Université Batna2
Département de Mathématiques
2eme année master statistique et analyse des données
Année universitaire 2022-2023

Mathématiques pour l’assurance

Master 02 SAD

(Partie 1)

Dans l'industrie, lorsqu'un nouveau produit est mis sur le marché, on Connaît avec précision le coût de production et peut en conséquence proposer un prix de vente pour son bien. Par contre, le chiffre d'affaire de l'entreprise est a priori inconnu, car il dépend de la capacité de celle-ci à vendre ses produits. Au contraire, lorsque l'assureur fixe la prime, il ne connaît pas le montant des sinistres ni les frais de gestion occasionnes par ceux-ci. Sa situation est donc toute différente des autres industries, puisque l'assureur connaît a priori son chiffre d'affaire (le volume des primes encaissées), mais pas le prix de revient des produits qu'il commercialise (i.e. ses charges futures).
Il s'agit du principe de **l'inversion du cycle de production**. Il faut parfois de longues années avant que la compagnie ne soit en mesure d'évaluer exactement le profit engendré par un produit d'assurance du fait des cadences de règlement très lentes des sinistres dans certaines branches.
Cette spécificité a deux conséquences.
— La première est la nécessité de mettre en place des outils mathématiques sophistiqués afin d'évaluer le montant de la prime à demander à l'assuré pour le protéger du risque et éviter les pertes pour l'assureur. Ces outils sont principalement statistiques et probabilistes et utilisent les données historiques pour cerner la variabilité des risques.
— la seconde est que l'assurance est extrêmement dépendante des données connues par l'assureur d'une part et l'assuré d'autre part sur le risque couvert par un contrat. Des asymétries d'informations entre les deux protagonistes sont régulièrement constatées expliquent une partie des règles qui encadrent l'activité d'assurance.

1 Modélisation probabiliste du risque (Description probabiliste du risque)

1.1 Évènements

Une notion fondamentale pour la suite sera celle **d'évènement aléatoire**. Il s'agit d'évènements dont on ne peut prédire avec certitude s'ils se réaliseront ou pas. C'est à ce genre d'évènements que sont subordonnées les prestations de l'assureur. Clairement,
1. les évènements considérés dépendent de la définition des garanties promises par l'assureur.
2. la description des évènements doit être exhaustive, mais sans double emploi.

Évènements élémentaires

Définissons les évènements élémentaires $e_1, e_2, e_3, \ldots$ comme étant ceux satisfaisant aux deux propriétés suivantes:

- deux évènements élémentaires distincts e_i et e_j , $j \neq i$, sont incompatibles, c'est-à dire qu'ils ne peuvent se réaliser simultanément.
- La réunion $\varepsilon = \{e_1, e_2, \ldots\}$ de tous les évènements élémentaires $e_1, e_2, \ldots$ correspond à la certitude, c'est-à-dire à l'évènement qui se réalise toujours.

Associons à présent à chaque évènement l'ensemble des résultats élémentaires qui conduisent à sa réalisation. Tout évènement E peut ainsi s'exprimer comme la réunion d'évènements élémentaires e_{i1} ,…e_{ik} , c'est-à-dire que E se réalise lorsque soit e_{i1}, soit e_{i2} ,… soit e_{ik} se réalise. Ceci se notera dorénavant $E = \{e_{i1}, e_{i2}, \dots, e_{ik}\}$.

Notons $\mathcal{A}$ l'ensemble des évènements qui conditionnent le risque (C'est-_à-dire ceux nécessaires pour déterminer les prestations de l'assureur). Ainsi, $\mathcal{A}$ est une collection de sous-ensembles de $\mathcal{E}$, collection à laquelle nous imposons de satisfaire les trois conditions suivantes

(qui font de $\mathcal{A}$ ce que les probabilistes appellent une sigma-algèbre ou tribu):

(i) l'évènement impossible $\emptyset \in \mathcal{A}$ et l'évènement certain $\mathcal{E} \in \mathcal{A}$

(ii) si $E_1, E_2, \ldots \in \mathcal{A}$ alors $\cup_{i\geq 1} E_i \in \mathcal{A}$

(iii) si $E \in \mathcal{A}$ alors $\overline{E} \in \mathcal{A}$

Les trois conditions imposées ci-dessus garantissent en fait la cohérence du raisonnement, autorisant à parler d'évènements qui se produisent lorsque plusieurs se produisent simultanément (intersection) ou lorsqu'au moins l'un d'entre eux se produit (union).

À chaque évènement E de $\mathcal{A}$ on souhaite associer un nombre, noté Pr[E] et appelé probabilité de E. Ce nombre mesure le degré de vraisemblance qu'on accorde a priori à la réalisation de E: il est non-négatif, et est d'autant plus grand que l'évènement est jugé vraisemblable. Le triplet (E,$\mathcal{A}$, Pr) est appelé espace probabilisé alors que le couple (E,$\mathcal{A}$) est appelé espace probabilisable en théorie des probabilités.

La probabilité Pr[E] associée à l'évènement E peut s'envisager comme la **prime** à payer pour recevoir l'indemnité versée par l'assureur en cas de réalisation de E (Les sommes versées par le preneur d'assurance (ou souscripteur) pour avoir droit à la garantie de l'assureur sont appelées primes ou parfois cotisations) pourvu que le marché de l'assurance réponde à un minimum de rationalité (à savoir qu'il est impossible de faire du profit sans prendre de risque). Ainsi, si E « il pleut sur Batna le 31 Décembre », Pr[E] est la somme à payer pour recevoir 1000 Da si, le 31 Décembre, la pluie tombe. De manière évidente, si l'assuré ne veut plus recevoir 1000 DA mais c DA en cas de réalisation de l'évènement E, il lui suffit de payer une prime c Pr[E]; Pr[.] s'apparente donc à un taux de prime.

Clairement, la probabilité de tout évènement est inférieure à 1. En effet, aucun assuré rationnel ne serait prêt à payer une prime supérieure à l'éventuelle indemnité versée par l'assureur. Ainsi, quel que soit $E \in \mathcal{A}$, $Pr[E] \leq 1$. On cherche donc à attribuer à toute éventualité une probabilité, nombre compris entre 0 et 1, d'autant plus faible que la survenance de l'évènement est rare.

Nous supposerons dorénavant le marché de l'assurance complet, c'est-à-dire qu'il est possible d'acheter ou de vendre une police garantissant le versement d'1000DA en cas de réalisation de n'importe quel évènement aléatoire.

1.2 Absence d'opportunité d'arbitrage.

Comme les financiers en ont l'habitude, on appelle opportunité d'arbitrage la possibilité de faire du profit sans risque. L'idée est la suivante: si le système de prime Pr[.] adopté par les acteurs du marché ne satisfait pas à un minimum de rationalité, il devient possible pour certains acteurs de réaliser des opérations générant un bénéfice certain, sans mise initiale. Une telle situation, même si elle n'est pas formellement exclue en pratique, ne pourrait en tout cas perdurer sur un marché sain. Grâce à ce principe simple, nous allons établir une série de propriétés auxquelles doivent satisfaire les probabilités.

Propriété d'additivité pour évènements incompatibles

Si E et F sont deux évènements incompatibles (i.e. qui ne peuvent survenir simultanément, ce qui se notera désormais $E \cap F = \emptyset$)
pour lesquels on a défini deux probabilités Pr[E] et Pr[F], i.e. on a déterminé les sommes à payer pour recevoir 1DA en cas de survenance de ces évènements, alors

$Pr[E \cup F] = Pr[E] + Pr[F]$

Montrons qu'il ne peut en être autrement. Pour ce faire, voyons ce qu'il adviendrait si $Pr[E \cup F] > Pr[E] + Pr[F]$. Dans ce cas, l'assureur émet une police garantissant le versement de 1DA en cas de réalisation de E ou de F. Avec la prime $Pr[E \cup F]$ encaissée, il souscrit deux polices (auprès d'un concurrent), la première garantissant le versement d'1DA si E se réalise et la seconde le même versement si F se réalise. Cela lui coûte Pr[E]+Pr[F] et ne nécessite aucune mise initiale. En fin de période, trois situations peuvent se présenter, Dans tous les cas, l'assureur réalise un profit positif s'élève donc a,

$Pr[E \cup F] - Pr[E] - Pr[F] > 0$

Pour une mise initiale nulle, ce qui contredit l'hypothèse d'absence d'opportunité d'arbitrage.

Le raisonnement s'adapte facilement au cas où $Pr[E \cup F] < Pr[E] + Pr[F]$.
Il suffit en effet d'émettre deux polices, l'une prévoyant le versement d'1DA si E se réalise, et l'autre le même versement en cas de survenance de F. L'assureur encaisse alors Pr[E] +Pr [F], somme avec laquelle il souscrit une police lui procurant 1DA si E ou F se réalise

Plus généralement, considérant une suite d'événements E1, E2, E3,… deux a deux incompatibles (i.e. quels que soient i et j, E_i et E_j ne peuvent se réaliser simultanément, $E_i \cap E_j = \emptyset$; pour tout $i \neq j$)
on a:

$$Pr[U_{k\geq 1}E_k] = \sum_{k\geq 1} Pr[E_k]$$

Ceci nous permet de calculer la probabilité de tout évènement $E = \{e_{i1}, \dots, e_{ik}\}$ en fonction des probabilités des événements élémentaires $e_{i1}, \dots, e_{ik}$ qui le composent

$$Pr[E] = Pr[\{e_{i1}\} + \cdots + Pr\{e_{ik}\}]$$

Il suffit donc de définir la probabilité de chaque évènement élémentaire pour en déduire la probabilité de tout évènement E.

Croissance des primes avec le risque

La condition d'additivité garantit également la croissance des primes avec la vraisemblance des évènements assurés. Précisément, considérons deux évènements E et F. Si la réalisation de E entraine la réalisation de F (i.e. si E a plus de chance de se produire que F), on doit raisonnablement avoir Pr[E] ≤Pr[F]. En effet, l'assureur ayant plus de chance de devoir verser l'indemnité de 1DA, il réclamera une prime de montant plus élevé. Nous noterons dorénavant E ⊆ F lorsque la réalisation de F implique celle de E. Dans ce cas,

$$F = (F \cap E) \cup (F \cap \bar{E}) = E \cup (F \setminus E)$$

Qui exprime le fait que F se réalise si soit E, soit $F\setminus E$ se réalise.

On a donc exprimé F comme la réunion de deux évènements incompatibles.

La relation d'additivité nous permet alors d'écrire

Pr [F] = Pr[E] + Pr[$F\setminus E$] ≥ Pr[E]:

De sorte qu'il sera alors toujours plus onéreux de s'assurer contre la réalisation de F, i.e.

$$E \subseteq F \Rightarrow Pr[E] \leq Pr[F]$$

Propriété d'équité

Si E et F sont deux évènements complémentaires (i.e. F = $\bar{E}$)

alors

Pr[E] + Pr[F] = 1:

En effet, ayant souscrit une police prévoyante le versement d'1DA si E se réalise et une autre le même versement si E ne se réalise pas, on est certain de toucher 1DA quoi qu'il arrive. Il serait donc inéquitable que pour cette opération certaine on ait payé plus ou moins d'1DA.

Formellement, comme $E \cap F = \emptyset, Pr[E \cup F] = Pr[E] + Pr[F]$

On obtient le résultat annoncé en remarquant que $Pr[E \cup F] = Pr[\varepsilon] = 1$. Ainsi, quel que soit l'_évènement E, Pr[E] = 1 - Pr[$\bar{E}$].

Comme$\bar{\varepsilon} = \theta$, on déduit de la propriété d'équité qu'il ne coûte rien de souscrire une couverture d'assurance contre un _évènement qui ne se réalise jamais, i.e. Pr[∅] = 0. Notez que cette propriété est parfois violée par les tarifs appliqués par les assureurs. C'est le cas notamment lorsque l'état impose une certaine dose de solidarité entre assurés. Ainsi, le législateur a institué dans certains pays l'obligation pour les assureurs de couvrir les inondations dans le cadre des polices incendie. Outre l'aspect discutable d'une telle mesure (le plus souvent justifiée par des considérations budgétaires), un assuré occupant un appartement au 44eme étage d'un gratte-ciel contribuera à la réparation des dommages dus aux inondations frappant un autre assuré dont l'immeuble est situé en zone inondable.

Propriété de sous-additivité

Quels que soient les évènements E et F, nous avons Pr [E∪F] ≤ Pr[E] + Pr [F]

En effet, si on acceptait Pr [E∪F] $>$ Pr[E] + Pr [F], on créerait une opportunité d'arbitrage (c'est-à-dire une possibilité de s'enrichir sans risque), qui ne peut subsister sur un marché sain

En itérant ce résultat, on obtient facilement que quels que soient les évènements aléatoires $E_1, E_2, \ldots, E_n$, l'inégalité

$$Pr[E_1 \cup E_2 \cup \ldots \cup E_n] \leq \sum_{i=1}^{n} Pr[E_i]$$

est toujours satisfaite.

Égalité de Poincaré

Conclure deux polices, l'une sur E ∪ F et l'autre sur E∩F conduit toujours au même flux financier que la conclusion de deux polices, l'une sur E et l'autre sur F. Les primes doivent alors être égales, i.e.

Pr [E∪F] + Pr [E∩F] = Pr[E] + Pr[F]

Notons que cette relation s'écrit encore

Pr [E∪F] = Pr[E] + Pr[F] - Pr [E∩F]

Cette dernière relation est connue comme l'égalité de Poincaré

1.3 Probabilité conditionnelle

Souvent, l'assureur est amené à revoir le montant de sa prime lorsqu'il dispose d'information supplémentaire. Ayant à notre disposition de l'information, matérialisée par la connaissance qu'un évènement F s'est produit, comment réévaluer en conséquence la probabilité qu'un évènement E se réalise? Cette réévaluation est appelée probabilité conditionnelle de E sachant F, notée $Pr[E|F]$. Notons tout d'abord que comme F s'est réalisé, il est naturel d'avoir Pr[F] > 0. Sous cette condition, on peut définir la probabilité conditionnelle Comme

$$Pr[E|F] = \frac{Pr[E \cap F]}{Pr[F]} \quad \text{pour autant que } Pr[F] > 0$$

Dès que l'on sait que l'évènement F s'est produit, seuls les évènements inclus dans F conservent une chance de se réaliser. En effet, quel que soit l'évènement E, $E \cap \bar{F}$ devient impossible (et par conséquent $Pr[E \cap \bar{F}|F] = 0$) et seul E∩F conserve une chance de se produire. Ainsi, sachant que F s'est réalisé, seule la réalisation simultanée de E et F est possible, ce qui explique le numérateur. De plus, F devient un évènement certain de sorte qu'il faut « renormaliser » les probabilités de façon à ce que $Pr[F|F] = 1$, ce qui justifie la division par Pr[F].

1.4 Évènements indépendants

Bien entendu, certaines informations n'amènent pas l'assureur à réévaluer le montant de sa prime. On parle dans ce cas d'indépendance.

Ceci se traduit comme suit: les évènements E et F sont indépendants lorsque la prime réclamée pour le versement d'1DA si E se réalise est la même que l'on sache ou non si F s'est réalisé, i.e.

$$Pr[E|F] = Pr[E|\bar{F}] \Leftrightarrow Pr[E|F] = Pr[E]$$

Notez qu'en vertu de la loi de la probabilité conditionnelle ceci est encore équivalent à

$$Pr[F|E] = Pr[F]$$
ou encore à
$Pr[E \cap F]$ = Pr[E] Pr[F]:
C'est cette dernière condition qui est le plus souvent retenue pour définir l'indépendance de deux évènements (car elle présente l'avantage de s'étendre aux évènements impossibles et d'être symétrique en E et F).
Étendons à présent le concept d'indépendance à plus de deux évènements. De manière générale, ayant des évènements E_1, E_2, …,E_n, ceux-ci seront dits indépendants lorsque
$$Pr[\cap_{i\in I} E_i] = \prod_{i\in I} Pr[E_i]$$
quel que soit le sous-ensemble$I\ de\ \{1,\dots,n\}$.

1.5 Règle de multiplication (de Bayes)

On déduit facilement de l'équation de la probabilité conditionnelle que
$$Pr[E \cap F] = Pr[E|F]Pr[F]$$
Cette identité permet parfois un calcul aisé de $Pr[E \cap F]$. Cette règle s'étend à toute collection E_1, …, E_n d'évènements tels que
$$Pr[E_1, E_2, \dots, E_n] > 0$$
comme suit:
$$Pr[E_1 \cap E_2 \cap \dots \cap E_n] = Pr[E_1]Pr[E_2|E_1] \dots Pr[E_n|E_1 \cap \dots \cap E_{n-1}].$$

1.6 Évènements conditionnellement indépendants

Deux évènements E et F sont indépendants conditionnellement a un troisième évènement G si
$$Pr[E \cap F|G] = Pr[E|G]Pr[F|G]$$
Ceci est encore équivalent à exiger que
$$Pr[F|E \cap G] = Pr[F|G]$$
Puisque
$$Pr[E \cap F|G] = \frac{Pr[F|E \cap G]Pr[E \cap G]}{Pr[G]} = Pr[F|E \cap G]Pr[E|G]$$

1.7 Théorème des probabilités totales

Ayant deux évènements E et F tels que 0 < Pr[F] < 1, on voit facilement que
$$\begin{aligned} Pr[E] &= Pr[E \cap F] + Pr[E \cap \bar{F}] \\ &= Pr[E|F]Pr[F] + Pr[E|\bar{F}]Pr[\bar{F}] \end{aligned}$$
ce qui permet de déduire Pr[E] de $Pr[E|F]$ et de $Pr[E|\bar{F}]$.
Exemple. Notons E l'évènement « l'assuré cause au moins un sinistre sur l'année » et F l'évènement « l'assuré est une femme ».
Si 10% des femmes et 15% des hommes causent au moins un sinistre sur l'année, quelle est la proportion de polices causant au moins un sinistre sur l'année dans un portefeuille comportant 2/3 d'hommes et 1/3 de femmes?
Cette proportion vaut
$$Pr[E] = Pr[E|F]Pr[F] + Pr[E|\bar{F}]Pr[\bar{F}]$$

$$= 0.1 \times \frac{1}{3} + 0.15 \times \frac{2}{3} = 0.133$$

Une extension naturelle de ce résultat est connue sous le nom de théorème des probabilités totales et décrite ci-après. Considérons un système exhaustif d'évènements$\{F_1, F_2, ...\}$; un tel système est tel que deux quelconques des F_i ne peuvent se réaliser simultanément et tel qu'ils recouvrent tous les cas de figure (i.e. $Pr[\cup_i \geq F_i] = 1 \, avec \, Pr[Fi] > 0$ pour tout *i*). Alors, quel que soit l'évènement E,

$$Pr[E] = Pr[E \cap (\cup_i \geq F_i)]Pr[\cup_i \geq 1(E \cap F_i)]$$
$$= \sum_{i \geq 1} Pr[E \cap F_i] = \sum_{i \geq 1} Pr[E|F_i]Pr[F_i].$$

1.8 Théorème de Bayes

Un des résultats les plus intéressants faisant intervenir les probabilités conditionnelles est sans nul doute le théorème de Bayes.

Étant donné un système exhaustif d'évènements$\{F_1, F_2, ...\}$, et E un évènement quelconque de probabilité positive, la formule

$$Pr[F_i|E] = \frac{Pr[E|F_i]Pr[F_i]}{Pr[E]}$$
$$= \frac{Pr[E|F_i]Pr[F_i]}{\sum_{j \geq 1} Pr[E|F_j]Pr[F_j]} \qquad i = 1, 2, ...,$$

est valable, où le dénominateur est obtenu à l'aide de la formule des probabilités totales.

Exemple.

Afin de détecter la fraude, beaucoup d'assureurs recourent à des programmes informatiques spécialement conçus à cet effet. Considérons une compagnie D'assurance qui soumet les dossiers sinistre de ses assurés à un tel outil. S'il y a effectivement fraude, le programme la détecte dans 99% des cas. Cependant, l'expérience montre que l'outil informatique range parmi les fraudes 2% des dossiers en règle. Des extrapolations au niveau du marché donnent à penser que 1% des dossiers sinistre donnent effectivement lieu à une fraude. Quelle est la probabilité qu'un dossier rangé parmi les fraudes par le programme informatique ait effectivement donne lieu à cette pratique condamnable?

Afin d'évaluer cette probabilité, définissons les évènements

E = « l'outil informatique décèle une fraude »

F_1 = « l y a fraude »= $\bar{F}_2$

Dans ce cas : $Pr[F_1] = 0.01, Pr[E|F_1] = 0.99, Pr[E|F_2] = 0.02$

La probabilité cherchée vaut

$$Pr[F_1|E] = \frac{Pr[E|F_1]Pr[F_1]}{Pr[E|F_1]Pr[F_1] + Pr[E|F_2]Pr[F_2]}$$

$$= \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.02 \times 0.99} = \frac{1}{3}$$

2. Variables aléatoires

2.1 Définition

Formellement, une variable aléatoire est définie par rapport à un espace probabilisé (Ɛ,A, Pr) même si nous verrons que l'actuaire peut le plus souvent faire l'économie de la définition précise de cet espace.
Une variable aléatoire X est une fonction définie sur l'ensemble Ɛ des évènements élémentaires et a valeurs dans R:

$$X: \varepsilon \to R; e \to x = X(e)$$

À laquelle on impose une condition technique (appelée condition de mesurabilité): quel que soit le réel x, on veut que

$$\{X \le x\} \equiv \{e \in \varepsilon | X(e) \le x\} \in A$$

i.e. $\{X \le x\}$ doit être un évènement.
La condition de mesurabilité garantit qu'il soit possible de souscrire une police prévoyant le versement d'1DA si l'évènement « X est inférieure ou égale à x" se produit.
Les exemples sont multiples, du plus élémentaire où X décrit la face montrée par une pièce de monnaie une fois retombée (X vaut alors 0 si on obtient pile et 1 si on obtient face), au cas plus compliqué où X est le nombre de blessés suite à un tremblement de terre dans une région ou le temps séparant deux tempêtes frappant une ville. Par extension, nous permettrons également à X d'être constante.

Exemple

Considérons une assurance couvrant le vol ou la perte des bagages lors d'un déplacement déterminé, l'assureur verse un forfait de 25000 DA si les bagages viennent à être perdus ou volés au cours du voyage. Les seules éventualités à prévoir sont: soit le vol ou la perte des bagages lors du déplacement couvert (auquel cas l'assureur verse le forfait de 25000 DA), soit l'absence d'un tel évènement (auquel cas l'assureur ne doit effectuer aucune prestation).
Les deux évènements élémentaires suivants peuvent se produire:
e_1 = "les bagages ne sont ni volés ni perdus lors du voyage"
e_2 = "les bagages sont perdus ou volés lors du voyage."
Ainsi, $\varepsilon = \{e_1, e_2\}$ et $\mathcal{A} = \{\emptyset, \{e_1\}, \{e_2\}, \varepsilon\}$
Notez que la définition de ε est aussi concise que possible. Ainsi, on ne s'intéressera pas aux circonstances précises entourant la perte ou le vol des bagages, puisqu'elles n'influencent pas la prestation de l'assureur.
La charge financière X de l'assureur, inconnue à la conclusion de la police, vaut donc

$$X_{(e)} = \begin{cases} 0\,DA, si\ e = e_{1,} \\ 25000\,DA, si\ e = e_2 \end{cases}$$

Cette fonction est bien une variable aléatoire puisque

$$\left\{X \leq x = \begin{cases} \emptyset, si\ x < 0, \\ e_1, si\ 0 \leq x < 25000DA, \\ \varepsilon, si\ x \geq 25000\ DA. \end{cases}\right.$$

fait bien partie de $\mathcal{A}$ quelle que soit la valeur de x.

Souvent, X(e) représente la dépense totale de l'assureur relativement à une police du portefeuille au cours d'une période déterminée lorsque e décrit la réalité observée. Il s'agit de la variable d'intérêt pour l'assureur quel que soit le type de prestations promises par lui. En effet, même si la prestation revêt différentes formes du point de vue de l'assuré, elle représente toujours un coût financier pour l'assureur. Ceci est d'autant plus vrai que les prestations en nature sont le plus souvent confiées à une société tierce, afin de mieux en maitriser les coûts. Ainsi, le volet assistance compris dans la plupart des polices automobiles est sous-traité à un spécialiste de ce genre de couverture.

2.2 Fonction de répartition

On sent bien que pour beaucoup de polices l'énumération des éléments de ε s'avérera fastidieuse et difficile: ceci est dû au grand nombre de situations pouvant conduire à un sinistre (il s'agit d'un accident qui donne lieu à une demande de remboursement) et à l'incertitude entourant les conséquences financières de celui-ci. En pratique, la définition de ε n'est pas nécessaire aux calculs actuariels et il suffit de connaitre la fonction de répartition de X pour être en mesure d'évaluer les quantités dont l'actuaire a besoin pour gérer le portefeuille.

Supposons que la variable aléatoire X représente le montant des indemnités que la compagnie devra verser. Afin de gérer ce risque, l'actuaire dispose d'informations à propos des sinistres antérieurs causés par des polices du même type. Ceci lui permet d'obtenir une fonction $F_X: R \rightarrow [0,1]$, appelée fonction de répartition, donnant pour chaque seuil x ∈ IR, la probabilité (i.e. la "chance") que X soit inférieure ou égale à x.

Définition. La fonction de répartition F_X associée à la variable aléatoire X est définie comme

$$F_X(x) = Pr\left[\left\{e \in \varepsilon \middle| X_{(e)} \leq x\right\}\right]$$
$$\equiv Pr[X \leq x], x \in R.$$

Notez que la condition de mesurabilité revêt ici toute son importance, en garantissant que l'évènement "X est inférieure ou égale à x" est bien un évènement dont on peut mesurer la probabilité. On peut voir $F_X(x)$ comme la prime à payer pour recevoir l'indemnité si l'évènement $\{X \leq x\}$ se réalise.

Exemple : La fonction de répartition de la variable aléatoire X définie à l'Exemple précédent vaut

$$F_X(x) = \begin{cases} \Pr[\emptyset] = 0, si\ x < 0, \\ \Pr[e_1]\,, si\ 0 \leq x < 25000DA, \\ \Pr[\varepsilon] = 1\,, si\ x \geq 25000\ DA. \end{cases}$$

Supposons que X représente le coût d'un sinistre. La connaissance de F_X ne signifie nullement que l'actuaire dispose du montant du sinistre. La donnée de

F_X fournit pourtant à l'actuaire une connaissance complète du comportement stochastique de la variable aléatoire X. En effet, quel que soit le niveau x, il sait combien il faut payer pour souscrire une police prévoyant le paiement d'1DA lorsque l'évènement $\{X \leq x\}$ se réalise (i.e. lorsque le montant du sinistre est inférieur ou égal au seuil x). Grâce aux polices élémentaires couvrant les évènements$\{X \leq x\}$, l'actuaire est en mesure de construire et de tarifer n'importe quel produit plus complexe.
Il est possible d'effectuer tous les calculs actuariels à l'aide de F_X, sans définir explicitement l'ensemble ε des évènements élémentaires et $\mathcal{A}$ des évènements conditionnant le risque. Ceci facilite grandement les développements actuariels.

2.3 Support d'une variable aléatoire

L'ensemble de toutes les valeurs possibles pour une variable aléatoire est appelé son support. Cette notion est précisément définie comme suit.
Définition. Le support d'une variable aléatoire X de fonction de répartition F_X est l'ensemble de tous les points $x \in R$ Où F_X est strictement croissante, i.e.
support(X) = $\{x \in R | F_X(x) > F_X(x - \epsilon) \; pour \; tout \; \epsilon > 0\}$

2.4 Fonction de queue

En plus de la fonction de répartition, nous utiliserons souvent son complémentaire, appelé fonction de queue en actuariat non-vie (en biostatistique et en assurance vie, cette même fonction est appelée fonction de survie lorsque X représente la durée de vie d'un individu). Notée $\bar{F}_X$, celle-ci est définie comme suit.
Définition Etant donnée une variable aléatoire X, la fonction de queue associée vaut

$$\bar{F}_X(x) = 1 - F_X(x) = Pr[X > x], x \in R.$$

$\bar{F}_X(x)$ représente donc la probabilité que X prenne une valeur supérieure à x.
On peut voir $F_X(x)$ comme la prime à payer pour recevoir une somme de 1DA si X excède x. Clairement, F_X est décroissante puisque l'évènement X > x est plus probable que X > x' lorsque x < x', i.e. $\{X > x'\} \subseteq \{X > x\}$.
Lorsque le support de la fonction de répartition F du montant de sinistre est R^+, on mesure généralement le risque associé à telle ou telle fonction de répartition par l'épaisseur des queues de distribution (c'est-_à-dire par la masse de probabilité repartie sur les régions (c,+∞), pour de grandes valeurs de c). Ainsi, on parlera de queue épaisse (ou lourde) lorsque $\bar{F}_x$ ne tend que lentement vers 0 lorsque $x \to +\infty$.

2.5 Égalité en loi

Souvent les actuaires sont davantage intéressés par la fonction de répartition d'une variable aléatoire que par cette variable aléatoire elle-même. Il est essentiel ici de bien saisir la nuance entre ces deux êtres mathématiques. La variable aléatoire S peut par exemple représenter le montant total de sinistre généré par monsieur Untel durant l'année qui vient. D'autre part, supposons que T représente le montant de sinistre généré par monsieur Unautre. Si ces deux

individus sont indiscernables pour l'assureur, cela signifié que $F_S(t) = F_T(t)$ quel que soit $t \in IR$. La compagnie leur appliquera alors le même tarif. Bien évidemment, il n'y aucune raison pour que les réalisations de S et T obtenues en fin de période soient identiques. Ainsi, le fait pour S et T de posséder la même fonction de répartition ne signifie pas que S = T.
Pour l'actuaire qui ne connaît ni monsieur Untel, ni monsieur Unautre, il n'y a d'intérêt que dans la fonction de répartition commune de S et T.

3. Variables aléatoires de comptage

Comme leur nom l'indique, ces variables comptent un nombre d'évènements, par exemple le nombre de sinistres survenus au cours d'une période. Dorénavant, nous noterons les variables de comptage par des lettres capitales du milieu de l'alphabet: I, J, K, L, N, M, . . . Leur support est donc (contenu dans) IN = $\{0,1,2,\ldots\}$.
Une variable aléatoire de comptage N est caractérisée par la suite des probabilités $\{pk, k \in N\}$ associée aux différents entiers, i.e. p_k = Pr[N = k]. La fonction de répartition de N est "en escalier", les sauts d'amplitude p_k se produisant aux entiers k, i.e.

$$F_N(x) = Pr[N < x] = \sum_{K=0}^{\lfloor x \rfloor} p_k \,, x \in R,$$

Où $\lfloor x \rfloor$ désigne la partie entière du réel x. Clairement, $F_N(x) = 0$ si x < 0.
Remarque. Plus généralement, on parle de variable discrète lorsque celle-ci prend ses valeurs dans un ensemble $\{a_1, a_2, a_3, \ldots\}$, qui est fini ou peut être mis en bijection avec IN. Les résultats à propos des variables de comptage s'étendent facilement au cas discret en substituant a_k à l'entier k.
Parmi les variables de comptage, nous utiliserons surtout celles dont les lois de probabilité sont présentées dans les paragraphes suivants.

3.1 Variable uniforme discrète

Une variable aléatoire N est dite de loi uniforme discrète sur$\{0, 1, \ldots, n\}$, ce qui se notera désormais N ~ DUni(n), lorsque
Pr [N = k] =$\frac{1}{n+1}$, pour k = $\{0, 1, \ldots, n\}$
Les masses de probabilité accordées aux entiers$0, 1, \ldots, n$ sont donc toutes identiques, d'où le qualificatif uniforme: aucune valeur du support $\{0, 1, \ldots, n\}$ n'est plus probable que les autres. Clairement, la fonction de répartition associée à cette variable aléatoire est donnée par

$$F_N(x) = \begin{cases} 0, si\ x < 0 \\ \dfrac{\lfloor x \rfloor + 1}{n+1}, si\ 0 \le x < n \\ 1, si\ x \ge n \end{cases}$$

Il s'agit donc d'une fonction dont le graphe effectue des sauts d'amplitude $\frac{1}{n+1}$ à chaque entier $0, 1, \ldots, n$.

3.2 Variables de Bernoulli

Une variable aléatoire N est dite de Bernoulli de paramètre 0 < q < 1, ou dite obéir à cette loi, ce qui se notera N ~ Ber(q), lorsque
Pr[N = 1] = q = 1 - Pr[N = 0]
Une variable de Bernoulli indique si l'_évènement d'intérêt s'est réalisé lors d'une répétition de l'expérience dans un schéma de Bernoulli (un tel schéma

suppose qu'une expérience aléatoire soit réalisée et qu'on s'intéresse à la réalisation d'un _évènement donné à cette occasion).
En sciences actuarielles, on associe classiquement la loi de Bernoulli à l'indicatrice d'un évènement aléatoire. L'indicatrice vaut 1 lorsque l'évènement est réalisé (et indique donc la réalisation de cet évènement). De tels évènements sont « la police a produit au moins un sinistre au cours de la période de référence » ou « l'assuré est décédé dans l'année", par exemple.

3.3 Variable binomiale

Une variable aléatoire N est dite Binomiale d'exposant m et de paramètre q, m ∈ IN et 0 < q < 1, ou dite obéir à cette loi, ce qui se notera N ∼Bin(m,q), lorsque N prend ses valeurs dans $\{0, 1, \ldots, m\}$ et

$$Pr[N = K] = C_m^k q^k (1-q)^{m-k}, k = 0, 1, \ldots, m$$

où la notation C_m^k désigné le coefficient binomial de Newton défini par

$$C_m^k = \frac{m!}{k!\,(m-k)!}$$

Notez que lorsque m = 1, on a pour $k \in \{0,1\}$

$$Pr[N = k] = q^k (1-q)^{1-k} = \begin{cases} q, & si\ k = 1 \\ 1-q, & si\ k = 0 \end{cases}$$

et on retrouve la loi de Bernoulli, qui apparaît ainsi comme un cas particulier de la Binomiale, i.e. Bin(1,q) = Ber(q).
La loi binomiale est classiquement associée au nombre de succès dans un schéma de Bernoulli: si une expérience aléatoire est répétée m fois indépendamment et sous des conditions identiques, le nombre de répétitions au cours desquelles un certain évènement E se réalise est de loi Bin(m, Pr[E]). En effet, en posant q = Pr[E].
La loi binomiale se prête bien à la modélisation du nombre de sinistres touchant le portefeuille dans les formes d'assurance pour lesquelles au plus un sinistre par période est possible pour chaque police (comme l'assurance sur la vie, l'assurance annulation ou l'assurance rapatriement conclue pour un voyage spécifie, par exemple). Si q est la probabilité qu'une des m polices du portefeuille donne lieu à un sinistre, et si ces m polices sont identiques et ne s'influencent pas mutuellement, le nombre N de polices sinistrées est de loi Bin(m,q).

3.4 Variable géométrique

La variable de comptage N est dite géométrique de paramètre q, 0 < q < 1, ou dite obéir à cette loi, ce qui se notera N ∼ Geo(q) lorsque

$$Pr[N = k] = q(1-q)^k, k \in IN$$

Si on considère un schéma binomial et qu'on parle de succès lorsqu'un évènement E se réalise lors d'une des répétitions de l'expérience aléatoire (et d'échec dans le cas contraire), la loi Geo(q) peut se voir comme celle du nombre d'échecs précédant le premier succès.
Lorsque N = k il aura donc fallu k + 1 répétitions de l'expérience pour obtenir un premier succès.
Exemple. Considérons un inspecteur de la compagnie qui contrôle des dossiers sinistres en vue de déceler d'_éventuelles fraudes. Si q désigne la proportion de dossiers où il y a eu fraude, la probabilité qu'il en contrôle k avant de tomber sur un premier dossier ayant donné lieu à une fraude est $q(1-q)^k$.

3.5 Variable de Poisson

Aussi appelée loi des évènements rares, la loi de Poisson fut dès l'origine associée à des comptages d'accidents ou de pannes.

La loi de Poisson fut introduite comme une approximation de la loi binomiale lorsque m était très grand et q très petit. En effet, considérons une variable aléatoire$N_m \sim Bin\left(m, \frac{\lambda}{m}\right)$. Nous avons

$$Pr[N_m = 0] = \left(1 - \frac{\lambda}{m}\right)^m \to e^{-\lambda}\, si\ m \to +\infty.$$

De plus

$$\frac{Pr[N_m = k+1]}{Pr[N_m = k]} = \frac{\frac{m-k}{k+1}\frac{\lambda}{m}}{1 - \frac{\lambda}{m}} \to \frac{\lambda}{k+1} si\ m \to +\infty$$

de sorte que

$$\lim_{n\to\infty} Pr[N_m = k] = e^{-\lambda}\frac{\lambda^k}{k!}$$

La probabilité apparaissant dans le membre de droite de cette dernière équation est celle définissant la loi de Poisson. Plus précisément, lorsque

$$Pr[N = k] = e^{-\lambda}\frac{\lambda^k}{k!}, k \in IN$$

La variable de comptage N est dite de loi de Poisson de paramètre λ, ce qui se notera N ~ Poi(λ). La loi de Poisson peut donc se voir comme celle du nombre de succès dans un schéma de Bernoulli, lorsque le nombre de répétitions est très grand (m> 50) et la probabilité de succès négligeable (q≤ 0.05).

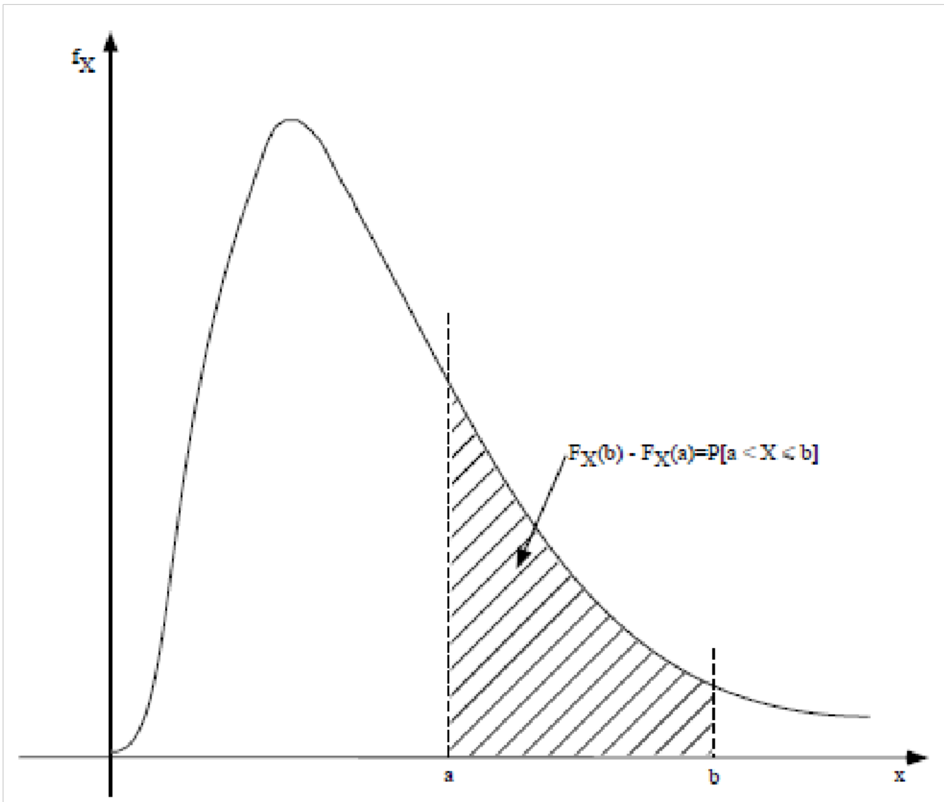
3.6 Variables aléatoires continues.

Une variable aléatoire X est dite continue lorsque sa fonction de répartition F_X admet la représentation

$$F_X(x) = \int_{y \le x} fX(y)dy, x \in R$$

pour une fonction intégrable f_X : IR → IR^+ appelée la densité de probabilité de X. Une variable aléatoire continue admet donc une fonction de répartition F_X continue.
La fonction f_X intervenant dans la représentation suivante possède une signification concrète: si nous traçons le graphe de f_X, l'aire de la surface délimitée dans le plan par le graphe de f_X, l'axe des abscisses et les verticales en a and b (a < b) est la probabilité que X prenne une valeur dans l'intervalle (a;b], i.e.

$$Pr[a < X \le b] = F_X(b) - F_X(a) = \int_{x=a}^{b} f_X(x)dx.$$



La Figure illustre la signification de la densité de probabilité. En faisant tendre la longueur b-a de l'intervalle vers 0, on voit clairement que

$$\lim_{b \to a} Pr[a < x \leq b] = Pr[x = a] = 0$$

quel que soit le réel a; une variable aléatoire continue ne possède donc pas de points massiques (qui sont l'apanage des variables discrètes).