Pub. Inst. Stat. Univ. Paris
1977, XXII - 45 - 53

DISTRIBUTION Z-POISSON

Andrzej GORALSKI

(Conférence à l'Institut de Statistique des Universités de Paris, 27 Mai 1974)

1. PROBLEME

J'ai considéré le modèle d'épreuves consécutives, sous l'hypothèse suivante :

si j-ième événement est arrivé dans la r-ième épreuve ($r = 1, 2, \ldots$) la possibilité que le même évènement arrivera dans s-ième épreuve ($s = r+1, r+2, \ldots, n$) est :

$$p_{j,s} = z\, p_{j,1} \qquad j = 1, 2, \ldots, m,$$

où $0 \leq z$ est le paramètre du modèle .

Quelle est la probabilité $P(n,k,p,z)$ que sur un nombre n d'épreuves, pendant lesquelles la probabilité de l'apparition d'un évenement E est : $p = p_{E,t}$ ($t = 1, 2, \ldots, n$), E arrive k fois ?

2. THEOREME

$$P(n,k,p,z) \underset{\substack{n \to \infty \\ \lambda z < C \\ \lambda = n\, p_{E,1} \\ p_{E,1} = 1/m \quad 1}}{\sim} \begin{cases} 1 - \dfrac{1 - e^{-\lambda z}}{z}, & k = 0 \\[2ex] z^{k-1} \dfrac{\lambda^k}{k!} e^{-\lambda z} & k \geq 1 \end{cases}$$

3. DEMONSTRATION

3.1. Soit $k = 0$

Considérons l'événement $j = a$.

Désignons par b l'événement qui est apparu dans la première épreuve.

On a :

$$p_{a,1} = \frac{1}{m} ,$$

$$p_{b,2} = \frac{z}{m} ,$$

$$p_{a,2} = \frac{m-z}{m(m-1)} .$$

Dans la deuxième épreuve on a :

. ou $E = b$ et $p_{a,3} = \frac{m-z}{m(m-1)}$,

. ou $E = x$ et $p_{a,3} = \frac{m-2z}{m(m-2)}$.
$x \neq a$
$x \neq b$

La valeur espérée de $p_{a,3}$ est donc :

$$p_{a,3} = \frac{m-2z}{m(m-2)} \left(1 - \frac{z(1-z)}{m(m-1)(m-2z)} + \ldots \right)$$

Si nous considérons de la même façon la $(r-1)$ -ième épreuve nous trouverons :

$$p_{a,r} = \frac{m-(r-1)z}{m(m-r+1)} \left(1 - \frac{z(1-z)}{m(m-r+2)(m-(r-1)z)} + \ldots \right) .$$

Mais :

$$\lim_{\substack{m \to \infty \\ n \to \infty \\ z < C}} (1 - \ldots) = 1 ,$$

donc finalement on a :

$$P_{a,r} = \frac{m-(r-1)z}{m(m-r+1)} \quad .$$

Ensuite :

$$P(n,0,p,z) = \prod_{r=1}^{n} (1 - P_{a,r})$$

$$= \prod_{r=1}^{n} \frac{1 - \frac{rm - (r-1)z}{m\,m}}{1 - \frac{r-1}{m}} \quad ,$$

et :

$$\lim_{\substack{n \to \infty \\ \lambda z < C}} \ln P(n,0,p,z) = \ln(1 - \frac{1 - e^{-nz/m}}{z}) \; .$$

3.2. Soit $k = 1$

Considérons l'événement $j = a$.

Il faut trouver :

$$P(n,1,p,z) = \sum_{r=1}^{n} P_r(1) \; ,$$

où $P_r(1)$ est la probabilité que l'événement a apparaitra exactement dans la r-ième épreuve .

On a : $P_1(1) = \frac{1}{m}\left(\frac{m-z}{m}\right)^{n-1}$,

$P_2(1) = P_1(1)$,

$P_3(1) = P(1)\left(1 + \frac{z(1-z)}{m(m-2)}\right)$,

.

.

.

$P_r(1) = P_1(1)\left(1 + \frac{(r-1)(r-2)z(1-z)}{m(2m-(r+1)(r-2))} + \ldots\right)$.

Mais:

$$\lim_{\substack{m \to \infty \\ n \to \infty \\ z < C}} (1 + \dots) = 1 ,$$

donc on a :

$$P_r(1) = \frac{1}{m}\left(\frac{m-z}{m}\right)^{n-1}$$

et :

$$\lim_{\substack{n \to \infty \\ \lambda z < C}} P(n,1,p,z) = \frac{n}{m} e^{-nz/m}$$

3.3. Soit $k \geq 2$

Si nous avons considéré parallèlement aux considérations faites dans le cas où $k = 1$, nous pourrons déduire que :

$$P(n,k,p,z) = \binom{n}{k} z^{k-1} m^{-k} \left(\frac{m-z}{m}\right)^{n-k}$$

et :

$$\lim_{\substack{n \to \infty \\ \lambda z < C}} P(n,k,p,z) = z^{k-1} \left(\frac{n}{m}\right)^{k} \frac{1}{k!} e^{-nz/m}$$

4. DISTRIBUTION Z-POISSON

Nous disons qu'une variable aléatoire **X** a la distribution z-Poisson si :

$$P(X=k) = \begin{cases} 1 - \dfrac{1 - e^{-\lambda z}}{z} & k = 0, \\[2ex] z^{k-1} \dfrac{\lambda^{k}}{k!} e^{-\lambda z} & k \geq 1. \end{cases}$$

Valeur espérée X :

$$E(X) = \sum_{k=0}^{\infty} k\, P(X=k) = \frac{e^{-\lambda z}}{z} \sum_{k=1}^{\infty} k \frac{(\lambda z)^k}{k!} = \lambda ,$$

variance X:

$$D^2(X) = E(X^2) - E^2(X) = E(X(X-1)) - E(X)\,(E(X)-1)$$

$$= \frac{e^{-\lambda z}}{z} \sum_{k=1}^{\infty} k(k-1) \frac{(\lambda z)^k}{k!} - \lambda\,(\lambda-1) = \lambda(1-\lambda\,(1-z)).$$

5. APPROXIMATION DE LA DISTRIBUTION DE POLYA PAR LA DISTRIBUTION Z-POISSON

Nous savons que la distribution de Polya pour une variable aléatoire X est :

$$P(X=k) = P(n,k,p,a)$$

$$= \binom{n}{k} \prod_{i=0}^{k-1} (p+ia) \prod_{j=0}^{n-k-1} (1-pja) \prod_{r=0}^{n-1} (1+ra)^{-1} ,$$

où $P(n,k,p,a)$ est la probabilité que sur un nombre n d'épreuves l'événement considéré E arrivera k fois; p est la probabilité que E arrivera dans la première épreuve, et a est un paramètre.

Si p est petite on peut faire une approximation de la distribution de Polya par la distribution z-Poisson, avec :

$$z = 1 + \frac{a}{1+a} \; \frac{1-p}{p} .$$

Donc on a :

$$P(n,k,p,a) \underset{p \ll 1}{\approx} P(n,k,p,z).$$

Si nous posons, par exemple, $n=10$, $p=0,02$ et $a=-0,01$ nous aurons : $z = 0,505$

et:

k =	0	1	2	3
$P(n,k,p,a)$ =	0,809	0,182	0,009	0,000
$P(n,k,p,z)$ =	0,810	0,181	0,009	0,000

6. TROIS EXEMPLES D'APPLICATIONS DE LA DISTRIBUTION Z-POISSON

6.1. Distribution des accidents de travail

Considérons la répartition des nombres des accidents qui sont arrivés aux travailleurs d'une certaine usine pendant un délai de 13 mois (B.Biegelsein-Zelazowski: Zarys psychologii pracy, PWN,1967,p.180) :

nombres des accidents (k) =	0	1	2	3	4
nombres des travailleurs(n(k)) =	428	132	62	21	5

Posons l'hypothèse suivante : la répartition des accidents de travail s'effectue selon la distribution z-Poisson.

Nous aurons ici :

$\lambda = 0,526$ $\qquad$ $z=1,79$

et des nombres espérés dans l'hypothèse :

k =	0	1	2	3	≥ 4
$\hat{n}(k)$ =	427,0	133,4	62,6	19,6	5,4

Le test du χ^2 donne ici : $\chi^2 = 0,15$ pour $\upsilon = 2$.

On peut dire aussi que ce n'est pas une bonne usine -z est égale à 1,79.

6.2. Comparaison des stratégies

Les données analysées ici proviennent de l'étude effectuée en 1968 dans l'Institut de Psychologie de l'Université de Varsovie (M. Materska : Tresc przygotowania teoretycznego a struktura czynnosci pratycznych, Wroclaw, 1972, Ossolineum).

Voici le tableau expérimental :

On a deux groupes (W,M) très semblables à ce point de vue de facteurs tels que : âge, sexe, niveau d'intelligence, motivations, etc., et très distinctifs au point de vue du facteur principal qui peut être dénommé "stratégie pour poser les questions". Dans chacun des groupes on a trente personnes. Chaque personne peut poser n fois ($n \leq 41$) chacune des m questions ($m=133$).

Etant donné les répartitions de questions :

	groupe W				groupe M			
k =	0	1	2	3	0	1	2	3
n(k) =	3123	785	77	5	3045	905	40	-

comparer les stratégies.

Si on suppose que les répartitions suivent la distribution z-Poisson on a:

groupe W	groupe M
$\lambda = 0{,}239$	$\lambda = 0{,}247$
$z = 0{,}81$	$z = 0{,}33$

Il est évident que les valeurs moyennes sont proches, c'est-à-dire : non distinctives. Mais voilà ce sont les valeurs du paramètre z qui font la taxinomie!

6.3. Un exemple historique

J'ai examiné récemment (A.Goralski : Semi-deterministyczny model zmiennosci resu i jego aplikacja histoeyczna,Procc.Hist. Int.Coll., Varsovie, Novembre 1973) les répartitions de nombres de changements territoriaux de mon pays dans les successives "portions" du temps, égales à cinq années, et j'ai obtenu le tableau suivant :

k =	0	1	2	3	période
n(k) =	48	16	3	1	961-1300
	35	13	4	1	1301-1565
	32	9	5	-	1566-1795

Faisant les calculs nécessaires on voit que :

période	λ	z
961 - 1300	0,368	1,30
1301 - 1565	0,453	1,31
1566 - 1795	0,413	1,29

Alors on constate que la densité de changements est variable et que la valeur z reste constante, ce qui est d'ailleurs étonnant.

BIBLIOGRAPHIE

(1) A.GORALSKI : O pewnym schemacie losowania zwrotnego, Zeszyty Naukowe PW, Matematyka 14, Warszawa, 1968.

(2) S.D.POISSON : Recherches sur la probabilité des jugements, Bachelier, Paris, 1837.

(3) G.POLYA : Sur quelques points de la théorie des probabilités, Ann. Inst.H.Poincaré, 1 ,Paris, 1931.

Reçu en Juin 1974

02-643 WARSZAWA
Etiudy R. 40/3
Pologne