M. F I S Z (Warszawa)

ROZKŁAD GRANICZNY ZMIENNEJ LOSOWEJ, BĘDĄCEJ RÓŻNICĄ DWÓCH NIEZALEŻNYCH ZMIENNYCH LOSOWYCH O ROZKŁADZIE POISSONA

J. Oderfeld zakomunikował mi interesującą obserwację: zmienna losowa $x = x_1 - x_2$, będąca różnicą dwóch niezależnych zmiennych losowych x_1 i x_2 o rozkładach Poissona

$$P\{x_1 = k\} = \frac{\lambda_1^k}{k!} e^{-\lambda_1}, \qquad P\{x_2 = k\} = \frac{\lambda_2^k}{k!} e^{-\lambda_2}, \tag{1}$$

gdzie $\lambda_1 > 0$, $\lambda_2 > 0$ i $k = 0,1,2,\ldots$, ma często rozkład zbliżony do normalnego.

Zagadnienie to w przypadku różnicy dwóch takich samych zmiennych ($\lambda_1 = \lambda_2$) było rozważane przez Irwina [1]. Stwierdził on, że rozkład różnicy dwóch takich niezależnych zmiennych zmierza do rozkładu normalnego, gdy wartość przeciętna zmiennej składowej wzrasta do ∞. W pracy tej zajmuję się rozkładem granicznym różnicy dwóch niezależnych zmiennych o rozkładzie Poissona w przypadku ogólnym.

Niech x_1 i x_2 będą dwiema niezależnymi zmiennymi losowymi o rozkładach Poissona (1) i niech $x = x_1 - x_2$. Funkcje charakterystyczne zmiennych x_1 i $-x_2$ są następujące:

$$\varphi_{x_1}(t) = e^{\lambda_1(e^{it}-1)},$$
$$\varphi_{-x_2}(t) = e^{\lambda_2(e^{-it}-1)}.$$

Wskutek niezależności zmiennych losowych x_1 i x_2, a więc także x_1 i $-x_2$, zmienna losowa x ma funkcję charakterystyczną

$$\varphi_x(t) = \varphi_{x_1}(t)\varphi_{-x_2}(t) = e^{\lambda_1(e^{it}-1)+\lambda_2(e^{-it}-1)}.$$

Rozwijając $e^{\pm it}$ w szeregi potęgowe, po dokonaniu redukcji znajdujemy

$$\varphi_x(t) = e^{(\lambda_1-\lambda_2)\frac{it}{1!}+(\lambda_1+\lambda_2)\frac{(it)^2}{2!}+(\lambda_1-\lambda_2)\frac{(it)^3}{3!}+\dots} \tag{2}$$

Z (2) wynika, że wszystkie nieparzyste półniezmienniki zmiennej x są $\lambda_1 - \lambda_2$, a wszystkie parzyste są $\lambda_1 + \lambda_2$.

Rozpatrzmy teraz standardyzowaną zmienną losową

$$X = \frac{x - (\lambda_1 - \lambda_2)}{\sqrt{\lambda_1 + \lambda_2}}.$$

Z (2) na mocy wzoru

$$\varphi_{ax+b}(t) = \varphi_x(at)\,e^{ibt}$$

otrzymujemy funkcję charakterystyczną zmiennej X:

$$\varphi_X(t) = e^{-\frac{t^2}{2}+\frac{(\lambda_1-\lambda_2)}{(\lambda_1+\lambda_2)^{3/2}}\cdot\frac{(it)^3}{3!}+\frac{1}{\lambda_1+\lambda_2}\cdot\frac{(it)^4}{4!}+\frac{(\lambda_1-\lambda_2)}{(\lambda_1+\lambda_2)^{5/2}}\cdot\frac{(it)^5}{5!}+\dots} \tag{3}$$

Z (3) widać, że dla λ_1 i λ_2 dostatecznie dużych $\varphi_X(t)$ różni się dowolnie mało od $e^{-t^2/2}$, a więc od funkcji charakterystycznej standardyzowanej zmiennej normalnej.

Weźmy z kolei pod uwagę zmienną losową

$$y = x_1 + x_2.$$

Zmienna y, jako suma niezależnych zmiennych losowych o rozkładzie Poissona, ma także rozkład Poissona, przy czym wszystkie półniezmienniki są $\lambda_1 + \lambda_2$.

Standardyzując zmienną losową y, otrzymamy zmienną

$$Y = \frac{y - (\lambda_1 + \lambda_2)}{\sqrt{\lambda_1 + \lambda_2}}.$$

Funkcją charakterystyczną tej zmiennej jest

$$\varphi_Y(t) = e^{-\frac{t^2}{2}+\frac{1}{(\lambda_1+\lambda_2)^{1/2}}\cdot\frac{(it)^3}{3!}+\frac{1}{\lambda_1+\lambda_2}\cdot\frac{(it)^4}{4!}+\frac{1}{(\lambda_1+\lambda_2)^{3/2}}\cdot\frac{(it)^5}{5!}+\dots} \tag{4}$$

Porównując (3) i (4) widzimy, że rozkład zmiennej X jest dla tych samych wartości λ_1 i λ_2 bliższy rozkładu normalnego niż rozkład zmiennej Y. Różnica ta jest tym wyraźniejsza, im mniej się różnią λ_1 i λ_2. Można więc stwierdzić, że wspomniana na wstępie obserwacja Oderfelda jest teoretycznie uzasadniona, gdy λ_1 i λ_2 nie są zbyt małe.

Udowodnione twierdzenie pozwala w praktyce stosować następujące przybliżone wzory na gęstość i dystrybuantę zmiennej losowej x:

$$P\{x=n\} \approx f(n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(n-a)^2}{2\sigma^2}} \quad (n = \text{liczba całkowita}), \tag{5}$$

$$P\{x \leqslant n\} \approx F(n) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{n+1/2} e^{-\frac{(t-a)^2}{2\sigma^2}} dt \quad (n = \text{liczba całkowita}), \tag{6}$$

otrzymane z rozkładu normalnego o wartości średniej $a = \lambda_1 - \lambda_2$ i wariancji $\sigma^2 = \lambda_1 + \lambda_2$. Dla zobrazowania dokładności tego przybliżenia podajemy dwie tablice dokładnych i przybliżonych wartości gęstości i dystrybuanty zmiennej losowej x dla dwu par wartości $\lambda_1 = \lambda_2 = 1{,}5$ oraz $\lambda_1 = 2$, $\lambda_2 = 1$. Z tablic I i II wynika, że dla $\lambda_1 + \lambda_2 = 3$ błąd przybliżenia nie osiąga wartości 0,02. Należy przy tym zwrócić uwagę, że błędy przybliżenia maleją szybko w miarę oddalania się od wartości oczekiwanej.

Tablica I

Wartości gęstości i dystrybuanty zmiennej losowej x oraz ich przybliżenia na podstawie wzorów (5) i (6) ($\lambda_1 = \lambda_2 = 1{,}5$, $\sigma^2 = \lambda_1 + \lambda_2 = 3$).

n	$P\{x=n\}$	$f(n)$	$P\{x=n\}-$ $-f(n)$	$P\{x\leqslant n\}$	$F(n)$	$P\{x\leqslant n\}-$ $-F(n)$
−9	0,00001	0,00000	0,00001	0,00001	0,00000	0,00001
−8	0,00004	0,00001	0,00003	0,00005	0,00001	0,00004
−7	0,00022	0,00007	0,00015	0,00027	0,00009	0,00018
−6	0,00108	0,00057	0,00051	0,00135	0,00075	0,00060
−5	0,00454	0,00357	0,00097	0,00589	0,00469	0,00120
−4	0,01622	0,01600	0,00022	0,02211	0,02166	0,00045
−3	0,04778	0,05139	−0,00361	0,06989	0,07446	−0,00457
−2	0,11178	0,11826	−0,00648	0,18167	0,19324	−0,01157
−1	0,19683	0,19497	0,00186	0,37850	0,38641	−0,00791
0	0,24300	0,23033	0,01267	0,62150	0,61359	0,00791
1	0,19683	0,19497	0,00186	0,81833	0,80676	0,01157
2	0,11178	0,11826	−0,00648	0,93011	0,92554	0,00457
3	0,04778	0,05139	−0,00361	0,97789	0,97834	−0,00045
4	0,01622	0,01600	0,00022	0,99411	0,99531	−0,00120
5	0,00454	0,00357	0,00097	0,99865	0,99925	−0,00060
6	0,00108	0,00057	0,00051	0,99973	0,99991	−0,00018
7	0,00022	0,00007	0,00015	0,99995	0,99999	−0,00004
8	0,00004	0,00001	0,00003	0,99999	1,00000	−0,00001
9	0,00001	0,00000	0,00001	1,00000	1,00000	0,00000

Tablica II

Wartości gęstości i dystrybuanty zmiennej losowej x oraz ich przybliżenia na podstawie wzorów (5) i (6) ($\lambda_1=2$, $\lambda_2=1$, $\sigma^2=\lambda_1+\lambda_2=3$).

n	$P\{x=n\}$	$f(n)$	$P\{x=n\}-$ $-f(n)$	$P\{x\leqslant n\}$	$F(n)$	$P\{x\leqslant n\}-$ $-F(n)$
-7	0,00001	0,00001	0,00000	0,00001	0,00001	0,00000
-6	0,00009	0,00007	0,00002	0,00010	0,00009	0,00001
-5	0,00057	0,00057	0,00000	0,00067	0,00075	$-0,00008$
-4	0,00306	0,00357	$-0,00051$	0,00373	0,00469	$-0,00096$
-3	0,01338	0,01600	$-0,00262$	0,01711	0,02166	$-0,00455$
-2	0,04624	0,05139	$-0,00515$	0,06335	0,07446	$-0,01111$
-1	0,11923	0,11826	0,00097	0,18258	0,19324	$-0,01066$
0	0,21171	0,19497	0,01674	0,39429	0,38641	0,00788
1	0,23846	0,23033	0,00813	0,63275	0,61359	0,01916
2	0,18496	0,19497	$-0,01001$	0,81771	0,80676	0,01095
3	0,10701	0,11826	$-0,01125$	0,92472	0,92554	$-0,00082$
4	0,04891	0,05139	$-0,00248$	0,97363	0,97834	$-0,00471$
5	0,01839	0,01600	0,00239	0,99202	0,99531	$-0,00329$
6	0,00586	0,00357	0,00229	0,99788	0,99925	$-0,00137$
7	0,00162	0,00057	0,00105	0,99950	0,99991	$-0,00041$
8	0,00039	0,00007	0,00032	0,99989	0,99999	$-0,00010$
9	0,00009	0,00001	0,00008	0,99998	1,00000	$-0,00002$
10	0,00002	0,00000	0,00002	1,00000	1,00000	0,00000

Dla $\lambda_1=\lambda_2=0,5$ błąd przybliżenia gęstości wynosi 0,00001 dla $n=-6$ i $n=6$, a osiąga maksymalną wartość 0,06682 dla $n=0$. Dla $\lambda_1=\lambda_2=3$ błąd ten wynosi 0,00001 dla $n=-11$ i $n=12$, a wartość maksymalną 0,00379 osiąga dla $n=0$. Wreszcie dla $\lambda_1=5$, $\lambda_2=1$ błąd wynosi 0,00001 dla $n=-7$ i $n=17$, a maksymalną wartość 0,01156 osiąga dla $n=2$.

Jest jasne, że w zagadnieniach weryfikacji hipotez statystycznych, wymagających zazwyczaj jedynie znajomości krańcowych części rozkładu, korzystanie z rozkładu normalnego nie budzi zastrzeżeń.

Państwowy Instytut Matematyczny

(Praca wpłynęła dnia 4. 6. 1951 r.)

[1] J. O. Irwin, *The frequency distribution of the difference between two independent variates following the same Poisson distribution*, Journal of the Royal Statistical Society, 1937.

М. ФИШ (Варшава)

ПРЕДЕЛЬНОЕ РАСПРЕДЕЛЕНИЕ РАЗНОСТИ ДВУХ НЕЗАВИСИМЫХ ПЕРЕМЕННЫХ РАСПРЕДЕЛЕННЫХ ПО ЗАКОНУ ПУАССОНА

РЕЗЮМЭ

Случайные величины x_1 и x_2 независимы и распределены по закону Пуассона, а их средние величины равны соответственно λ_1 и λ_2. В работе доказывается, что предельным законом разности $x_1 - x_2$ — когда λ_1 и λ_2 стремятся к бесконечности — является нормальный закон.

Из таблиц I и II следует, что для целей статистической проверки гипотез где уровень значимости по обыкновению мал, можно уже для сравнительно небольших значений λ_1 и λ_2 пользоваться предельным распределением.

M. FISZ (Warszawa)

THE LIMITING DISTRIBUTION OF THE DIFFERENCE OF TWO POISSON RANDOM VARIABLES

SUMMARY

The random variables x_1 and x_2 are independent and have a Poisson distribution with the respective mean values λ_1 and λ_2.

It is shown that as λ_1 and λ_2 converge to infinity the distribution of the difference $x_1 - x_2$ converges to the normal distribution.

It results from tables I and II that when testing statistical hypotheses on a low significance level one can use the limiting normal distribution for relatively small values of λ_1 and λ_2.
