

# LA LOI DES PETITS NOMBRES

*Arthur Charpentier*

*Professeur d'actuariat, Université du Québec, Montréal*

*Comme nous l'avons observé dans un article précédent [Charpentier, 2011], la « loi des grands nombres » est souvent évoquée pour justifier la mutualisation des risques indépendants : plus la mutualité sera grande, plus petite sera la variabilité par tête. À condition que les risques ne soient pas trop grands. Les risques catastrophiques constituent un cas problématique : ils sont rares, et parfois tellement (potentiellement) coûteux que l'hypothèse d'existence de la variance doit être remise en cause. La modélisation de ces événements rares repose sur la « loi des petits nombres », pour reprendre le titre de l'ouvrage de Ladislaus Bortkiewicz. Nous allons voir que la loi de Poisson est à la loi des petits nombres ce que la loi normale est à la loi des grands nombres. Nous étudierons donc en détail l'importance de la loi de Poisson pour modéliser les événements « rares ». Et nous verrons pourquoi la probabilité qu'un événement ne survienne pas est toujours de 37 %. Ou presque.*

## La loi de Poisson

**S**iméon Denis Poisson a travaillé sur le calcul des probabilités pendant vingt ans, de 1820 à 1840. Comme beaucoup de ses contemporains, il a effectué ses premiers travaux sur des problèmes de jeux, avec son *Mémoire sur l'avantage du banquier au jeu de trente et quarante*, lu à l'Académie des sciences le 13 mars 1820. Mais son premier travail important a porté sur un problème classique de statistique, déjà longuement étudié par Laplace (qui a été l'un de ses professeurs) : il s'agit de son *Mémoire sur la proportion des naissances des filles et des garçons* de 1829. Poisson y présente en particulier une démonstration de la loi des grands nombres pour des variables suivant une loi de Bernoulli, qu'il modifiera par la suite. Dans ce

mémoire, il énonce une loi, qui sera appelée plus tard « loi de Poisson ». Pour reprendre la terminologie de son fameux traité, paru en 1837, *Recherches sur la probabilité des jugements en matière criminelle et en matière civile*, « la probabilité qu'un événement dont la chance à chaque épreuve est la fraction très petite  $x/\mu$  n'arrivera pas plus de  $n$  fois dans un très grand nombre  $\mu$  d'épreuves » (ce qui correspond à la fonction de répartition avec une terminologie plus contemporaine) est :

$$P = \left( 1 + x + \frac{x^2}{1.2} + \frac{x^3}{1.2.3} + \dots + \frac{x^n}{1.2 \dots n} \right) e^{-x}$$

En particulier, dans le chapitre VIII, il obtient sa loi comme limite de la loi binomiale  $B(T, \lambda/T)$  lorsque  $T$  devient grand, mais il passe assez rapidement

à d'autres considérations. Il n'étudie pas cette loi limite et ne propose pas vraiment de l'utiliser. Il faudra attendre les travaux de Ladislaus Bortkiewicz, presque un demi-siècle plus tard, pour voir des applications, dans son ouvrage *Das Gesetz der kleinen Zahlen* (« la loi des petits nombres »). Et, comme toujours en mathématiques, attribuer tout le crédit à Poisson est un peu excessif, puisque, en 1718, un siècle auparavant, de Moivre avait obtenu la même loi, également comme limite de la loi binomiale. Si le nom de Poisson est passé à la postérité, c'est essentiellement parce que le citent Boltzmann en 1868, Seidel en 1876, et surtout Tchebychev. Cela dit, Poisson était loin d'être un inconnu, en tant que scientifique. Il a travaillé sur des problèmes d'électrostatique (la fameuse équation de Poisson) et sur l'équation de la chaleur avec Fourier, il s'est opposé à Fresnel sur des problèmes d'optique et il a présidé à deux reprises l'Académie des sciences, à Paris.

## La loi des petits nombres

La distribution d'un nombre d'événements obtenu comme somme de variables de Bernoulli, dans une grande population, peut s'approcher par la loi de Poisson (ou lorsque la probabilité de survenance est faible, relativement à la taille de l'échantillon). Formellement, si  $N$  suit

une loi binomiale  $B(n,p)$ , avec  $p \approx \lambda / n$  alors :

$$P[N = k] = \binom{n}{k} p^k (1-p)^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!}$$

On parle de « loi des petits nombres » car on compte ici des événements rares, la probabilité de survenance d'un événement étant inversement proportionnelle à  $n$ .

Ce résultat se traduit de la manière suivante (voir figure 1) : si, sur un damier 10 x 10, on lance 100 pièces et que l'on compte le nombre de pièces par case, la distribution suivra une loi de Poisson de moyenne 1 (le ratio entre le nombre de pièces et le nombre de cases). Notons dès à présent qu'environ 37 % des cases sont vides.

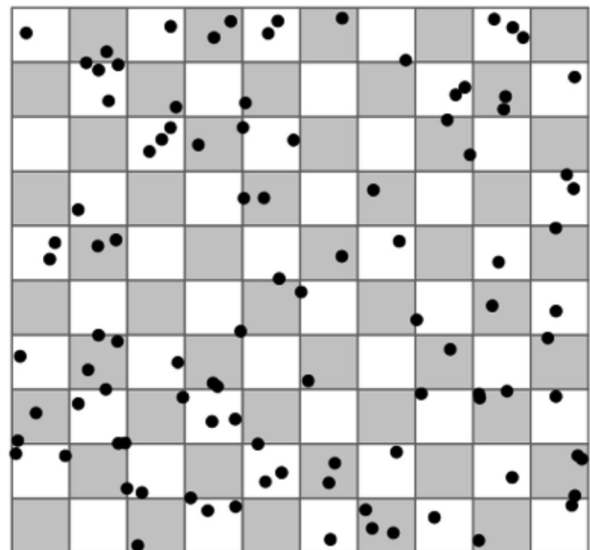
## L'utilisation de la loi de Poisson

À la fin du XIX<sup>e</sup> siècle, à l'université de Göttingen, Wilhelm Lexis (connu en démographie pour les diagrammes qui portent aujourd'hui son nom) souhaitait étudier les applications de la loi de Poisson, notamment dans un contexte démographique. Un de ses étudiants,

Figure 1 - Nombre de pièces par case, après le lancement de 100 pièces sur un damier 10 x 10

Nombre de pièces par case	Fréquence	Loi de Poisson
0	36	36,78
1	39	36,78
2	16	18,39
3	7	6,13
4	2	1,53
5 et plus	0	0,37

Source : Arthur Charpentier.



Ladislaus Bortkiewicz, proposa de l'utiliser pour modéliser des nombres d'accidents. Dans son exemple fameux, en 1898, ce dernier a étudié le nombre de cavaliers morts par ruade de cheval, entre 1875 et 1894, dans 10 corps d'armée (si la référence est un corps, par an, on a alors 200 observations). Il a obtenu la distribution donnée dans la table de la figure 2. La colonne de droite donne le nombre de décès donné par une loi de Poisson de même moyenne. On note que l'ajustement est très bon, et on comprend vite la fascination de Bortkiewicz pour cette loi.

Maintenant, pour être tout à fait honnête, la loi que nous avons énoncée comme une loi des petits nombres est très différente de celle proposée par

Bortkiewicz dans son ouvrage, qui lui a valu un cinglant commentaire de Corrado Gini, qui écrivait en 1907, de manière provocatrice, que la loi des grands nombres n'existe pas. Mais l'utilisation de la loi de Poisson pour modéliser les accidents, et plus généralement les *petits* nombres, venait de débiter.

Dans les années 1930, Filip Lundberg avait noté l'intérêt (théorique) du processus de Poisson pour modéliser l'arrivée des sinistres, à l'instar de plusieurs actuaires de l'école scandinave (Esscher en 1932, Carl-Otto Segerdahl en 1939, etc.). Depuis, tous les actuaires ont utilisé cette loi pour modéliser toutes sortes d'événements « rares », comme la survenance annuelle d'ouragans aux États-Unis (figure 3).

Figure 2 - Nombre de décès par ruade de cheval

Nombre de décès par corps d'armée	Fréquence	Loi de Poisson
0	109	108,67
1	65	66,21
2	22	20,22
3	3	4,11
4	1	0,63
5 et plus	0	0,08

Source : Bortkiewicz [1898].

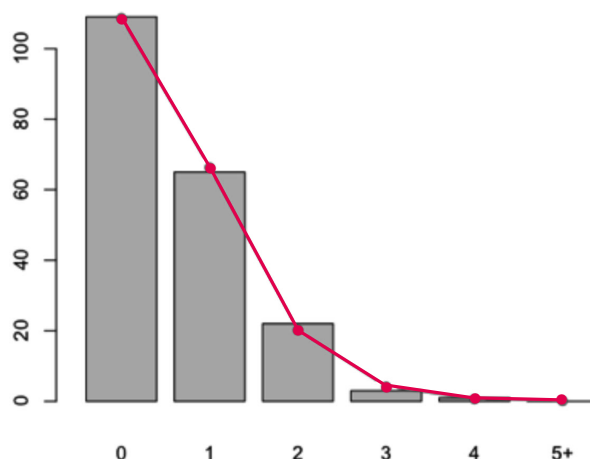
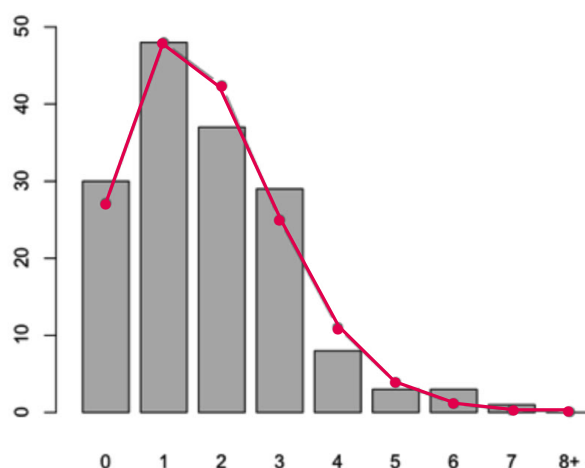


Figure 3 – Nombre annuel d'ouragans ayant touché la côte est des États-Unis

Nombre d'ouragans	Fréquence	Loi de Poisson
0	30	27,16
1	48	47,99
2	37	42,41
3	29	24,98
4	8	11,03
5	3	3,90
6	3	1,15
7	1	0,29
8 et plus	0	0,08

Source : Levi & Partrat [1989].



## De la loi de Poisson au mélange de lois de Poisson

**M**ais, comme souvent, la pratique de la modélisation est plus complexe. Les actuaires savent bien que la loi de Poisson ne convient pas pour modéliser la fréquence d'accidents de la route, par exemple. En témoigne la distribution obtenue sur la figure 4, à partir de données (anciennes) de La Royale Belge. La loi de Poisson sous-estime le nombre d'assurés non accidentés et, en même temps, le nombre d'assurés souvent accidentés.

Plusieurs solutions sont envisagées en pratique. Le premier modèle (A) consiste à supposer qu'une partie des sinistres ne sont pas déclarés (ce qui peut arriver dans le cas de systèmes de type malus), pour une proportion  $\pi$ , et, pour la proportion restante, une loi de Poisson conviendra.

$$P[N = k] = \begin{cases} \pi + (1 - \pi) \cdot e^{-\lambda} & \text{si } k = 0 \\ (1 - \pi) \cdot e^{-\lambda} \frac{\lambda^k}{k!} & \text{si } k \geq 1 \end{cases}$$

Le second modèle (B) consiste à supposer que la population est hétérogène, avec des classes de risques différentes. Pour une proportion  $\pi_A$ , on a une fréquence de sinistres de moyenne  $\lambda_A$  ; pour une proportion  $\pi_B$ , on a une fréquence de sinistres de moyenne  $\lambda_B$ , etc.

$$P[N = k] = \pi_A \cdot e^{-\lambda_A} \frac{\lambda_A^k}{k!} + \pi_B \cdot e^{-\lambda_B} \frac{\lambda_B^k}{k!} + \dots$$

Ce genre de modèle permet justement de corriger la sous-estimation du nombre d'assurés non sinistrés et celle du nombre d'assurés souvent sinistrés.

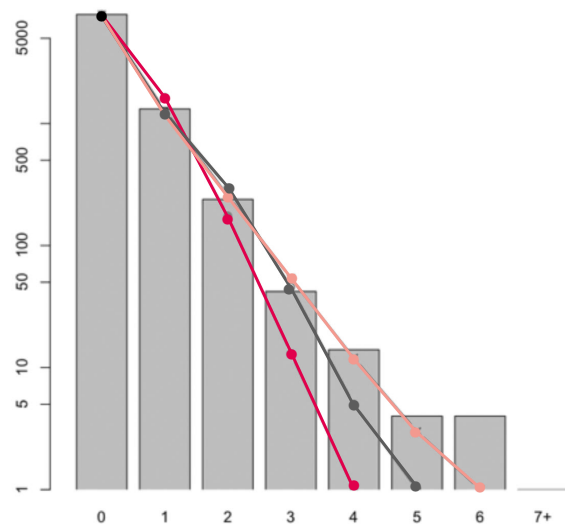
## De la loi de Poisson à la période de retour

**U**n concept fondamental en gestion des risques extrêmes a été introduit par Emil Gumbel en 1958, liant le temps qui s'écoule (en années) entre deux événements consécutifs et la probabilité (annuelle) de survenance. Pour des événements se produisant avec une probabilité annuelle de  $1/T$ , indépendamment

Figure 4 – Nombre annuel d'accidents de la route, pour des assurés belges

Nombre d'accidents	Fréquence	Loi de Poisson	(A)	(B)
0	7 840	7 636	7 840	7 847
1	1 317	1 637	1 272	1 288
2	239	175	297	257
3	42	13	46	54
4	14	1	5	12
5	4	0	1	3
6	4	0	0	1
7 et plus	1	0	0	0

Source : Simar [1976].



les uns des autres, le temps moyen d'attente entre deux événements est  $T$ , appelé « période de retour », et la probabilité qu'aucun événement ne survienne pendant  $n$  années (consécutives) est alors :

$$P[N > n] = \left(1 - \frac{1}{T}\right)^n$$

On peut illustrer ceci dans la figure 5, qui montre la probabilité qu'un événement ne survienne pas pendant  $n$  années (en ligne) en fonction de la période de retour  $T$ . Pour un événement centennal, il y a 36,60 % de chance pour qu'il ne survienne pas en cent ans. On retrouve une nouvelle fois nos 37 %.

Figure 5 - Période de retour et probabilité de ne pas avoir de catastrophe

Nombre d'années sans catastrophe	Période de retour (en années)				
	10	20	50	100	200
10	<b>34,86 %</b>	59,87 %	90,43 %	90,43 %	95,11 %
20	12,15 %	<b>35,84 %</b>	66,76 %	81,79 %	90,46 %
50	0,51 %	7,69 %	<b>36,41 %</b>	60,50 %	77,83 %
100	0,00 %	0,59 %	13,26 %	<b>36,60 %</b>	60,57 %
200	0,00 %	0,00 %	1,75 %	13,39 %	<b>36,69 %</b>

Source : Arthur Charpentier.

## La probabilité qu'un événement ne survienne pas est de... 37 %

Comme on le voit dans la figure 5, en suivant la diagonale, la probabilité qu'un événement ne survienne pas pendant  $T$  années quand sa période de retour est  $T$  est de l'ordre de 37 %. Si l'on revient un instant à notre damier 10 x 10 évoqué auparavant, la probabilité de n'avoir aucune pièce sur une case était de... 37 %. On retrouve ici l'exponentielle de -1, tout simplement.

Reprenons ici un exemple qui avait fait polémique, il y a quelques années, à propos du risque nucléaire.

Dans un article [Laponche et Dessus, 2011], on apprenait que la probabilité d'avoir un incident majeur sur réacteur nucléaire, pour une année, était de l'ordre de 0,0003 (3 chances sur 10 000) avec 5 incidents pour 14 000 années-réacteur (31 années pour 450 réacteurs). Or, comme il y a 143 réacteurs nucléaires dans l'Union européenne, sur 30 ans, la probabilité d'avoir un incident majeur est (selon le calcul des auteurs) :

$$\underbrace{143 \times 30}_n \times \underbrace{0,0003}_p = 129 \%$$

D'où la conclusion savoureuse : « la probabilité d'occurrence d'un accident majeur sur ces parcs serait donc [...] de plus de 100 % pour l'Union européenne ». Comme on l'a noté, la probabilité d'avoir un incident majeur sur une période  $n$  lorsque la probabilité annuelle est  $p$  s'écrit

$$P[N \leq n] = 1 - (1 - p)^n \approx np$$

en utilisant un développement limité non justifié ici ! Il convient d'utiliser le modèle de Poisson. La probabilité d'avoir au moins un incident majeur est

$$P[N \leq n] = 1 - (1 - p)^n \approx 1 - e^{-np}$$

soit ici 72,47 %. Si la probabilité était traduite en termes de durée,

$$p = \frac{1}{T} \text{ avec } T = \frac{1}{0,0003 \times 143} \approx 23,31$$

On voit que le temps moyen d'attente entre deux incidents majeurs en Europe est de 23 ans. Ce qui fait que, sur 23 ans, la probabilité de n'avoir aucun incident majeur est de l'ordre de 37 %.

## Conclusion

La loi de Poisson est présente partout en assurance, car c'est la loi centrale pour modéliser le comptage d'événements rares. Le nombre de décès dans un

portefeuille d'assurance vie suit une loi de Poisson, tout comme le nombre d'accidents par unité de temps en assurance automobile. On retrouve même l'estimateur *Chain Ladder* du montant de provisions pour sinistres à payer quand on utilise une régression de Poisson sur les incréments de paiements.

### Bibliographie

BORTKIEWICZ L., *Das Gesetz der kleinen Zahlen*, Teubner, 1898. Disponible en PDF : <https://archive.org/details/dasgesetzderklei00bortrich>

CHARPENTIER A., « La loi des grands nombres et le théorème central limite comme base de l'assurabilité ? », *Risques*, n° 86, juin 2011, pp. 111-116.

GUMBEL E. J., *Statistics of Extremes*, Columbia University Press, 1958.

LAPONCHE B. ; DESSUS B., « Accident nucléaire : une certitude statistique », *Libération*, 3 juin 2011. Disponible en ligne : [http://www.liberation.fr/politiques/2011/06/03/accident-nucleaire-une-certitude-statistique\\_740208](http://www.liberation.fr/politiques/2011/06/03/accident-nucleaire-une-certitude-statistique_740208)

LEVI C. ; PARTRAT C., “Statistical Analysis of Natural Events in the United States”, *Astin Bulletin*, journal de l'Association actuarielle internationale, vol. 21 (2), Ceuterick, novembre 1991, pp. 253-276. Disponible en PDF : <https://www.casact.org/library/astin/vol21n2/253.pdf>

SIMAR L., “Maximum Likelihood Estimation of a Compound Poisson Process”, *The Annals of Statistics*, vol. 4 (6), novembre 1976, pp. 1200-1209. Disponible en PDF : [http://projecteuclid.org/download/pdf\\_1/euclid.aos/1176343651](http://projecteuclid.org/download/pdf_1/euclid.aos/1176343651).