＜多項分布＞

例 6　＜多項分布＞

二項分布を多次元に一般化したのが多項分布である。二項分布のコイン投げをサイコロ投げに変えたと考えれば良い。いま k 個の面をもつ仮想のサイコロを考える。第 i 番目の面の出る確率を p_i とする。$p_i > 0$，$p_1+p_2+\cdots+p_k=1$ である。このサイコロを N 回投げたときに i 番目の面が出た回数を X_i とおく。このとき

$$P(X_1 = n_1,\ X_2 = n_2,\ \cdots,\ X_k = n_k)=\frac{N!}{n_1!n_2!\cdots n_k!}p_1^{n_1}p_2^{n_2}\cdots {p_k}^{nk}$$

となる。ただし $n_1+n_2+\cdots+n_k=N$ である。この分布を**多項分布** $\boldsymbol{M(N,\ (p_i))}$ という。多項分布と呼ばれるのは，多項展開式

$$(p_1+p_2+\cdots+p_k)^N = \sum_{n_1+n_2+\cdots+n_k=N} \frac{N!}{n_1!n_2!\cdots n_k!}\ p_1^{n_1}p_2^{n_2}\cdots {p_k}^{n_k}$$

の各項を確率としているからである。

$(X_1,\ X_2,\ \cdots,\ X_k)$ の分布は k 次元分布であり，各 i に対し X_i の分布は 1 次元分布である。この 1 次元分布を多次元分布の**周辺分布**という。X_i の分布は二項分布 $B(N,\ p_i)$ であるから，その平均と分散は

$$E[X_i] = Np_i, \qquad V[X_i] = E[(X_i - E[X_i])^2] = Np_i(1-p_i)$$

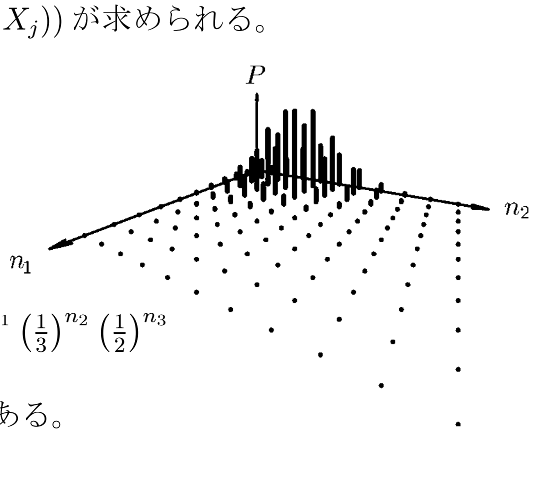
である。また $X_i + X_j$ は二項分布 $B(N,\ p_i+p_j)$ に従うから，

$$E[X_i + X_j] = N(p_i+p_j), \qquad V[X_i + X_j] = N(p_i+p_j)(1-p_i-p_j)$$

である。さらに共分散 $C_{ov}(X,\ Y)=E[(X-E[X])(Y-E[Y])]$ は

$$C_{ov}(X_i,\ X_j) = \frac{1}{2}\{V(X_i+X_j) - V(X_i) - V(X_j)\} = -Np_ip_j$$

となる。これによって分散共分散行列 $(C_{ov}(X_i,\ X_j))$ が求められる。ただし $C_{ov}(X_i,\ X_i)=V(X_i)$ である。



右図は $k=3$，$p_1=\dfrac{1}{6}$，$p_2=\dfrac{1}{3}$，$p_3=\dfrac{1}{2}$，$N=10$ のとき確率

$$P(X_1 = n_1,\ X_2 = n_2,\ X_3 = n_3)=\frac{10!}{n_1!n_2!n_3!}\left(\tfrac{1}{6}\right)^{n_1}\left(\tfrac{1}{3}\right)^{n_2}\left(\tfrac{1}{2}\right)^{n_3}$$

を $(n_1,\ n_2)$ 平面上の棒の高さで表現したものである。ただし $n_3 = 10 - n_1 - n_2$ である。