

12 2項分布のPoisson分布による近似

12.1 Poisson 分布

任意の時刻において同じ確率でランダムに起こる事象があって、1時間あたりの平均生起回数が λ であるとします。この事象が1時間中に k 回起きる確率はどうなるでしょうか。

連続時間は難しいので、離散時間で近似的に考えます。例えば1分に1回通常の6面サイコロが振られる状況で『5が出る』事象を考えると（5以外が出た場合は『何も起きなかった』、スルーです）、1時間のうちでは60回（独立に）振られることになりますから、60回中5が出た回数の平均値は10回です。これはつまり $\lambda = 10$ であることを意味します。

このとき1時間における生起回数を表す確率変数を X とすれば、 X は2項分布 $B(60, \frac{1}{6})$ に従い、

$$P[X = k] = \binom{60}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{60-k}$$

です。

次に時刻の区切りをもう少し小さくします。毎秒ごとに1回6面サイコロが振られることにしてしまうと、1時間では3600回も振られることになってしまい、そのうち5が出る回数の平均値 λ は600回であり、最初の仮定、『1時間あたりの平均生起回数が $\lambda (= 10)$ である』に反します。したがって $\lambda = 10$ を固定したままにするためには、1回あたりの5の出る確率を変えるしかありません。そこでどの面も同じ確率で出る360面のサイコロ（そんなものが現実是否存在するかどうかは別として）を用意して毎秒ごとにこれが振られることにすれば、1回あたり5の出る確率は $\frac{1}{360}$ であって、3600回中に5の出る回数の平均は10回です。この場合も1時間における生起回数表す確率変数を X とすれば、 X は2項分布 $B(3600, \frac{1}{360})$ に従い、

$$P[X = k] = \binom{3600}{k} \left(\frac{1}{360}\right)^k \left(\frac{359}{360}\right)^{3600-k}$$

です。

更に時刻の区切りを小さくしましょう。1時間を10000等分して、各時刻に1000面サイコロが振られます（5の出る確率は $\frac{1}{1000}$ です）。すると1時間中での5の出る回数の平均値は10です。この場合も1時間における生起回数表す確率変数を X とすれば、 X は2項分布 $B(10000, \frac{1}{1000})$ に従い、

$$P[X = k] = \binom{10000}{k} \left(\frac{1}{1000}\right)^k \left(\frac{999}{1000}\right)^{10000-k}$$

です。

一般に1時間を N 等分して、 $\frac{\lambda}{N}$ の確率で当たりの出るくじが等分した時間ごとに1時間に N 回引かれることにすれば、1時間あたりの『当たり』の発生回数の平均値は λ であり、この場合は1時間における当たり発生回数表す確率変数を X_N とすれば、 X_N は2項分布 $B(N, \frac{\lambda}{N})$ に従い、

$$P[X_N = k] = \binom{N}{k} \left(\frac{\lambda}{N}\right)^k \left(1 - \frac{\lambda}{N}\right)^{N-k}$$

です。これは $N \rightarrow \infty$ において

$$\begin{aligned} P[X_N = k] &= \binom{N}{k} \left(\frac{\lambda}{N}\right)^k \left(1 - \frac{\lambda}{N}\right)^{N-k} \\ &= \frac{N \cdot (N-1) \cdots (N-k+1)}{k!} \frac{\lambda^k}{N^k} \left(1 - \frac{\lambda}{N}\right)^N \left(1 - \frac{\lambda}{N}\right)^{-k} \end{aligned}$$

ここで

$$\lim_{X \rightarrow \infty} \left(1 + \frac{1}{X}\right)^X = e, \quad \lim_{X \rightarrow -\infty} \left(1 + \frac{1}{X}\right)^X = e$$

なので（後者で $-\frac{N}{\lambda} = X$ と考えれば）

$$\left(1 - \frac{\lambda}{N}\right)^N = \left\{ \left(1 + \frac{1}{-\frac{N}{\lambda}}\right)^{-\frac{N}{\lambda}} \right\}^{-\lambda} \rightarrow e^{-\lambda} \quad (N \rightarrow \infty)$$

となることに注意して

$$= \frac{\lambda^k}{k!} \underbrace{\frac{N \cdots (N-k+1)}{N \cdots N}}_{\downarrow 1} \underbrace{\left(1 - \frac{\lambda}{N}\right)^N}_{\downarrow e^{-\lambda}} \underbrace{\left(1 - \frac{\lambda}{N}\right)^{-k}}_{\downarrow 1}$$

となり、

$$\lim_{N \rightarrow \infty} P[X_N = k] = \frac{\lambda^k}{k!} e^{-\lambda}$$

が分かります。ここで e^x のTaylor展開から

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k = e^{-\lambda} e^{\lambda} = 1$$

が成り立っており、

k	0	1	2	$\dots$	k	$\dots$
$P[X = k]$	$e^{-\lambda}$	$\lambda e^{-\lambda}$	$\frac{\lambda^2}{2!} e^{-\lambda}$	$\dots$	$\frac{\lambda^k}{k!} e^{-\lambda}$	$\dots$

であるような確率変数の分布を考えることが出来、これをパラメーター λ のポアソン分布 (Poisson distribution) $P_O(\lambda)$ と言います。これは、今見たように、ある単位時間内に平均 λ 回起こる事象が、単位時間内に起こる回数に対応した確率変数と考えられます。

パラメータ λ のポアソン分布に従う X の平均値は

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

であり、分散も

$$\begin{aligned} \sum_{k=0}^{\infty} k^2 \frac{\lambda^k}{k!} e^{-\lambda} - \lambda^2 &= \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} + \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} - \lambda^2 \\ &= \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} e^{-\lambda} + \lambda - \lambda^2 \\ &= \lambda \\ &= Var[X] \end{aligned}$$

となっています。

例 12.1.1 Maria Dolores Ugarte らの研究 (2016) によれば、サッカー W 杯における 1 試合の平均ゴール数は 2.5 であり、ポアソン分布が上手く適用されるようです：

k	1 試合のゴール数が k である確率
0	$e^{-2.5} \approx 0.082$
1	$\frac{2.5}{1} e^{-2.5} \approx 0.205$
2	$\frac{2.5^2}{2!} e^{-2.5} \approx 0.257$
3	$\frac{2.5^3}{3!} e^{-2.5} \approx 0.213$
4	$\frac{2.5^4}{4!} e^{-2.5} \approx 0.133$

例 12.1.2 1 時間に平均 4 回電話が掛かってくるコールセンターで、1 時間に電話が掛かってくる回数も、ポアソン分布で求めることが出来ます。

1 時間当たりの平均コール数は 4 ですからパラメータ 4 のポアソン分布で計算します。

k	1 時間に k 回コールがある確率
0	0.0183
1	0.0732
2	0.1465
3	0.1954
4	0.1954
5	0.1563
6	0.1042
7	0.0595
8	0.0298
合計	0.9786

1 時間の業務中に 1 回もコールがない確率が 2 %、9 回以上コールがある確率も 2 % あります。

つまり、1 時間業務を 50 回こなせば、全くコールのない日も、逆にコールが 9 回以上あるような日もどちらも 1 回ぐらいは経験するのが『普通』と云うことです。

問題 12.1.3 10 年に平均 1 回氾濫する川が 1 年に 2 回氾濫する確率を求めてください。ただし、1 年に氾濫する回数はポアソン分布に従うものと仮定し、 $e^{-0.1} \approx 0.904$ とします。

『10 年に 1 回』というのを『1 年に 0.1 回』と解釈して考えましょう。

パラメータ 0.1 のポアソン分布に従うと考えれば、1 年間で 2 回氾濫する確率は

$$\frac{0.1^2}{2!} e^{-0.1} \approx 0.00452$$

です。

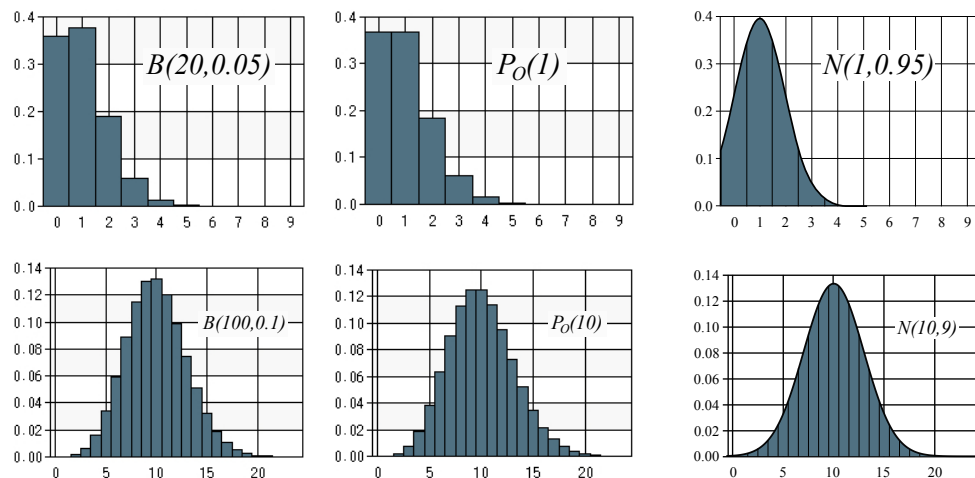
□

12.2 2項分布の Poisson 分布による近似

Poisson 分布を導入した経緯を考えれば、2項分布は以下のように近似されます：

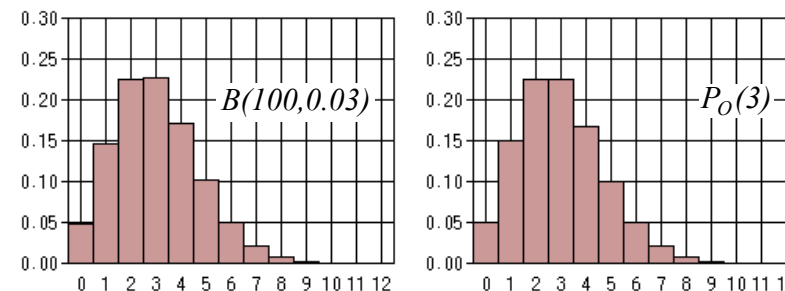
事実 12.2.1 2項分布 $B(n, p)$ は、 n が十分大きく、 p が十分小さければ、具体的には $n \geq 20, p \leq 0.05$ であるか、あるいは、 $n \geq 100, np \leq 10$ であればパラメータ np の Poisson 分布 $P_O(np)$ で近似されます。

$$P[B(n, p) = j] \approx P[P_O(np) = j], \quad j = 0, 1, 2, \dots, n$$



問題 12.2.2 非常に多くの同種の製品の中に 3% の不良品が含まれています。今この製品の中から任意に 100 個抽出するとき、不良品が 2 個以下である確率を求めてください。ただし、製品は非常に多いので抽出は復元抽出であるとみなして良いものとします。 $e^{-3} \approx 0.0498$ 。

含まれる不良品の数 X は 2 項分布 $B(100, 0.03)$ に従います。基準を満たしている程度に n は大きく p は小さいので、パラメータ 3 の Poisson 分布 $P_O(3)$ で近似します。



すると

$$P[X \leq 2] \approx \sum_{k=0}^2 \frac{3^k}{k!} e^{-3} \approx (1 + 3 + 4.5) 0.0498 \approx 0.423$$

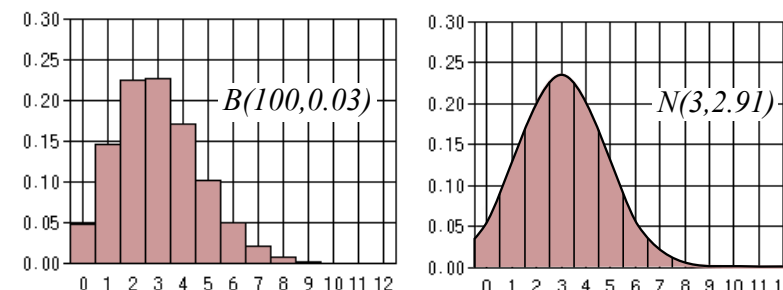
が得られます。

ちなみにコンピュータを使って 2 項分布で直接正確な値を計算すると 0.420 です。

また、 $np = 3 < 5$ なので正規分布による近似は利用できませんが、そこを敢えて近似してみると、分散は

$$100 \cdot 0.03 \cdot 0.97 = 2.91$$

ですから



$$P[N(100 \cdot 0.03, 100 \cdot 0.03 \cdot 0.97) \leq 2 + 0.5] \approx 0.3847$$

です。やはり Poisson 分布の方が良い近似になっています。

12.3 問題演習

● Poisson 分布

基本演習 12.1 自動車の通行台数が単位時間につき平均 5 台である道路において、単位時間に通行する自動車の台数を X とします。 X はポアソン分布に従うものとして、単位時間に通過する自動車の台数が 2 台以下である確率を求めてください。ただし、 $e^{-5} \approx 0.00674$ とします。

基本演習 12.2 ある家では 1 日に平均 10 回の電話がかかるといいます。1 日にかかる電話の回数 X がポアソン分布に従うものとして、 X が 5 以上である確率を求めてください。ただし、 $e^{-10} \approx 0.0000454$ とします。

● 2 項分布の Poisson 分布による近似

基本演習 12.3 袋の中に 1000 個の玉が入っていて、そのうち 4 個が赤玉です。いま袋の中から 1 個ずつ 500 回復元抽出するとき、赤玉がちょうど 5 回出る確率を求めてください。ただし、 $e^{-2} \approx 0.135$ とします。

基本演習 12.4 当たりくじが 1000 本中 3 本入っているくじを 100 本買ったとき、当たりくじが 2 本以上含まれている確率を求めてください。ただし、このくじは極めて多数売られているため、復元抽出したとみなせるとし、 $e^{-0.3} \approx 0.741$ とします。

基本演習 12.5 りんごを 250 個ずつ箱に詰めます。箱詰めされたリンゴは平均して 0.8 % が腐ります。箱詰めリンゴを 1 箱開けたとき、腐ったリンゴが 3 個以上入っている確率を求めてください。ただし、 $e^{-2} \approx 0.135$ とします。

● 2 項分布の正規分布による近似 (先週の演習問題・再掲)

基本演習 11.1 2 項分布 $B(3, p)$ に従う確率変数 X の平均値 (期待値)・分散を 2 項分布の分布表から計算して下さい。

基本演習 11.2 [教科書 例題 15.6] 正常な硬貨を 400 回投げて表の出る回数を確率変数 X とするとき、 $P[190 \leq X \leq 210]$ を求めて下さい (半整数補正あり)。

基本演習 11.3 [教科書 練習問題 16-7] 正常なサイコロをくりかえし投げて実際に 1 の目が出る割合とその数学的確率 $\frac{1}{6}$ との差が 0.1 以下になる確率を考えま

す。その確率が 95 % 以上であるようにするためには少なくとも何回投げれば良いでしょうか。ただし、正規分布で近似するときに $\frac{1}{2}$ の補正はしなくて結構です。

基本演習 11.4 [教科書 練習問題 15-5] 袋の中に赤玉 3 個と白玉 7 個が入っています。その中から復元抽出法で 1 個ずつ玉を繰り返し取り出すとき赤玉の出た回数を X で表すことにします。

(1) 100 回取り出すとき $P[X < 27 \text{ 又は } 33 < X]$ を求めて下さい。

(2) 1000 回取り出すとき $P[X < 270 \text{ 又は } 330 < X]$ を求めて下さい。

問題演習解答例

基本演習 12.1 自動車の通行台数が単位時間につき平均 5 台である道路において、単位時間に通行する自動車の台数を X とします。 X はポアソン分布に従うものとして、単位時間に通過する自動車の台数が 2 台以下である確率を求めてください。

ただし、 $e^{-5} \approx 0.00674$ とします。

X はパラメータ 5 のポアソン分布に従いますから、

$$P[X \leq 2] = \sum_{k=0}^2 P[X = k] = \sum_{k=0}^2 \frac{5^k}{k!} e^{-5} \approx \left(1 + 5 + \frac{25}{2}\right) 0.00674 \approx 0.125$$

です。

基本演習 12.2 ある家では 1 日に平均 10 回の電話がかかるといいます。1 日にかかる電話の回数 X がポアソン分布に従うものとして、 X が 5 以上である確率を求めてください。ただし、 $e^{-10} \approx 0.0000454$ とします。

X はパラメータ 10 のポアソン分布に従いますから

$$\begin{aligned} P[X \geq 5] &= 1 - P[X \leq 4] \\ &= 1 - \sum_{k=0}^4 P[X = k] \\ &= 1 - \sum_{k=0}^4 \frac{10^k}{k!} e^{-10} \\ &= 1 - \left(1 + 10 + \frac{100}{2} + \frac{1000}{6} + \frac{10000}{24}\right) 0.0000454 \\ &\approx 0.971 \end{aligned}$$

です。

基本演習 12.3 袋の中に 1000 個の玉が入っていて、そのうち 4 個が赤玉です。いま袋の中から 1 個ずつ 500 回復元抽出するとき、赤玉がちょうど 5 回出る確率を求めてください。ただし、 $e^{-2} \approx 0.135$ とします。

500 回引いた中の赤玉の数を表す確率変数 X は 2 項分布 $B(500, \frac{4}{1000})$ に従います。

$np = 2$ なので正規分布で近似することは出来ませんが、パラメータ 2 のポアソン分布で近似することが出来ます。

$$P[X = 5] \approx P[Po(2) = 5] = \frac{2^5}{5!} e^{-2} \approx \frac{4}{15} 0.135 = 0.036.$$

基本演習 12.4 当たりくじが 1000 本中 3 本入っているくじを 100 本買ったとき、当たりくじが 2 本以上含まれている確率を求めてください。ただし、このくじは極めて多数売られているため、復元抽出したとみなせるとし、 $e^{-0.3} \approx 0.741$ とします。

100 本中の当たりくじの数を表す確率変数 X は 2 項分布 $B(100, \frac{3}{1000})$ に従うとみなせます。 $np = 0.3$ なので正規分布では近似できませんが、パラメータ 0.3 のポアソン分布で近似することができます。

$$\begin{aligned} P[X \geq 2] &= 1 - \sum_{k=0}^1 P[X = k] \\ &\approx 1 - P[Po(0.3) = 0] - P[Po(0.3) = 1] \\ &= 1 - (1 + 0.3)e^{-0.3} \\ &\approx 1 - 1.3 \cdot 0.741 \\ &\approx 0.037. \end{aligned}$$

基本演習 12.5 りんごを 250 個ずつ箱に詰めます。箱詰めされたリンゴは平均して 0.8 %が腐ります。箱詰めリンゴを 1 箱開けたとき、腐ったリンゴが 3 個以上入っている確率を求めてください。ただし、 $e^{-2} \approx 0.135$ とします。

250 個の中の腐ったリンゴの数を X とすれば、これは 2 項分布 $B(250, 0.008)$ に従います。 $np = 2$ なので正規分布で近似することはできませんが、パラメータ 2 のポアソン分布で近似することが出来、

$$P[X \geq 3] = 1 - \sum_{k=0}^2 P[X = k] \approx 1 - \sum_{k=0}^2 \frac{2^k}{k!} e^{-2} = 1 - (1 + 2 + 2) 0.135 \approx 0.325$$

です。