黒木玄 Gen Kuroki
@genkuroki

二項分布などの正規分布での近似はPoisson分布の正規分布による近似を経由すると楽になる。

1. Poisson分布を正規分布で近似

$k = np + \sqrt{n}\,x = np(1+z)$, $z = x/(\sqrt{n}\,p)$ とおくと、$n\to\infty$ のとき、

$$
\begin{aligned}
&np((1+z)\log(1+z) - z) = \frac{x^2}{2p} + o(1),\\
&(1+z)^{np(1+z)}e^{-npz} = e^{x^2/(2p)}(1+o(1)),\\
&k! \approx k^k e^{-k}\sqrt{2\pi k}\\
&\ = (np)^k e^{-np}(1+z)^{np(1+z)}e^{-npz}\sqrt{2\pi np(1+z)}\\
&\ \approx (np)^k e^{-np} e^{x^2/(2p)}\sqrt{2\pi np},\\
&e^{-np}\frac{(np)^k}{k!} \approx \frac{e^{-x^2/(2p)}}{\sqrt{2\pi np}}.
\end{aligned}
$$

2017年06月02日 19:11 · Web · 0 · 1 · Webで開く

黒木玄 Gen Kuroki @genkuroki Friday at 7:17pm

2. 二項分布を正規分布で近似

$p, q > 0, p + q = 1, k + l = n,$

$$
\begin{aligned}
k &= np + \sqrt{n}\,x,\\
l &= nq + \sqrt{n}\,y,
\end{aligned}
$$

とおくと、$x + y = 0$ となり、Poisson分布の場合の結果より、

$$
\begin{aligned}
&\binom{n}{k} p^k q^l\,dk\\
&= \frac{n!}{n^n e^{-n}}\,e^{-np}\frac{(np)^k}{k!}\,e^{-nq}\frac{(nq)^l}{l!}\sqrt{n}\,dx\\
&\approx \sqrt{2\pi n}\frac{e^{-x^2/(2p)}}{\sqrt{2\pi np}}\,\frac{e^{-y^2/(2q)}}{\sqrt{2\pi nq}}\sqrt{n}\,dx\\
&= \frac{e^{-x^2/(2pq)}}{\sqrt{2\pi pq}}dx.
\end{aligned}
$$

最後の等号で $p+q=1, x+y=0$ を使った。

黒木玄 **Gen Kuroki** @genkuroki Friday at 7:18pm

3. 多項分布の場合

$p_i > 0, p_1 + \cdots + p_r = 1, k_1 + \cdots + k_r = n,$

$$k_i = np_i + \sqrt{n}\, x_i$$

とおくと、

$$\begin{aligned}
&\frac{n!\prod_{i=1}^r p_i^{k_i}}{\prod_{i=1}^r k_i!} dk_1 \cdots dk_{r-1} \\
&\approx \sqrt{2\pi n} \prod_{i=1}^r \frac{e^{-x_i^2/(2p_i)}}{\sqrt{2\pi n p_i}} \prod_{j=1}^{r-1} (\sqrt{n}\, dx_j) \\
&= \frac{\exp\left[-\frac{1}{2}\sum_{i=1}^r x_i^2/p_i\right]}{\sqrt{(2\pi)^{r-1} p_1 \cdots p_r}} dx_1 \cdots dx_{r-1}
\end{aligned}$$

続く

黒木玄 **Gen Kuroki** @genkuroki Friday at 7:18pm

続き

指数函数の中身にPearsonのカイ二乗統計量

$$\sum_{i=1}^r \frac{x_i^2}{p_i} = \sum_{i=1}^r \frac{(k_i - np_i)^2}{np_i}$$

が見えているのは、Poisson分布を経由した計算のおかげである。

黒木玄 **Gen Kuroki** @genkuroki Friday at 9:06pm

二項分布の正規分布による近似をスターリングの公式を使って出すときのコツは、

(1) 最初から頑張って標準正規分布で近似しようとしないこと

(2) 二項係数の分母の $k!$ と $l! = (n-k)!$ を最初のうちはあたかも独立変数であるかのように扱って、Poisson分布の中心極限定理の結果を使うこと。$k+l=n$ は最後に使えばよい。