二項分布の正規近似

定理 (ド・モアブル-ラプラスの定理)

X が $B(n,p)$ に従うとき, $Z = \dfrac{X-np}{\sqrt{np(1-p)}}$ と標準化すると, $n \to \infty$ のとき Z は $N(0,1)$ に従う.

- この定理を証明するために, 「階乗の近似」である **Stirling (スターリング) の公式**

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}} = 1 \qquad \text{を使う.}$$

(証明)　スターリングの公式から,

$$\begin{aligned} n! &\approx \sqrt{2\pi n} e^{-n} n^n \quad (n\to\infty), \\ k! &\approx \sqrt{2\pi k} e^{-k} k^k \quad (k\to\infty), \\ (n-k)! &\approx \sqrt{2\pi(n-k)} e^{-(n-k)} (n-k)^{n-k} \quad (n-k\to\infty) \end{aligned}$$

が成り立つから

$$\begin{aligned} &{}_nC_k\, p^k(1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &\approx \frac{\sqrt{2\pi n} e^{-n} n^n\, p^k (1-p)^{n-k}}{\sqrt{2\pi k} e^{-k} k^k \sqrt{2\pi(n-k)} e^{-(n-k)} (n-k)^{n-k}} = \frac{\sqrt{n}}{\sqrt{2\pi}\sqrt{k(n-k)}} \left(\frac{np}{k}\right)^k \left(\frac{nq}{n-k}\right)^{n-k} \end{aligned}$$

となる. $\dfrac{k-np}{\sqrt{npq}} = t \quad (q=1-p)$ とおくと, $k = np + \sqrt{npq}t$, $n-k = n(1-p) - \sqrt{npq}t = nq - \sqrt{npq}t$ となるから,

$$\begin{aligned} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} &\approx \frac{\sqrt{n}}{\sqrt{2\pi}\sqrt{k(n-k)}} \cdot \left(\frac{np}{np+\sqrt{npq}t}\right)^k \left(\frac{nq}{nq-\sqrt{npq}t}\right)^{n-k} \\ &= \frac{\sqrt{n}}{\sqrt{2\pi}\sqrt{k(n-k)}} \cdot \left(\frac{1}{1+\frac{\sqrt{q}t}{\sqrt{np}}}\right)^k \left(\frac{1}{1-\frac{\sqrt{p}t}{\sqrt{nq}}}\right)^{n-k} \\ &= \frac{\sqrt{n}}{\sqrt{2\pi}\sqrt{k(n-k)}} \cdot \left(1+\frac{\sqrt{q}t}{\sqrt{np}}\right)^{-k} \left(1-\frac{\sqrt{p}t}{\sqrt{nq}}\right)^{k-n} \end{aligned}$$

が成り立つ. 一方,

$$\begin{aligned} &\log\left[\left(1+\frac{\sqrt{q}t}{\sqrt{np}}\right)^{-k} \left(1-\frac{\sqrt{p}t}{\sqrt{nq}}\right)^{k-n}\right] = -k\log\left(1+\frac{\sqrt{q}t}{\sqrt{np}}\right) + (k-n)\log\left(1-\frac{\sqrt{p}t}{\sqrt{nq}}\right) \\ &= (-np-\sqrt{npq}t)\log\left(1+\frac{\sqrt{q}t}{\sqrt{np}}\right) + (-nq+\sqrt{npq}t)\log\left(1-\frac{\sqrt{p}t}{\sqrt{nq}}\right) =: \text{(A)} \end{aligned}$$

で, $\log x$ のマクローリン展開

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots \quad (|x|<1)$$

より,

$$
\begin{aligned}
\text{(A)} &= (-np-\sqrt{npq}t)\cdot\left(\frac{\sqrt{q}t}{\sqrt{np}}-\frac{qt^2}{2np}+\frac{1}{3}(\frac{\sqrt{q}t}{\sqrt{np}})^3+\cdots\right)\\
&\quad+(-nq+\sqrt{npq}t)\cdot\left(-\frac{\sqrt{p}t}{\sqrt{nq}}-\frac{pt^2}{2nq}-\frac{1}{3}(\frac{\sqrt{p}t}{\sqrt{nq}})^3+\cdots\right)\\
&=(-\sqrt{np}-\sqrt{q}t)\cdot\left(\sqrt{q}t-\frac{qt^2}{2\sqrt{np}}+\frac{1}{3}\frac{(\sqrt{q}t)^3}{np}+\cdots\right)\\
&\quad+(-\sqrt{nq}+\sqrt{p}t)\cdot\left(-\sqrt{p}t-\frac{pt^2}{2\sqrt{nq}}-\frac{1}{3}\frac{(\sqrt{p}t)^3}{nq}+\cdots\right)\\
&=-\sqrt{npq}t-qt^2+\frac{qt^2}{2}+\frac{(\sqrt{q}t)^3}{2\sqrt{np}}-\frac{(\sqrt{q}t)^3}{3\sqrt{np}}\\
&\quad+\sqrt{npq}t-pt^2+\frac{pt^2}{2}-\frac{(\sqrt{p}t)^3}{2\sqrt{nq}}+\frac{(\sqrt{p}t)^3}{3\sqrt{nq}}+\cdots\\
&=-\frac{(p+q)t^2}{2}+\frac{(\sqrt{q}t)^3}{6\sqrt{np}}-\frac{(\sqrt{p}t)^3}{6\sqrt{nq}}+\cdots\\
&=-\frac{t^2}{2}+O(\frac{1}{\sqrt{n}})\quad(n\to\infty)\qquad(O(\tfrac{1}{\sqrt{n}})\text{ とは, }n\to\infty\text{ のとき }\frac{1}{\sqrt{n}}\text{ と同程度ということ})
\end{aligned}
$$

となる. したがって

$$
\left(1+\frac{\sqrt{q}t}{\sqrt{np}}\right)^{-k}\left(1-\frac{\sqrt{p}t}{\sqrt{nq}}\right)^{k-n}\approx e^{-\frac{t^2}{2}}
$$

が成り立つ. また,

$$
\begin{aligned}
\frac{\sqrt{n}}{\sqrt{k(n-k)}} &= \frac{\sqrt{n}}{\sqrt{(np+\sqrt{npq}t)(nq-\sqrt{npq}t)}}=\frac{1}{\sqrt{(\sqrt{n}p+\sqrt{pq}t)(\sqrt{n}q-\sqrt{pq}t)}}\\
&=\frac{1}{\sqrt{npq+t\sqrt{npq}(q-p)-pqt^2}}=\frac{1}{\sqrt{npq\left\{1-\left[\frac{t}{\sqrt{npq}}(p-q)+\frac{t^2}{n}\right]\right\}}}\\
&=\frac{1}{\sqrt{npq}}\left\{1+\left[\frac{t}{\sqrt{npq}}(p-q)+\frac{t^2}{n}\right]+\left[\frac{t}{\sqrt{npq}}(p-q)+\frac{t^2}{n}\right]^2+\cdots\right\}\\
&\approx\frac{1}{\sqrt{npq}}\quad(n\to\infty)
\end{aligned}
$$

となるから,

$$
P(X=k)\approx\frac{1}{\sqrt{2\pi npq}}e^{-\frac{t^2}{2}}\quad(n\to\infty)
$$

となる. 左辺は k の関数で, 右辺は t の関数であるが, $\dfrac{k-np}{\sqrt{npq}}=t$ であったので, $\dfrac{dk}{dt}=\sqrt{npq}$ より右辺 (の積分) は

$$
\int\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt
$$

となるので, 密度関数は $\phi(t)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$, すなわち標準正規分布となった.