REVUE DE STATISTIQUE APPLIQUÉE

G. PILÉ

Processus de poisson et apparentés

Revue de statistique appliquée, tome 9, nº 4 (1961), p. 71-83
<http://www.numdam.org/item?id=RSA_1961__9_4_71_0>

PROCESSUS DE POISSON ET APPARENTÉS

G. PILÉ

Ancien Élève de l'École Polytechnique

On sait tout l'intérêt que présentent les processus de Poisson dans les applications les plus diverses, notamment dans l'étude des processus stochastiques auxquels s'intéresse la recherche opérationnelle.

Il n'est pas rare que des utilisateurs apparaissent insuffisamment informés à leur sujet ou qu'ils en aient une compréhension intuitive sujette à caution.

L'auteur, qui ne prétend d'ailleurs aucunement faire oeuvre originale, a rappelé sous le présent titre un certain nombre de notions de base gravitant autour de ce qu'il est convenu d'appeler les "processus de Poisson", lesquels s'inscrivent, comme on le sait, parmi des processus plus généraux dits "additifs".

Les considérations développées dans cet exposé peuvent aider à une meilleure vulgarisation de notions aussi importantes.

I - GENERALITES -

Rappelons comment, de façon générale, un processus peut tendre vers une loi de Poisson.

Considérons une suite indéfinie de V.A. (variables aléatoires) : ΔX_1, ΔX_2, ..., ΔX_n supposées mutuellement indépendantes quel que soit n et ne pouvant prendre que les valeurs 0 ou 1.

Les probabilités relatives à la valeur 1 étant notées p_i, on suppose que les p_i sont majorées par un infiniment petit de l'ordre de $\frac{1}{n}$ et que leur somme a pour limite, lorsque $n \longrightarrow \infty$, un nombre positif m.

Etudions la V.A. à valeurs entières positives ou nulles $Z_n = \sum_0^n \Delta X_i$.

Par application du théorème de multiplication, on peut écrire :

$$P(Z_n = 0) = \prod_1^n (1 - p_i)$$

Lorsque $n \longrightarrow \infty$ le logarithme du second membre tend vers $- \sum p_i = - m$ en vertu des hypothèses faites sur les p_i et

$$P(Z_n = 0) \longrightarrow e^{-m}$$

Généralisons :

$$P(Z_n = X) = \sum \prod_{r=1}^{r=x} p_{i_r} \prod_{i \neq \{i_1 \ldots i_x\}} (1 - p_i)$$

la somme étant étendue à tous les ensembles distincts possibles $i_1 \ldots i_x$ qui sont eux-mêmes les combinaisons x à x des n premiers entiers.

D'après ce qui précède, les deuxièmes facteurs ont une limite commune e^{-m} tandis que les premiers ont une moyenne égale, à la limite, à l'espérance mathématique du produit de x tirages indépendants d'une valeur p_i.

Par application du théorème de multiplication des variables indépendantes, on peut donc écrire :

$$P(Z_n = x) \longrightarrow C_n^x \, e^{-m} \left(\frac{m}{n}\right)^x = \frac{e^{-m} m^x}{x!} \left(1 - \frac{1}{n}\right) \ldots \left(1 - \frac{x-1}{n}\right) = \frac{e^{-m} m^x}{x!} \left(1 + \frac{a_x}{n}\right)$$

Le terme correctif compris entre 0 et $- \frac{(x-1)x}{2n}$ tend vers 0 lorsque $n \longrightarrow \infty$.

On obtient en définitive l'expression générale quel que soit l'entier $x \geqslant 0$:

$$P_x = e^{-m} \frac{m^x}{x!}$$

Dans le cas particulier où tous les p_i sont égaux l'expression précédente apparaît comme la limite de la loi binomiale de paramètres : n, p lorsque n croît indéfiniment et que le produit np est égal à une constante m.

II - POINT DE VUE CONCRET -

Désignons par X(t) le nombre d'événements définis (par exemple : arrivées d'usagers d'un service) survenant au cours d'un intervalle de temps (0, t).

X(t) peut être considéré comme une V.A. engendrant, t variant, une fonction aléatoire non décroissante à valeurs entières non négatives.

L'accroissement $\Delta X = X(t + \Delta t) - X(t)$ représente le nombre d'arrivées au cours de l'intervalle $(t, t + \Delta t)$. On est conduit très fréquemment en pratique, à admettre que ΔX est indépendant du nombre et des instants des arrivées antérieures à t, et on exclut par ailleurs la possibilité d'existence d'instants précis associés à une probabilité d'arrivées $\neq 0$ (cas mixte).

Dans ces hypothèses :

X(t) est une fonction aléatoire à accroissements indépendants et discontinus;

X(t) doit être continue en probabilité, sinon il faudrait admettre que ΔX n'est pas infiniment petit en même temps que Δt.

Faisons par ailleurs l'hypothèse d'arrivées isolées. De façon plus précise, supposons que la probabilité de plusieurs arrivées simultanées au cours de Δt soit d'ordre au moins égal à Δt^2. Alors les considérations précédentes qui ont conduit à la loi de Poisson sont applicables et l'on peut écrire que si M(t) est l'expérience mathématique du nombre d'arrivées au cours de la période t, la probabilité P_x de x arrivées au cours de t est donnée par la formule :

$$\boxed{P_x = \frac{[M(t)]^x}{x!} e^{-M(t)}}$$

1/ Si l'on ne s'intéresse qu'au nombre de réalisations au cours d'une période définie, il est inutile de faire une hypothèse sur la loi de densité.

La loi de Poisson apparaît comme une loi à un paramètre : l'espérance mathématique du nombre observé.

2/ Il est en revanche nécessaire de faire une hypothèse sur la loi de densité lorsque la période d'observation joue le rôle d'un paramètre. Si cette loi ne constitue pas une simple hypothèse de travail et doit être définie en fonction de données d'observation, un problème d'estimation délicat peut se poser.

Il faut tester tel ou tel ajustement à des échantillons d'observations qu'il est souvent difficile de considérer comme statistiquement homogènes, c'est-à-dire issus d'une même population.

Nous verrons plus loin que, t étant défini, les moments de la distribution de x ont des valeurs simples, en particulier :

$$E(x) = \text{Var.}\, x = M(t)$$

(ces relations s'établissent d'ailleurs facilement ici par un calcul direct).

Cette condition nécessaire peut constituer un premier test de la validité d'ajustement d'un processus de Poisson.

La grande variabilité des variances d'échantillon ne permet malheureusement pas d'obtenir des résultats très significatifs. Des analyses plus fines ont été proposées qui ne seront pas développées ici. Nous nous contenterons d'indiquer plus loin un des moyens élémentaires utilisables pour tester l'uniformité du processus (cas $M(t) = \lambda t$).

3/ Il importe d'avoir une compréhension intuitive des diverses circonstances pouvant donner naissance à un processus de Poisson dans les cas d'application courante.

Faisons remarquer tout d'abord que les probabilités élémentaires p_i considérées dans le cas théorique peuvent être interprétées dans un processus stochastique ponctuel comme définissant :

- soit les probabilités, au cours d'intervalles élémentaires successifs, de réalisation d'un quelconque des événements susceptibles de se produire au cours de l'intervalle d'observation T ;
- soit les probabilités globales au cours de T, de réalisation d'événements "personnalisés".

Recourons à des exemples concrets justiciables de cette deuxième optique :

- supposons que chaque maison d'une cité ait une petite probabilité p_i de brûler au cours d'une période donnée. La somme $\sum_1^n p_i$ n'est autre que l'espérance mathématique du nombre de sinistres dans la cité, n désignant le nombre de ses maisons.

Les p_i peuvent être égaux ou inégaux, il n'est pas nécessaire de faire à ce sujet une hypothèse particulière pour la validité de l'assimilation à un processus de Poisson, le point essentiel étant seulement le caractère d'indépendance entre les variables ΔX_i prises deux à deux.

Par exemple si le feu se déclarant à cette maison avait une probabilité notable de s'étendre à une maison voisine, l'hypothèse précédente ne pourrait plus être retenue et un modèle plus complexe (schéma de contagion) devrait être envisagé.

Supposons qu'entre deux instants définis délimitant un intervalle, 20 minutes par exemple, 4 arrivées d'avions réguliers, soient attendues sur un aérodrome. En fait, selon les services, des écarts plus ou moins importants (heure réelle - heure prévue) sont constatés, se situant par exemple dans un range (-10 +20).

Peuvent a priori se présenter au cours de T, tous les services tels que cet intervalle "supporte" une masse de probabilité appréciable p_i. Dans le cas choisi, on aura par exemple à considérer au maximum 15 réalisations possibles i avec :

$$p_i < 1 \qquad \text{quel que soit } i$$

et

$$\sum p_i \sim 4$$

On est dans ces conditions presque assuré d'un bon ajustement de la loi de Poisson.

Elargissons l'intervalle d'observation, disons à 1/2 heure, 3 ou 4 arrivées deviennent alors des événements pouvant être considérés comme certains (correspondant à des p_i égaux à 1).

L'assimilation à une loi de Poisson s'en trouve altérée, et ce, en dépit de l'invariance par addition de cette loi démontrée plus loin.

Il est à présumer que le processus est alors plus valablement représenté par une loi de Poisson dans laquelle un changement d'origine est effectué sur la variable.

La loi de Poisson apparaît ainsi comme fondamentalement adaptée à une analyse étroitement localisée.

III - LE CAS UNIFORME -

Il correspond au cas le plus habituellement considéré en pratique :

$$M(t) = \lambda t$$

où λ est une constante positive.

3.1/ P_x ne dépend plus des bornes de l'intervalle considéré mais seulement de sa longueur. Un raisonnement simple, qu'il est utile de connaître, suffit dans ce cas pour obtenir les P_x indépendants du temps ("probabilités stationnaires"). "Déplaçons" de dt l'intervalle d'observation supposé de longueur invariable T. La probabilité d'observer une transition de l'état (x - 1) à l'état x doit être égale à la probabilité d'observer la transition inverse. On peut ainsi écrire :

$$P_{x-1}\, \lambda\, dt(1 - \mu_{x-1}\, dt) = P_x\, \mu_x\, dt(1 - \lambda dt)$$

où $\lambda\,dt$ représente la probabilité d'une "acquisition" et $\mu_x\,dt$ celle d'une "perte", le système étant à l'état x.

Or $\mu_x = \frac{x}{T}$ (du fait d'une répartition uniforme de la densité de probabilité "a posteriori").

On en déduit en égalant les parties principales des deux membres :

$$x.\,P_x = \lambda T\,P_{x-1}$$

soit :

$$P_x = \frac{(\lambda T)^x}{x!}\,P_o$$

compte tenu de : $\sum P_x = 1$:

$$P_x = e^{-\lambda T}\,\frac{(\lambda T)^x}{x!}$$

3.2/ Propriétés caractéristiques de la loi de Poisson.

Fonction génératrice - Celle-ci est définie par :

$$G(s) = \sum_o^{\infty} e^{-m}\,\frac{(m\,s)^x}{x!} = e^{m(s-1)}$$

Faisons observer que l'expression de G(s) pouvait être déduite directement des propriétés de définition de l'alinéa 1.

La F.G. de la variable X considérée s'écrit en effet

$$G(s) = \prod_1^n (q_i + p_i\,s)$$

$$\text{Log } G(s) = \sum_1^n \log\{1 - p_i\,(1 - s)\}$$

Les p_i tendant vers 0, leur somme restant finie

$$\text{Log } G(s) \longrightarrow -\,(1 - s)\sum_1^n p_i = -\,(1 - s)m$$

Cette méthode d'obtention de l'expression de la loi de Poisson est certainement la plus simple.

Fonctions caractéristiques -

$$\varphi(s) = \sum_o^{\infty} e^{isx}\,P_x = e^{m(e^{is}-1)}$$

$$\psi(s) = m(e^{is} - 1)$$

Tous les cumulants d'une distribution de Poisson sont donc égaux à m.

Il en résulte que les moments centrés d'ordre 2 et 3 sont égaux à m.

De façon générale

$$\mu_{r+1} = m\,r\,\mu_{r-1} + m\,\frac{d\mu_r}{dm}$$

c'est ainsi que

$$\mu_4 = m + 3m^2 \qquad \mu_5 = m + 10m^2$$

La forme des fonctions caractéristiques met en évidence l'invariance de la loi de Poisson par addition. En d'autres termes, la composition de deux lois de Poisson de paramètres m et m' est une loi de Poisson de paramètre m + m'.

Cette propriété fondamentale doit en pratique être utilisée avec précaution, compte tenu des remarques in fine de l'alinéa 2.

Noter aussi qu'elle s'applique seulement à la somme des variables de Poisson, mais non à une combinaison linéaire quelconque.

IV - INTERVALLES ENTRE EVENEMENTS SUCCESSIFS : LOIS "Ek" -

4.1/ Une arrivée (0) étant prise pour origine, la probabilité qu'aucune arrivée ne survienne au cours de la période T s'obtient en faisant x = 0 dans la loi de Poisson de paramètre M(T), la probabilité d'une arrivée au cours de l'intervalle élémentaire dT étant égale à m(T)dt, il s'ensuit que la loi cherchée a pour expression :

$$e^{-M(T)}\, m(T)\, dt$$

qui est en fait la limite du produit étendu à tous les intervalles élémentaires dt_i de somme t :

$$\prod \{1 - m(t_i)dt_i\}\, m(T)dt$$

Supposons que M(t) = λt (cas uniforme) et recherchons de façon générale la loi de probabilité de l'intervalle compris entre l'arrivant et l'arrivant suivant d'ordre k. Cette loi peut être définie par l'intégrale :

$$\iint \ldots \int_{t_1 + \ldots + t_k = T} e^{-\lambda t_1}\, \lambda dt_1 \ldots e^{-\lambda t_k}\, \lambda\, dT$$

qui s'écrit :

$$\lambda dT\ \lambda^{k-1}\ e^{-\lambda T} \iint \ldots \int_{t_1 + \ldots + t_{k-1} \leqslant T} dt_1 \ldots dt_{k-1} = \frac{(\lambda T)^{k-1}}{(k-1)!}\, e^{-\lambda T}\ \lambda dT$$

Cette loi est couramment désignée dans la théorie des processus d'attente sous le nom "d'exponentielle à k phases" ou loi d'Erlang d'ordre k, en abrégé : Ek. On sait que des lois de ce type sont rencontrées en statistique : Lois de Pearson type III, loi de χ^2.

Il y a lieu de remarquer qu'il n'est pas nécessaire de prendre pour origine un instant d'arrivée, tous les raisonnements précédents demeurent valables si celle-ci est arbitrairement choisie. Ceci résulte de la propriété fondamentale des processus purement additifs, exprimée de façon imagée : "absence de mémoire".

4.2/ Prenons comme intervalle unité l'intervalle moyen $\frac{1}{\lambda}$ entre deux arrivants consécutifs et calculons les moments des divers ordres de la distribution Ek.

Ceux-ci peuvent être obtenus par un calcul direct :

$$m_r = \int_0^\infty T^r \frac{T^{k-1}}{(k-1)!}\, e^{-T} dT = \frac{(r+k-1)!}{(k-1)!}$$

On en déduit en particulier :

$$m_1 = k \qquad m_2 = k(k+1) \qquad m_3 = k(k+1)(k+2)$$
$$\mu_2 = k \qquad \mu_3 = 2k$$

Si l'unité était quelconque, on aurait :

$$\mu_1 = \frac{k}{\lambda} \qquad \mu_2 = \frac{k}{\lambda^2}$$

Il est généralement utile de prendre pour unité l'intervalle moyen entre arrivées "distantes" de k, c'est-à-dire m_1.

On a dans ce cas : $\mu_2 = \frac{1}{k}$ et

$$f(t)\,dT = \frac{(kT)^{k-1}}{(k-1)!}\,e^{-kT}\,dT$$

La forme de la fonction caractéristique d'une distribution E_k est utile à connaître :

$$\varphi(t) = \frac{\lambda^k}{(k-1)!}\int_0^{\infty} e^{x(-\lambda+it)}\,x^{k-1}\,dx = \left(1 - \frac{it}{\lambda}\right)^{-k}$$

En développant en série de puissances entières de t :

$$\varphi(t) = 1 + \frac{k}{\lambda}\,it + \frac{k(k+1)}{2}\left(\frac{it}{\lambda}\right)^2 + \ldots$$

On retrouve simplement à partir de ce développement les moments des divers ordres des E_k.

Si, au lieu de la fonction caractéristique on considère la Transformée de Laplace, celle-ci est obtenue par un calcul analogue :

$$L(s) = \left(1 + \frac{s}{\lambda}\right)^{-k}$$

4.3/ Le fait que, dans un processus de Poisson uniforme les intervalles successifs sont des V.A. indépendantes obéissant à une même loi exponentielle, peut servir de critère pour déceler si une suite d'événements fournie par l'observation peut être raisonnablement assimilée à une suite de réalisations d'un processus de Poisson uniforme.

Les intervalles successifs $(t_{i+1} - t_i)$ étant groupés par classes, on construit un diagramme où l'on porte en abscisse les limites de classes et en ordonnées le logarithme népérien de la fréquence N des intervalles observés supérieurs aux bornes inférieures de chaque classe. Si l'hypothèse est plausible, les points figuratifs sont approximativement alignés.

Dans l'hypothèse d'un processus de Poisson uniforme de densité λ, le nombre des intervalles observés de longueur supérieure à T est une variable aléatoire N dont la distribution a les moyenne et écart-type suivants :

$$E(N) = n\,e^{-\lambda T} \qquad \sigma_N = \left[n\,e^{-\lambda T}(1 - e^{-\lambda T})\right]^{\frac{1}{2}}$$

d'où :

$$\text{Log } E(N) = \log n - \lambda T$$

$$\text{Log } [E(N) \pm \sigma_N] \sim \log E(N) \pm \frac{\sigma_N}{E(N)} = \log n - \lambda T \pm \frac{(1 - e^{-\lambda T})^{\frac{1}{2}}}{(n e^{-\lambda T})^{\frac{1}{2}}}$$

Si les points figuratifs sont groupés pour la plupart dans le domaine compris entre les deux courbes venant d'être définies de part et d'autre de la droite de régression, on considèrera que l'assimilation est plausible.

L'assimilation d'un processus ponctuel à une loi (E_k) $k > 1$ est plus délicate. Nous n'entrerons pas dans le détail de ce problème. On doit avoir :

$$\lambda = \frac{m_1}{\mu_2} \qquad k = \frac{m_1^2}{\mu_2}$$

Il est donc nécessaire que $\frac{m_1^2}{\mu_2}$ soit voisin d'un nombre entier.

V - LOIS DIVERSES ASSOCIEES AUX LOIS E_k -

Il est utile de connaître les principales lois pouvant être définies en fonction des lois (E_k).

5.1/ Posons $\lambda T = m$.

Si $F_m(x)$ désigne la fonction de répartition de la loi de Poisson, on a :

$$F_m(x) = \sum_{k=o}^{k=x-1} e^{-\lambda} \frac{m^k}{k!} = \int_m^{\infty} e^{-u} \frac{u^{x-1}}{(x-1)!} du \qquad \text{(fonction gamma incomplète)}$$

En effet la probabilité d'observer moins de x événements dans l'intervalle m, doit être égale à la probabilité que l'intervalle entre l'arrivant 0 et l'arrivant x soit au moins égal à m.

5.2/ Lois conditionnelles.

La probabilité a priori de x arrivées de T étant $e^{-\lambda T} \frac{(\lambda T)^x}{x!}$ la loi conditionnelle des x variables $t_1 \ldots t_x$, instants de réalisation est donnée par :

$$\frac{x!}{T^x} dt_1 \ldots dt_x$$

En effet, une suite particulière de probabilité élémentaire $\frac{dt_1}{T} \ldots \frac{dt_x}{T}$ peut être obtenue de x! façons différentes.

On peut aussi remarquer que :

$$\iint \ldots \int_{t_1 + \ldots + t_x = T} dt_1 \ldots dt_x = \frac{T^x}{x!}$$

En fait, on s'intéresse surtout dans les applications aux lois conditionnelles discrètes de distribution dans i intervalles égaux divisant T. L'hypothèse d'une seule réalisation au plus par intervalle élémentaire dt ne peut être retenue dans ce cas.

i^x suites de répartition de probabilité uniforme sont a priori possibles, elles ne sont pas toutes topologiquement distinctes puisque l'on considère ici comme semblables celles caractérisées par une même distribution de fréquences. Leur dénombrement s'identifie avec celui de x objets distincts dans i cases.

La probabilité conditionnelle d'une suite particulière :

$$\{e_1 \ldots e_i\} \qquad (e_1 + \ldots + e_i = x)$$

est ainsi donnée par :

$$C_x^{e_1} C_{x-e_1}^{e_2} \ldots C_{e_i}^{e_i} \Big/ i^x = \frac{1}{i^x} \frac{x!}{e_1! \ldots e_i!}$$

5.3/ Lois β.

Considérons r + s intervalles indépendants $t_1 - t_o$, $t_2 - t_1$, ... suivant tous une même loi exponentielle dont la moyenne est prise pour unité. Intéressons-nous à la position de l'instant t_r connaissant celle de l'instant t_{r+s}. Nous devons pour cela considérer la variable

$$u_{rs} = \frac{t_r - t_o}{t_{r+s} - t_o}$$

On montre que sa loi de distribution est la suivante :

$$dA(u_{rs}) = \frac{u^{r-1}(i - u)^{s-1} du}{\int_o^1 u^{r-1}(1 - u)^{s-1} du}$$

(On reconnaît au dénominateur la fonction B(r, s) dite fonction β).

Les réalisations indépendantes $t_r - t_o$ et $t_{r+s} - t_r$ peuvent aussi être appréciées comparativement l'une à l'autre.

Si l'on pose

$$V_{rs} = \frac{t_r - t_o}{t_{r+s} - t_r}$$

On trouve que V_{rs} suit une loi d'expression :

$$dA(V_{rs}) = \frac{1}{B(r;\ s)} \frac{V^{r-1} dV}{(1 + V)^{r+s}}$$

VI - NOMBRE D'EVENEMENTS SURVENANT AU COURS D'UN INTERVALLE ALEATOIRE -

On a fréquemment à considérer la loi de probabilité du nombre de réalisations d'un processus de Poisson uniforme sur un intervalle V non plus constant mais aléatoire de distribution invariable quelconque dB(V).

La probabilité p_j de j réalisations est donnée par

$$p_j = \int_o^\infty e^{-v} \frac{v^j}{j!} dB(v)$$

p_j a pour fonction génératrice :

$$G(z) = \sum p_j z^j = \int_o^\infty e^{-v} \sum \frac{(v z)^j}{j!} d B v = \int_o^\infty e^{-v(1-z)} d B(v)$$

Soit L(t) la transformée de Laplace de B(v) :

$$L(t) = \int_o^\infty e^{-tv} d B(v)$$

On a la relation : G(z) = L(1 - z).

Remarquons que l'on peut l'écrire aussi sous la forme :

$$G(1 + z) = L(-z)$$

Identifiant terme à terme les développements en séries entières des deux membres par rapport à z, on obtient :

$$M_k = m_k$$

où M_k est le moment factoriel d'ordre k de la variable j et m_k le moment d'ordre k de la variable v.

Si l'unité mesurant v est $\frac{1}{\lambda}$; v doit être remplacé par $\frac{V}{\lambda}$
L(t) par $L(\lambda t)$

et :

$$M_k = \lambda^k m_k$$

En particulier :

$$E[j(j - 1)] = \lambda^2 E(v^2)$$

compte tenu de $E j = \lambda E(v)$, on en déduit, la relation suivante liant les variances μ_j et μ_v :

$$\mu_j = \lambda E(v) + \lambda^2 \mu_v$$

Exemple d'application - Supposons que d B(v) soit une exponentielle à k phases :

$$d B(v) = e^{-kv} \frac{(k v)^{k-1}}{(k - 1)!} k d v$$

La transformée de Laplace est $\left(1 + \frac{t}{k}\right)^{-k}$ et la fonction génératrice de la loi cherchée est donnée par :

$$G(s) = \left[1 + \frac{1 - s}{k}\right]^{-k} = p^k(1 - q s)^{-k}$$

où $q = \frac{1}{k + 1}$ $p = \frac{k}{k + 1}$.

G(s) n'est autre que la fonction génératrice d'une distribution de Pascal résultant de l'autocomposition d'ordre k d'une loi géométrique $P_i = q^i p$ (de Fonction génératrice : $\frac{p}{1 - qs}$).

On sait que le développement en série de puissances entières j de $(1 - x)^{-k}$ a pour coefficient général le nombre de combinaisons avec répétition de k objets j à j.

Finalement,

$$P_j = C_{j+k-1}^{j} \, q^j \, p^k$$

Résultat important pouvant être obtenu sans difficultés par un calcul direct.

On notera :

$$E_j = \frac{kq}{p} = 1$$

$$\mu_j = \frac{kq}{p^2} = 1 + \frac{1}{k}$$

VII - LOIS COMPOSEES AVEC UNE LOI DE POISSON -

Soit $\{X_k\}$ une séquence de variables aléatoires mutuellement indépendantes avec une distribution commune $P\{X_k = j\} = f_j$.

Intéressons-nous à la somme

$$S_n = X_1 + \ldots + X_n$$

où n est lui-même une variable aléatoire : nombre de réalisations observées dans un processus de Poisson, au cours d'un intervalle donné.

Nombre de processus importants en pratique se réclament d'un tel schéma. En voici quelques exemples :

Stocks - Les retraits opérés sur un stock surviennent à des instants aléatoires suivant un processus de Poisson et portent sur des quantités probabilisées.

Assurances - Les fréquences de sinistres, obéissent à un processus de Poisson, le montant des dommages associés à chacun d'eux étant probabilisé.

Services divers - Les demandes de transport ou de services tels que restaurant, théâtre, ...) ont lieu non pas isolément mais par "grappes" d'unités (familles, groupes, etc.).

On sait que la fonction génératrice G(S) de la variable S_n, se déduit des fonctions génératrices : P(s) de la loi de n, H(S) de la loi de X par la relation

$$G(S) = P\{H(S)\}$$

(Il suffit d'observer que G(S) n'est autre que l'espérance mathématique de $[H(S)]^n$ par rapport à la distribution de n).

Dans le cas actuel :

$$G(S) = e^{-\lambda(1-H(S))}$$

Remarquons que si H(S) est une loi binomiale q + ps

$$G(S) = e^{-\lambda p(1-s)}$$

La loi composée reste une loi de Poisson de paramètre λp (la connaissance de cette propriété peut être très utile).

Propriétés de la loi de Z et de sa fonction de répartition F(Z) - Des calculs simples montrent que :

$$E(Z) = \lambda\, EX$$
$$\text{Var } Z = \lambda\, E(X^2)$$

La généralisation de ces résultats est la suivante :

Les cumulants k_2 de la loi composée sont liés aux moments (ordinaires) m_2 de la loi d'effectif de grappe par la relation simple :

$$k_r = \lambda\, m_r$$

Dans le cas présent :

$$k_r = \lambda\, r!$$

On est donc conduit naturellement dans le cas d'une distribution la plus générale de l'effectif de grappe à une représentation de la loi de Z par un développement de Gram Charlier, lequel, comme on le sait, fait intervenir les cumulants de la distribution ajustée. On obtient :

$$F(Z) \sim \pi(\vartheta) - \frac{\lambda y_\theta''}{\sigma_z^3} + \frac{\lambda y_\theta'''}{\sigma_z^4} - \frac{\lambda y_\theta^{IV}}{\sigma_z^5} +$$

avec

$$\vartheta = \frac{Z - E(Z)}{\sigma_Z} \qquad \sigma_Z = \sqrt{\lambda\, m_2}$$

et $y^r_{(\theta)}$ dérivée d'ordre r de la loi normale réduite.

On en déduit que lorsque λ grandit indéfiniment F(Z) tend vers la loi normale $\pi(\vartheta)$ (ce résultat semble dû à l'Américain H. Robbins, en 1947, cet auteur ayant utilisé la méthode plus rigoureuse des fonctions caractéristiques).

Un autre type de distributions composées fréquemment utilisé est le suivant :

On considère des lois de Poisson de paramètre $\lambda_1 \ldots \lambda_i$, les probabilités de choix a priori correspondantes étant $u_1 \ldots u_i$, le nombre de réalisations observées n est alors régi par la loi de probabilité

$$p_n = \frac{1}{n!} \sum u_i\, e^{-\lambda_i} (\lambda_i)^n$$

On compare de la même façon les lois E_k. En particulier des lois E_1, on considère ainsi

$$p(t)\, dt = \sum_1^n u_i\, \lambda_i\, e^{-\lambda_i t}\, dt \qquad \text{(généralement } n = r)$$

Des distributions de ce type sont dites hyperexponentielles.

L'intérêt principal de ce type de distributions réside dans leur aptitude à "prolonger" en quelque sorte la gamme des lois E_k, d'un usage courant pour ajuster des lois théoriques à des distributions empiriques.

Il est en effet fréquent dans une distribution d'intervalles par exemple que les intervalles "courts" prédominent nettement sur les intervalles "longs" (tendance à l'agglutinement). De telles distributions sont alors caractérisées par un rapport $\frac{\sigma}{m} > 1$. Dans toutes les lois E_k, on a $\frac{\sigma}{m} \leqslant 1$ alors que dans les distributions de type hyperexponentiel on a nécessairement : $\frac{\sigma}{m} > 1$, ainsi qu'on peut le montrer aisément à partir de l'expression suivante :

$$\left(\frac{\sigma}{m}\right)^2 = \frac{2\sum_1^n \left(\frac{a_i}{\lambda_i}\right)^2}{\left[\sum_1^n \frac{a_i}{\lambda_i}\right]^2} - 1$$

La relative facilité avec laquelle les lois hyperexponentielles peuvent être introduites dans les calculs joints à cette dernière propriété en fait, un outil mathématique intéressant pour des ajustements. En outre, on peut leur trouver une signification théorique dans certains problèmes.