頻度分布の分布

M 個の乱数を N 個の bin を使って頻度分布を計算する場合の、その各 bin の高さの分布を考える。ある bin に k 個の乱数が入る確率は、 $p=1/N$ として次式で表される。

$$P(k)=\binom{M}{k}p^k(1-p)^{M-k}$$

乱数の総数 M が十分に大きく、各 bin への頻度の平均 pM も十分に大きいと仮定する。 k が平均 pM に近いときの挙動を考える。

大きな整数 n の階乗に対して、Stirling の公式が知られている。

$$\ln(n!)\simeq n\ln n-n$$

これを二項分布に使う。

$$\begin{aligned}\ln P(k)&=\ln\binom{M}{k}+k\ln p+(M-k)\ln(1-p)\\&\simeq+M\ln M-k\ln k(M-k)\ln(M-k)+k\ln p+(M-k)\ln(1-p)\end{aligned}$$

変数を x に変更する。

$$x=\frac{k-pM}{M}\ll 1$$

$$\begin{aligned}\ln p(x)\simeq&+M\ln M-M(p+x)\ln(p+x)M-M(1-p-x)\ln(1-p-x)M\\&+M(p+x)\ln p+M(1-p-x)\ln(1-p)\\=&M(p+x)[\ln p-\ln(p+x)]+M(1-p-x)[\ln(1-p)-\ln(1-p-x)]\end{aligned}$$

x で展開することで正規分布であることがわかる。

$$\begin{aligned}\ln p(x)\simeq&M(p+x)\left[-\frac{x}{p}+\frac{1}{2}\frac{x^2}{p^2}+O(x^3)\right]\\&+M(1-p-x)\left[\frac{x}{1-p}+\frac{1}{2}\frac{x^2}{(1-p)^2}O(x^3)\right]\\=&-\frac{M}{2}\frac{1}{p(1-p)}x^2+O(x^3)\\=&-\frac{(k-pM)^2}{2\sigma^2}+O\left(\left(\frac{k-pM}{M}\right)^3\right)\end{aligned}$$