir 1, afin de répondre à la question qu'on ne manque de se poser, à savoir s'il ne manquerait pas un signe négatif à cette densité (mais bon, une densité devrait être positive...) et on constate après calcul que non. Cette réflexion est bien sûr à refaire après chaque longue période passée sans avoir touché de loi exponentielle.

Quant à la fonction génératrice des moments de X , elle est donnée par

$$\phi(\tau) := E[e^{\tau X}] = \int_0^{\infty} e^{\tau x} dV(x) = \int_0^{\infty} e^{\tau x} e^{-\frac{x}{\mu}} \frac{dx}{\mu} = \frac{1}{1 - \mu\tau} \quad \text{pour } \tau < \frac{1}{\mu}.$$

Ici, même chose : si on n'a plus touché de fonctions génératrices depuis longtemps, il faut se convaincre à nouveau qu'une fonction qu'il suffit de dériver une fois en $\tau = 0$ pour obtenir l'espérance, deux fois pour obtenir la variance, trois fois pour... quoi déjà ? La dissymétrie, c'est bien cela ? Et ça continue comment déjà ? il faut se convaincre qu'une telle fonction, c'est bien pratique. Et probablement que, la genèse de cette notion de fonction génératrice restant mystérieuse, on peine à l'intégrer dans sa boîte à outils mathématiques.

En sus de ce qu'on a exigé de X , on veut que le nombre de sinistres N soit une variable aléatoire de loi de Poisson de paramètre λ . La variable aléatoire qui est la plus intéressante pour la compagnie d'assurance est en effet le montant total à payer Z et celui-ci est

$$Z = \sum_{i=1}^N X_i,$$

c'est-à-dire une somme de variables aléatoires dont la borne supérieure de sommation est elle-même une variable aléatoire. Grâce à des considérations sur les fonctions génératrices des moments, on exprime la fonction génératrice de Z – a priori compliquée – comme une composition des fonctions génératrices de X et de N .

³ Le choix que je propose n'est bien sûr pas exhaustif. Il est cependant en rapport avec les thématiques présentées dans ce petit document et il correspond surtout aux quelques questions que je me pose, comme un rituel de mise en route, à chaque fois qu'il m'arrive de me replonger dans la théorie des probabilités après l'avoir laissée longtemps de côté. Notion de rituel de réactivation de connaissances.

Répétons-le : la première dérivée d'une fonction génératrice des moments évaluée en 0 est égale à l'espérance (on dit aussi le premier moment) de la variable aléatoire et la seconde dérivée évaluée en 0 est égale à sa variance (on dit aussi son deuxième moment centré). On reconnaît là le grand intérêt à la fois pratique et théorique des fonctions génératrices : celui de générer les moments. Il est à noter que certains calculs importants trouvent une expression plus simple si l'on considère le logarithme de la fonction génératrice, appelé la fonction log-généralrice. Si l'on note $\varphi(\tau) := \log(\phi(\tau))$, alors on a le résultat central du modèle d'Erlang :

$$\varphi_Z(\tau) = \varphi_N(\varphi_X(\tau)).$$

Ce résultat reste vrai quelle que soit la loi de X et tant que N est de loi de Poisson; on dit que Z suit une loi de *Poisson composée*.

Pour comprendre cela, il faut avoir auparavant compris que pour deux variables i.i.d. X_1 et X_2 ,

$$\phi_{X_1+X_2}(\tau) = (\phi_{X_1}(\tau))^2.$$

De fil en aiguille, on en vient invariablement à se demander une $n^{\text{ième}}$ fois comment on fait pour montrer la loi de la somme $X + Y$, et en quoi la loi de la différence de deux variables $X - Y$ s'en différencie.

Il est à noter que jamais on n'a l'occasion de faire cette réflexion en partant de zéro. Celle-ci est en effet provoquée par la rencontre, au hasard de la $n^{\text{ième}}$ lecture du même texte mathématique de référence, d'une certaine formule dont la substance est la suivante (les minuscules sont des constantes) et qui, puisqu'elle n'est jamais déduite mais toujours donnée comme préalable à la réflexion, conserve une légère odeur de mystère⁴ :

$$\begin{aligned} X + Y \leq c & \quad \Leftrightarrow \quad (X \leq c - y) \wedge (Y \leq y) \\ \{X + Y \leq c\} & = \quad \{X \leq c - y\} \cap \{Y \leq y\} \\ P(X + Y \leq c) & = \quad P(X \leq c - y) \cdot P(Y \leq y) \end{aligned}$$

La première ligne est l'énoncé en formalisme logique, la deuxième est sa traduction ensembliste et la troisième en est la traduction probabiliste, en supposant l'indépendance stochastique entre X et Y . Ce partitionnement est vrai pour tous nombres c et y . C'est alors qu'invariablement, on se demande comment on fait pour passer de cette formule-ci à la densité de convolution de X et Y .

On veut en fait pouvoir employer le calcul intégral et pour cela, il faut remplacer les inégalités par autant de sommes d'égalités (l'intégration discrète, c'est cela, non ?). Dans le cas continu, on veut sommer des tranches⁵ de la forme $P(y < Y \leq y + dy)$ (ici, y est fixé), dont on sait que chacune de ces tranches, notées par exemple $dV(y)$, est égale à $v(y)dy$, où $v(y) := V'(y)$ est la densité de probabilité de Y au point y .

⁴ Parfois il faut avoir obtenu soi-même un résultat pour pouvoir l'accepter totalement. Sinon, la question demeure « Mais comment en arrive-t-on à poser ceci et cela ? Y serais-je arrivé par moi-même ? ».

⁵ Le recours à des éléments infinitésimaux est en réalité une approche mathématique peu formelle et en toute rigueur on devrait avoir recours à la théorie de la mesure. Le V dans l'expression $dV(y)$ est alors la mesure-image par Y de l'ensemble des valeurs que peut prendre Y (ici, \mathbb{R}). Soit en effet une variable aléatoire $Z : \Omega \rightarrow E$; alors $V_Z(z) := P(Z^{-1}(z))$ est une mesure qui associe à chaque partie de E la mesure de sa préimage par Z .

Remplacer les deux facteurs du membre de droite de l'équation conduit à considérer le produit de deux intégrales. Or il n'y a pas de théorème permettant de transformer un produit d'intégrales en une seule intégrale (on aurait alors la densité de la somme $X + Y$). Fubini ne parle pas de produit d'intégrales, ni Cavalieri d'ailleurs, et on cherchera en vain.

Remplacer le membre de gauche de l'équation semble plus prometteur. On se met à considérer que

$$P(X + Y \leq c) = \int_{-\infty}^c P(z < X + Y \leq z + dz),$$

en réfléchissant bien évidemment sur le cas discret

$$P(X + Y \leq c) = \sum_{z \leq c} P(X + Y = z).$$

A ce stade, on a probablement le sentiment d'avoir compris car on fait désormais le lien avec la convolution à partir du cas discret. De

$$X + Y = z \quad \leftrightarrow \quad (X = z - y) \wedge (Y = y)$$

on tire que

$$P(X + Y = z) = \sum_{y \leq z} P((X = z - y) \wedge (Y = y)) = \sum_{y \leq z} P(X = z - y) \cdot P(Y = y)$$

et donc que

$$P(X + Y \leq c) = \sum_{z \leq c} \sum_{y \leq z} P(X = z - y) \cdot P(Y = y).$$

Il reste à effectuer le passage du cas discret au cas continu. Il convient à ce point d'être au clair avec ce qu'on entend par un tel passage, sans quoi on risque effectivement d'écrire n'importe quoi, et ceci d'autant plus que l'écriture adoptée par le livre de Straub (livre de référence, voir bibliographie) et qui est reprise dans ce texte est malheureuse du point de vue de sa maniabilité mathématique.

De façon générale, on fait **passer du discret au continu** une expression de la forme $\sum_{z \in M} P(K = z)$, pour M une partie⁶ de \mathbb{N} , en premier lieu en linéarisant la fonction de répartition cumulative⁷ de la variable aléatoire discrète K . C'est en effet la notion de fonction de répartition (cumulative) qui permet de faire le lien entre les variables aléatoires discrètes et continues.

Par linéarisation, on entend ici le remplacement d'une fonction en escaliers, discontinue à chaque saut, par une fonction affine par morceaux continue partout, de la même façon dont on procède en statistique descriptive pour obtenir graphiquement les quartiles associés à une fonction de répartition cumulative empirique.

⁶ Exiger ceci de M n'est pas une restriction à la généralité mais une simplification bienvenue. On peut en effet, si M est un ensemble discret quelconque, indiquer ses éléments et dire que ce sont les indices de ces éléments qui sont pris dans une partie de \mathbb{N} et non les éléments de M eux-mêmes.

⁷ On précise ici qu'elle est cumulative parce que dans le cas discret, il est possible de donner une fonction de répartition non cumulative qui, dans le cas continu, correspond à la densité.

En deuxième lieu, on écrit $\sum_{z \in M} P(K = z)$ comme $\sum_{z \in M} V_K(z) - V_K(z - 1)$, où V_K est la fonction de répartition cumulative linéarisée de K .

En troisième lieu, on suppose qu'un partitionnement plus fin de l'ensemble $M \subset \mathbb{N}$ fait sens (on s'imagine pour cela qu'on le complète par des éléments de \mathbb{Q}) et on écrit le partitionnement⁸ suivant (l'expression ci-dessous est en fait une nouvelle définition de M pour chaque choix d'un nombre réel positif Δ_z),

$$\mathbb{Q} \supset M_{\Delta_z} := \{n_{\min}, \dots, n_{\max}\} = \mathbb{Q} \cap \bigcup_{i=0}^N]n_{\max} - (i+1)\Delta_z, n_{\max} - i\Delta_z],$$

dont on attend qu'il fasse sens pour tout Δ_z suffisamment petit, de façon à ce que l'on puisse en prendre la limite lorsque $\Delta_z \rightarrow 0$. C'est-à-dire qu'on veut pouvoir considérer M comme un intervalle d'un seul tenant $[n_{\min}, n_{\max}]$.

En dernier lieu, on écrit

$$\sum_{z \in M} P(K = z)$$

comme

$$\lim_{\Delta_z \rightarrow 0} \sum_{z \in M_{\Delta_z}} \frac{V_K(n_{\max} - i\Delta_z) - V_K(n_{\max} - (i+1)\Delta_z)}{\Delta_z} \cdot \Delta_z = \int_M V_K'(z) dz = \int_M v_K(z) dz.$$

L'entier n_{\max} étant la borne supérieure de M , on a que

$$\sum_{z \in M} P(K = z) = P(n_{\min} < K \leq n_{\max})$$

et donc que

$$\int_M v_K(z) dz = \int_{n_{\min}}^{n_{\max}} v_K(z) dz,$$

si bien que, en vertu du théorème fondamental du calcul intégral, la densité s'obtient par

$$v_K(n_{\max}) = \frac{\partial}{\partial n_{\max}} P(n_{\min} < K \leq n_{\max}).$$

Cette digression étant terminée, il s'agit à présent de trouver le pendant continu de l'équation suivante :

$$P(X + Y \leq c) = \sum_{z \leq c} \sum_{y \leq z} P(X = z - y) \cdot P(Y = y).$$

Le membre de gauche de cette équation devient

⁸ Par le partitionnement d'un ensemble, on entend son expression comme réunion disjointe.

$$\sum_{z \leq c} P(z < X + Y \leq z + dz) = \sum_{z \leq c} dV_{X+Y}(z) = \sum_{z \leq c} v_{X+Y}(z) dz.$$

Quant au membre de droite, il devient

$$\sum_{z \leq c} \sum_{y \leq z} \underbrace{(V_X(z-y) - V_X(z-y-\Delta_z))}_{=P(X=z-y)} \cdot \underbrace{(V_Y(y) - V_Y(y-\Delta_y))}_{=P(Y=y)}.$$

On remarque que la borne supérieure de l'une des sommes est l'indice de sommation de l'autre, ce qui permet de ne pas se tromper dans l'ordre de prise de limites. L'expression ci-dessus est ainsi égale à

$$\begin{aligned} \lim_{\Delta_z \rightarrow 0} \sum_{z \leq c} \lim_{\Delta_y \rightarrow 0} \sum_{y \leq z} \frac{V_Y(y) - V_Y(y-\Delta_y)}{\Delta_y} \Delta_y \cdot \frac{V_X(z-y) - V_X(z-y-\Delta_z)}{\Delta_z} \Delta_z \\ = \int_{-\infty}^c \left(\int_{-\infty}^z V_Y'(y) dy \right) V_X'(z-y) dz \\ = \int_{-\infty}^c \left(\int_{-\infty}^z V_Y'(y) V_X'(z-y) dy \right) dz \\ = \int_{-\infty}^c \left(\int_{-\infty}^z v_X(z-y) v_Y(y) dy \right) dz. \end{aligned}$$

La densité de la somme de deux variables aléatoires indépendantes est donc la dérivée $\frac{\partial}{\partial c}$ de cette dernière expression, à savoir (application du théorème fondamental du calcul intégral)

$$\begin{aligned} f_{X+Y}(z) &:= \int_{-\infty}^z v_X(z-y) v_Y(y) dy \\ &= \int_{-\infty}^{\infty} v_X(z-y) v_Y(y) dy, \end{aligned}$$

La dernière égalité venant du fait que v_X est nulle pour un argument négatif. On reconnaît bien là la convolution des densités de X et de Y .

On voit que le passage du discret au continu n'est pas facile et qu'un vilain sentiment de confusion peut persister si l'on se contente de comprendre les choses à peu près. Le passage des sommes aux intégrales nécessite de savoir jongler entre les différentes écritures intégrales, puisqu'une fois le « dx » est l'expression d'une mesure (intégrale de Lebesgue) et une autre fois c'est un élément infinitésimal. Par exemple, on peut se demander à quoi correspond le x entre parenthèses dans l'expression « $dV_X(x)$ », dès lors que celle-ci exprime un élément infinitésimal et non une fonction de x .

Question : Quelle est la loi de la différence de deux variables aléatoires indépendantes X et Y ?

Comment calculer une prime à partir de données concrètes ?

Soit Z la variable aléatoire d'un risque financier pour lequel nous cherchons à déterminer une prime d'assurance correcte (i.e. ni trop basse ni trop élevée). Celle-ci doit au moins valoir l'espérance de Z (prime pure) sans quoi la ruine actuarielle est inévitable sur le long terme (voir chapitre suivant).

Une première façon d'envisager une telle prime P pour Z est de poser

$$P = E[Z] + \delta \cdot \sigma(Z),$$

où δ est un petit pourcentage (disons 10%). Si l'écart-type de Z est grand, il est en effet normal de demander une prime plus élevée que s'il est petit.

Une deuxième façon d'envisager une prime correcte pour Z est de poser

$$\begin{aligned} P &= E[Z] + \delta \cdot Var(Z) + \delta \cdot Cov(Z, W) \\ &= E[Z] + \delta \cdot Cov(Z, W + Z), \end{aligned}$$

où W est le risque déjà assuré et pour lequel une prime existe déjà. Cette formule est en substance la même que la précédente sauf que l'espérance de Z est augmentée d'un pourcentage non pas de son écart-type mais de sa variance, une considération sur l'éventuelle dépendance entre W et Z venant ensuite s'ajouter à la précédente considération. Il est à noter que la question de savoir s'il faut ajouter à $E[Z]$ un pourcentage de l'écart-type de Z ou au contraire de sa variance peut diviser les esprits. En fait, il faut considérer que dans le cas où Z et W sont stochastiquement indépendantes, les variances s'additionnent tandis que dans le cas où Z et W sont totalement dépendantes – leur coefficient de corrélation valant alors 1 – ce sont les écarts-types qui s'additionnent. A partir de là, on peut choisir.

Le risque déjà assuré W n'est pas nécessairement stochastiquement indépendant du nouveau risque Z que nous voulons ajouter à notre portefeuille de risques et pour lequel nous cherchons à déterminer la prime P . Si l'un des risques Y de notre portefeuille est fortement corrélé à Z , le coefficient de corrélation valant – disons – un peu moins de 1, alors la probabilité de leur occurrence conjointe est, à peu de choses près, égale à $\text{Prob}(Z)$, alors que la somme des primes de Y et de Z est pensée pour couvrir une occurrence conjointe de $\text{Prob}(Y) \cdot \text{Prob}(Z) \ll \text{Prob}(Z)$. Cette réflexion explique pourquoi il faut majorer la prime de Z d'une grandeur qui dépend de la corrélation ou de la covariance.

La détermination de la covariance promet par contre d'être difficile et on préférera dans la pratique la première approche. Là où la considération de la covariance fait par contre vraiment sens, c'est dans la détermination d'une prime de réassurance, étant clair que les établissements de réassurance peuvent, au travers de montages pouvant être complexes, se réassurer indirectement plusieurs fois sans que cela ne saute aux yeux.

Quoi qu'il en soit, il faut pouvoir déterminer, à partir des données à disposition sur les occurrences passées de la variable aléatoire Z , son espérance $E[Z]$.

En pratique, nous avons à déterminer l'espérance de Z pour une catégorie de risque donnée. Pour fixer un cadre, on se donne un indice $i \in \overline{n}$ qui court sur les années, l'année 0 se situant dans le futur et étant l'année pour laquelle il faut calculer une prime et l'année 1 étant l'année actuelle et la première pour laquelle des données statistiques sont disponibles. On se donne aussi un indice $j \in \overline{m}$ qui court sur les catégories de risques. On dispose des données statistiques des occurrences de chacune des variables aléatoires Z_{ij} et notre but est de déterminer $E[Z_{.j}]$, la moyenne espérée sur plusieurs années par catégorie de risque.

Comment estimer $E[Z_{.j}]$? La réponse semble pourtant aller de soi : le meilleur estimateur statistique $\hat{\mu}_j$ pour l'espérance μ_j est la moyenne arithmétique $z_{.j} := \frac{1}{N_j} \sum_{i=1}^n z_{ij}$, où chaque z_{ij} est la moyenne empirique des occurrences de la variable aléatoires Z_{ij} pour l'année i et $N_j = \sum_i n_{ij}$ est la somme des nombres d'occurrences annuels. Y aurait-il ici matière à difficulté ?

En fait, on peut ici prendre une moyenne d'au moins deux façons qui font sens.

On peut d'une part et bien entendu estimer $E[Z_{.j}]$ avec $z_{.j}$. Ce faisant, on considère implicitement que les individus de deux catégories de risque différentes diffèrent du point de vue de leur profil de risque de façon essentielle, dans le sens que cette différence serait inscrite dans leur nature (ou essence) même. Les femmes seraient par nature plus ou moins personnes à risques que les hommes, par exemple. En effet, cette façon de calculer l'estimateur $E[Z_{.j}]$ suppose que pour deux catégories de risque j_1 et j_2 , la différence entre $E[Z_{.j_1}]$ et $E[Z_{.j_2}]$ ne relève pas du tout du hasard. Ce point est important et justifie les développements ultérieurs, il convient donc de le méditer.

Y réfléchir conduit à interroger la pertinence du découpage d'une population en telles catégories de risque plutôt que d'autres. Il existe d'ailleurs des algorithmes de *clustering* qui ont pour but d'obtenir une répartition en classes d'une population qui soit optimale pour un critère donné, ce qui veut bien dire qu'une partie du problème de l'actuaire consiste à compenser le fait de devoir travailler à partir de catégories de risque fixées, par exemple par la loi.

On peut d'autre part considérer que toute différence entre classes de risques n'est que fortuite et prendre $z_{..} := \frac{1}{N} \sum_{j=1}^m \sum_{i=1}^n z_{ij}$, avec $N = \sum N_j$, comme estimateur de $E[Z_{.j}]$. Dans ce cas-là, on suppose que toutes les classes tendent à se ressembler les unes aux autres à mesure que chacune d'elle regroupe un nombre d'individus toujours plus grand.

La première façon de voir les choses est naïve du point de vue de la pure compréhension statistique (c'est la pratique naïve habituelle) et pour le moins discriminatoire. La seconde façon de voir les choses est naïve du point de vue de la vision sociale, trop prompt à niveler les différences.

Comme souvent, la vérité est ailleurs et voici comment il est possible de faire la part des choses.

Étape 1

Soit T_{ij} le volume⁹ correspondant à Z_{ij} . On considère la grandeur relative $X_{ij} := Z_{ij}/T_{ij}$, grandeur qui correspond au *loss ratio* lorsque Z_{ij} est une quantité de numéraire et qu'on nomme dorénavant ainsi dans tous les cas, par un abus de langage bien pratique. Le numérateur de ce loss ratio est une variable aléatoire tandis que son dénominateur est un nombre.

Nous voulons dès à présent estimer non plus $E[Z_{ij}]$ mais $E[X_{ij}]$, étant clair que si l'on connaît celui-ci, on revient facilement à celui-là.

Nous posons que chaque classe de risque est déterminée par un paramètre de risque ν_j de façon que

$$\text{Prob}(X_{ij} \leq x \mid \nu_j = \nu) = F_\nu(x, T_{ij}),$$

c'est-à-dire de façon que la fonction de répartition du loss ratio ne dépende que de ce paramètre et du volume T_{ij} (le membre de droite de l'équation ci-dessus ne dit rien d'autre), avec, et c'est là l'important

$$E[X_{ij} \mid \nu] = \mu(\nu) \quad (*)$$

et

$$\begin{aligned} \text{Var}(X_{ij} \mid \nu) &= \frac{\sigma^2(\nu)}{T_{ij}} \\ \Leftrightarrow \frac{1}{T_{ij}^2} \text{Var}(Z_{ij} \mid \nu) &= \frac{\sigma^2(\nu)}{T_{ij}} \\ \Leftrightarrow \frac{\text{Var}(Z_{ij} \mid \nu)}{T_{ij}} &= \sigma^2(\nu) \quad (**) \end{aligned}$$

c'est-à-dire de façon que l'espérance de cette fonction de répartition ne dépende que de ce paramètre ν (équation (*)) et de façon que la variance de Z_{ij} soit proportionnelle¹⁰ au volume (équation (**)).

Que la démarche soit ici claire : ces exigences qui concernent la fonction de répartition du loss ratio portent en réalité sur sa dépendance par rapport au volume : on fait varier le volume et on se demande comment doit varier cette fonction de répartition. Que la fonction de répartition du loss ratio dépende d'un paramètre semble par contre tout à fait clair et exiger ceci relève davantage du besoin de fixer les notations en termes d'espérance conditionnelle (dont l'utilité n'apparaîtra qu'à la prochaine étape du développement) que d'une limitation particulière.

⁹ Qu'est-ce donc que ce volume ? Si Z_{ij} est le nombre de cas d'accidents (ou la somme déboursée par l'assurance) durant l'année i pour la catégorie de risque j , alors le volume T_{ij} est le nombre total de personnes assurées (respectivement la somme totale assurée) dans la catégorie de risque j durant l'année i . De façon analogue, T_j est le volume correspondant à Z_j .

¹⁰ En effet, écrire que

$$\frac{a(t)}{b(t)} = \text{constante ne dépendant pas de } t$$

signifie que a varie proportionnellement à b .

Exemple : Le nombre de cas d'accidents pour l'année i et dans la catégorie j est $Z_{i,j} \sim \mathcal{P}(\lambda_j)$. Son espérance est λ_j . Si l'on se met à faire varier le volume $T_{i,j}$, on modifie la variable aléatoire $Z_{i,j}$ puisqu'elle est définie pour un volume donné. Si par exemple on double le volume, l'espérance de la variable ainsi modifiée est $2\lambda_j$. C'est pourquoi nous choisissons de fixer que le nombre $Z_{i,j}^u$ de cas d'accidents pour un volume unitaire est de loi $\mathcal{P}(\lambda_j)$ de façon à pouvoir ensuite affirmer que $E[Z_{i,j}] = T_{i,j} \cdot \lambda_j$. De là, on vérifie que l'espérance du loss ratio ne dépend pas du volume (condition (*)) puisque $E[X_{i,j}] = \frac{E[Z_{i,j}]}{T_{i,j}} = \lambda_j$ et que la variance du nombre de cas d'accidents est proportionnelle au volume (condition (**)). En effet,

$$\frac{Var(Z_{i,j})}{T_{i,j}} = \frac{Var\left(\sum_{u=1}^{T_{i,j}} Z_{i,j}^u\right)}{T_{i,j}} = \frac{T_{i,j} \cdot Var(Z_{i,j}^u)}{T_{i,j}} = \lambda_j,$$

où les $Z_{i,j}^u$ sont deux à deux stochastiquement indépendantes¹¹ pour deux unités de volume u différentes, chaque $Z_{i,j}^u$, $u \in \overline{T_{i,j}}$, décrivant le nombre de cas d'accidents pour une unité de volume et le nombre de cas d'accidents pour un volume $T_{i,j}$ s'obtenant en sommant les $Z_{i,j}^u$ pour u allant de l'unité de volume numérotée 1 jusqu'à celle numérotée $T_{i,j}$.

Étape 2

On considère que pour chaque catégorie de risque, le paramètre ν_j n'est pas simplement un nombre inconnu bien que déterminé mais est un nombre aléatoire, c'est-à-dire essentiellement indéterminé. On pose ainsi que les ν_j sont autant de variables aléatoires i.i.d. dont il faut déterminer la répartition ou du moins l'espérance et la variance.

Ce qu'on a écrit ci-dessus concernant l'espérance du loss ratio :

$$E[X_{ij} | \nu] = \mu(\nu)$$

se fait modifier de la façon suivante. Le ν , qui était un nombre considéré comme fixé, devient une variable aléatoire et alors que jusqu'ici on ne nécessitait pas le concept d'espérance conditionnelle, celui-ci est à présent incontournable¹² :

$$E[X_{ij} | \nu_j] = \mu(\nu_j).$$

Puisque les ν_j sont i.i.d., on peut fixer la classe de risque en posant $j = k$, pour k un nombre fixé. Mais comment progresser dans le raisonnement, puisqu'on ignore à la fois X_{ij} et ν_j ?

¹¹ Lorsque deux variables aléatoires sont stochastiquement indépendantes, la variance de leur somme est égale à la somme de leurs variances. C'est ce qui permet la deuxième égalité ci-dessus puisqu' alors $Var\left(\sum_{u=1}^{T_{i,j}} Z_{i,j}^u\right) = T_{i,j} \cdot Var(Z_{i,j}^u)$. L'exigence que la variance de $Z_{i,j}$ soit proportionnelle au volume est donc en fait une exigence d'indépendance stochastique des risques individuels à l'intérieur de chaque catégorie de risque.

¹² Alors qu'une espérance est un nombre déterminé, une espérance conditionnelle est quant à elle une variable aléatoire. Le lecteur doit ici s'efforcer d'être au clair avec cette notion qui revient sans cesse par la suite.

Dans pareil cas, la solution est d'exiger que l'objet mathématique recherché prenne une forme particulière (on appelle cela faire un Ansatz) : on exige que l'estimateur de cette espérance conditionnelle s'écrive sous la forme d'une combinaison linéaire

$$\hat{\mu}_k = \sum_{i,j} {}_k\alpha_{ij} X_{ij}$$

qui satisfasse la condition de carré minimum

$$E[(\hat{\mu}_k - \mu(v_k))^2] = \text{minimum}$$

et qui soit non biaisée

$$E[\hat{\mu}_k] = E[\mu(v_k)] =: m.$$

On cherche les coefficients ${}_k\alpha_{ij}$ qui satisfont à cette attente en formant l'expression suivante :

$$\mathcal{L}({}_k\alpha_{ij}, c) = \frac{1}{2} E \left[\left(\sum_{i,j} {}_k\alpha_{ij} X_{ij} - \mu(v_k) \right)^2 \right] + c \left(E \left[\sum_{i,j} {}_k\alpha_{ij} X_{ij} \right] - m \right)$$

et en posant égales à zéros ses dérivées partielles par rapport à ${}_k\alpha_{ij}$ et à c .

Le coefficient c est ce qu'on appelle un multiplicateur de Lagrange. Le premier terme de cette expression est une fonction dont on veut le minimum et le second terme, celui qui est un multiple de c , est une fonction de contrainte. La méthode des multiplicateurs (ici au singulier) de Lagrange est applicable puisque la fonction à minimiser et la fonction de contrainte sont toutes deux convexes.

Durant les calculs, pénibles, que l'on retrouve en détails dans E. Straub : *Non-Life Insurance Mathematics*, Springer Verlag, on procède essentiellement de la façon suivante :

1) On fait à tout bout de champ usage de l'identité suivante :

$$E[E[X | Y]] = E[X],$$

2) On pose $\frac{\partial \mathcal{L}}{\partial c} = 0$ et on déduit facilement que $\sum_{i,j} {}_k\alpha_{ij} = 1$.

3) On pose $\frac{\partial \mathcal{L}}{\partial {}_k\alpha_{ij}} = 0$ et on applique le résultat du point 2) pour transformer l'égalité posée de façon à obtenir cm .

4) On introduit la valeur obtenue pour cm dans l'équation $\frac{\partial \mathcal{L}}{\partial {}_k\alpha_{ij}} = 0$ et on obtient péniblement les coefficients ${}_k\alpha_{ij}$. On connaît alors tous les coefficients et on a déjà virtuellement fini.

5) On exprime l'estimateur $\hat{\mu}_k = \sum_{i,j} {}_k\alpha_{ij} X_{ij}$ à l'aide des coefficients obtenus et obtient ainsi une expression indigeste qu'on s'attache à simplifier. Il faut pour cela avoir identifié une certaine expression fractionnaire qui revient en plusieurs endroits dans l'expression indigeste; alors on peut alors avoir l'idée de donner un nom de coefficient à cette expression fractionnaire et réécrire le tout en plus compact.

L'estimateur $\hat{\mu}_k$ que l'on a ainsi obtenu donne la **prime pure** recherchée. Disposant de celle-ci, on procède ensuite selon les explications données à la page 8 de ce document.

Remarques : La méthode décrite correspond au fameux modèle Bühlmann-Straub (~1970) en *théorie de la crédibilité*. Il y a tout de même eu nombre de développements théoriques depuis, un point notamment pouvant être vu comme une limitation du modèle de Bühlmann-Straub étant que celui-ci pose (Ansatz) que l'estimateur recherché $\hat{\mu}_k$ est une combinaison linéaire des X_{ij} . C'est pourtant cette limitation qui permet l'application de la méthode des multiplicateurs de Lagrange en rendant la fonction de contrainte $(X_{ij}) \mapsto E[\hat{\mu}_k] - m$ linéaire et donc convexe.

La détermination de la probabilité de ruine

Soit Z_t , de loi de Poisson composée, les frais accumulés dus aux sinistres qui sont arrivés durant l'intervalle de temps $[0, t]$, soit $P_t := P \cdot t$ le volume de primes collecté durant cette période et U les réserves initiales. La ruine arrive lorsque il existe un moment t' dans $[0, t]$ tel que

$$Z_{t'} - P_{t'} > U.$$

On définit donc la probabilité de ruine durant l'intervalle de temps T de la façon suivante :

$$\Psi_T(U, P) := P(\sup_{t \in T} (Z_t - P_t) > U).$$

Selon la nature du temps que l'on considère, temps discret ou continu, à l'horizon borné ou temps infini, on distingue différentes expressions de la probabilité de ruine. Exemple de temps discret borné : $T = \{t ; t = ih, i = 0, 1, 2, \dots, n\}$ et exemple de temps continu infini : $T = [0, \infty[$.

Pour un temps continu et infini, on cherche une formule pour la probabilité de ruine $\Psi_T(U, P)$.

Puisque Z_t suit une loi de Poisson composée, il existe un paramètre λ pour lequel le temps entre chaque sinistre est une variable aléatoire de loi exponentielle de densité $f(t) = \frac{e^{-\lambda t}}{\lambda}$. Quant à la densité du montant d'un sinistre particulier, elle peut être quelconque et est dénotée $v(x)$.

Soit t l'instant et x le montant du premier sinistre du processus stochastique Z_t . Alors on peut affirmer qu'**on ne se trouve pas** en état de ruine actuarielle lorsque les deux événements suivants arrivent de façon conjointe :

$$\begin{aligned} x &\leq U + Pt && \text{et} \\ Z_s - x &\leq U - x + Ps, && \forall s > t. \end{aligned}$$

La première condition stipule que la ruine n'arrive pas lors du premier sinistre et la seconde condition, que la ruine n'arrive pas par après non plus. Or la densité de probabilité du premier de ces deux événements est

$$f_1(t, x) = \lambda e^{-\lambda t} \cdot v(x),$$

c'est-à-dire une densité définie sur le produit cartésien $[0, \infty] \times \mathbb{R}^+$ du temps et de l'ensemble des montants auxquels peut s'élever un sinistre seul.

Cette densité permet d'exprimer la probabilité que l'événement $\{x \leq U + Pt\}$ se produise pour Δx et Δt . La probabilité du second événement est donnée quant à elle par

$$1 - \Psi_{]x, \infty[}(U - x + Pt, P).$$

En notant $\delta(U) := 1 - \Psi_{]x, \infty[}(U, P)$, la probabilité conjointe est donnée par le produit des probabilités

$$\lambda e^{-\lambda t} \cdot v(x) \Delta x \Delta t \delta(U - x + Pt).$$

En intégrant sur toutes les possibilités pour le premier sinistre :

$$\delta(U) = \int_0^{\infty} \left(\int_0^{U+Pt} \lambda e^{-\lambda t} \cdot v(x) \delta(U - x + Pt) dx \right) dt.$$

$\delta(U)$ donne la probabilité de ne jamais être ruiné durant le temps infini continu $[0, \infty[$ en fonction des réserves initiales U . Cependant cette intégrale ne se laisse pas du tout évaluer et il faut encore déployer beaucoup d'ingéniosité pour lui donner une forme utilisable.

Un mot sur l'élégance du développement proposé : alors qu'on recherche une probabilité $\Psi_T(U, P)$ dont on ignore tout, on arrive tout de même à obtenir une équation implicite non triviale ! Cela passe par l'idée de « couper » l'événement inconnu en deux événements indépendants dont la densité de l'un des deux est connue et dont la conjonction doit être vérifiée, ce dernier point amenant à travailler avec la probabilité complémentaire $\delta(U) := 1 - \Psi_T(U, P)$, puisqu'on ne peut le faire directement avec $\Psi_T(U, P)$.

Le gros problème à résoudre à ce stade du développement est que l'équation ci-dessus ne donne la fonction δ que de façon implicite, puisque celle-ci se retrouve de part et d'autre de l'égalité.

La dérivation en U de part et d'autre de l'égalité fournit une nouvelle égalité. Afin de pouvoir dériver en U , il faut auparavant modifier la borne supérieure d'intégration du membre de droite de l'équation. C'est pourquoi on fait le changement de variable suivant : $r := U + Pt$.

Calculs intermédiaires : $t(r) = \frac{r-U}{P}$ et $t'(r) = \frac{1}{P}$. Quant aux bornes d'intégration, $t^{-1}(0) = U$ et $t^{-1}(\infty) = \infty$. Le changement de variable donne ainsi :

$$\delta(U) = \frac{\lambda}{P} \int_U^{\infty} \left(e^{-\frac{\lambda}{P}(r-U)} \cdot \int_0^r \delta(r-x)v(x)dx \right) dr.$$

Afin de calculer la dérivée en U de cette expression, il faut tout d'abord en remarquer la (une) structure. Ceci permet de décomposer le travail de dérivation et d'appliquer les bonnes règles dans le bon ordre sans rien oublier.

Puisque r dépend de U , on doit commencer par repérer dans l'expression à dériver toutes les occurrences de ces deux variables, et rien qu'elles. En procédant ainsi, on ne manque de remarquer que U se trouve à la borne inférieure de la première des deux intégrales. Or il se trouve qu'on ne sait traiter que la situation où U se trouve à la borne supérieure d'intégration. On s'arrange donc pour faire passer le U de la borne inférieure à la borne supérieure :

$$\delta(U) = \delta(U) = \frac{\lambda}{P} \int_U^{\infty} \dots = \frac{\lambda}{P} \left(\int_U^0 \dots + \int_0^{\infty} \dots \right) = \frac{\lambda}{P} \left(- \int_0^U \dots + \int_0^{\infty} \dots \right).$$

On peut dériver le premier de ces deux termes. En le dénotant par $F(U)$, cela donne :

$$\frac{d}{dU} F(U) = -\frac{\lambda}{P} \cdot \frac{d}{dU} \left[\int_0^U \left(e^{-\frac{\lambda}{P}(r-U)} \cdot \int_0^r \delta(r-x)v(x)dx \right) dr \right].$$

On traite le fait que U est à la fois borne supérieure d'intégration et variable sous l'intégrale en exprimant F comme la fonction composée suivante :

$$F(a) := F_1 \circ \begin{bmatrix} a \\ a \end{bmatrix}, \quad \text{où } F_1 : \mathbb{R}^2 \rightarrow \mathbb{R}; \quad F_1(a, b) := -\frac{\lambda}{P} \cdot \int_0^a \left(e^{-\frac{\lambda}{P}(r-b)} \cdot \int_0^r \delta(r-x)v(x)dx \right) dr.$$

On dérive ainsi F comme une fonction composée :

$$dF(a) \Big|_{a=U} = dF_1(a, b) \Big|_{\substack{a=U \\ b=U}} \circ \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{\partial}{\partial a} F_1(a, b) \Big|_{\substack{a=U \\ b=U}} + \frac{\partial}{\partial b} F_1(a, b) \Big|_{\substack{a=U \\ b=U}},$$

où, en définissant $F_2(a, b) := e^{-\frac{\lambda}{P}(b-a)} \cdot \int_0^b \delta(b-x)v(x)dx$,

$$\frac{\partial}{\partial a} F_1(a, b) \Big|_{\substack{a=U \\ b=U}} = \frac{\partial}{\partial a} \int_0^a F_2(b, r) dr \Big|_{\substack{a=U \\ b=U}} = F_2(a, a) = \underbrace{e^{-\frac{\lambda}{P}(U-U)}}_{=1} \cdot \int_0^U \delta(U-x)v(x)dx$$

et où

$$\frac{\partial}{\partial b} F_1(a, b) \Big|_{\substack{a=U \\ b=U}} = \int_0^a \left(\frac{\partial}{\partial b} F_2(b, r(b)) \right) dr \Big|_{\substack{a=U \\ b=U}}.$$

On dérive en b la fonction F_2 :

$$\frac{\partial}{\partial b} F_2(b, r(b)) = \left[\frac{\partial}{\partial b} F_2(b, r) \quad \frac{\partial}{\partial r} F_2(b, r) \right] \cdot \begin{bmatrix} 1 \\ r'(b) \end{bmatrix} = \frac{\partial}{\partial b} F_2(b, r) + \frac{\partial}{\partial r} F_2(b, r) \cdot r'(b).$$

Par suite, et en remarquant que la présence de la dérivée de la variable d'intégration r ne pose pas de problème puisque $r'(b)$ tient le rôle d'une constante (c'est r' évalué en b) :

$$\begin{aligned} \frac{\partial}{\partial b} F_1(a, b) \Big|_{\substack{a=U \\ b=U}} &= \int_0^a \left(\frac{\partial}{\partial b} F_2(b, r) + \frac{\partial}{\partial r} F_2(b, r) \cdot r'(b) \right) dr \Big|_{\substack{a=U \\ b=U}} \\ &= \int_0^U \left(\frac{\partial}{\partial b} F_2(b, r) \Big|_{b=U} \right) dr + \int_0^U \left(\frac{\partial}{\partial r} F_2(U, r) \cdot r'(U) \right) dr. \end{aligned}$$

La fonction F_2 est un produit de deux fonctions $F_2(a, b) := F_3(a, b) \cdot F_4(b)$ avec $F_3(a, b) := e^{-\frac{\lambda}{P}(b-a)}$ et $F_4(b) := \int_0^b \delta(b-x)v(x)dx$; on peut donc écrire :

$$\frac{\partial}{\partial b} F_2(b, r) \Big|_{b=U} = \frac{\partial}{\partial b} F_3(b, r) \Big|_{b=U} \cdot F_4(r)$$

et

$$\frac{\partial}{\partial r} F_2(U, r) = \frac{\partial}{\partial r} F_3(U, r) \cdot F_4(r) + F_3(U, r) \cdot \frac{\partial}{\partial r} F_4(r).$$

Ainsi,

$$\left. \frac{\partial}{\partial b} F_2(b, r) \right|_{b=U} = \frac{\lambda}{P} \cdot e^{-\frac{\lambda}{P}(r-U)} \cdot \int_0^r \delta(r-x)v(x)dx$$

et

$$\frac{\partial}{\partial r} F_2(U, r) = -\frac{\lambda}{P} \cdot e^{-\frac{\lambda}{P}(r-U)} \cdot \int_0^r \delta(r-x)v(x)dx + e^{-\frac{\lambda}{P}(r-b)} \cdot \int_0^b \left(\frac{\partial}{\partial r} \delta(b-x) \right) v(x)dx.$$

Bref, ces calculs sont compliqués. Au final, on doit obtenir

$$\delta'(U) = -\frac{\lambda}{P} \int_0^U \delta(U-x) v(x)dx + \frac{\lambda}{P} \delta(U).$$

On intègre cette expression sur l'intervalle $]0, y[$:

$$\delta(y) - \delta(0) = \frac{\lambda}{P} \int_0^y \delta(y-x) \left(1 - \int_0^x v(t)dt \right) dx.$$

On a l'identité suivante :

$$\int_0^\infty 1 - \left(\int_0^x v(t)dt \right) dx = \int_0^\infty x \left(\int_0^x v(t)dt \right) dx,$$

Ce qui est égal à $\mu = E[X]$, c'est-à-dire au montant espéré pour un sinistre individuel.

On a que

$$\Psi(0, P) = \frac{\lambda\mu}{P},$$

indépendamment de la fonction de répartition $\int_0^x v(t)dt$, ce qui signifie que lorsque les réserves sont nulles ($U = 0$), la probabilité de ruine est la prime pure $\lambda\mu$ divisée par la prime P . Ceci signifie que la ruine est certaine lorsque $P \leq \lambda\mu$, c'est-à-dire lorsque la prime est inférieure aux coûts attendus, ce qu'il fallait démontrer.

Source bibliographique : Erwin Straub : *Non-Life Insurance Mathematics*, réimpression 1997, Springer Verlag (136 pages)