２項分布，基本的な分布

1　場合の数

集合の要素を数え上げるとき，つぎの形がよく用いられる。正の整数 $m \geq 1$ で $m! = m(m-1)(m-2)\cdots 3{\cdot}2{\cdot}1$, $0! = 1$ とし，正の整数 n の場合には $\binom{n}{m} = \dfrac{n!}{m!(n-m)!} = {}_nC_m = C_m^n$ などと表す。このときには，n 個の中から m 個を取り出す組み合わせの数で，$(1+x)^n = \overbrace{(1+x)(1+x)\cdots(1+x)}^{n} = 1 + x + {}_nC_2x^2 + {}_nC_3x^3 + {}_nC_mx^m + \cdots$ が成り立ち，べき乗で”1” を $n-m$ 個，”x”を m 個取り出す，すなわち x^m の係数に等しい。これを拡張して正の整数 m について $\binom{\alpha}{m} = \dfrac{\overbrace{\alpha(\alpha-1)\cdots(\alpha-m+1)}^{m}}{m!}$ ここで α は正の整数とは限らなくてもよい。

2　コインとサイコロ

コインを投げると，表か裏のいずれか，2 通りのうちのひとつが結果としておこる。結果を表す変数を確率変数といい，この変数がとり得る値の集まりを考え，その起こりえる可能性を尺度化することで，確率分布が定められる。コイン投げの 2 通りの結果がある場合をベルヌーイ分布といい，サイコロ投げのように 6 通りの目の出方が同じ確からしさをもつ場合，一様分布とよぶ。

3　２項分布

表の出る確率が p のコインを一枚投げる実験では，$X = \begin{cases} 1, & p \\ 0, & 1-p \end{cases}$ で，これを n 枚投げる（繰り返す）ことは $Y_n = X_1 + X_2 + \cdots + X_n$ とおき，$k(k = 0, 1, 2, \cdots, n)$ 枚表が出る事象の確率 $P(Y_n = k)$ は $X_i(i = 1, 2, \cdots, n)$ のなかで 1 となっているものの数え上げであるから 2 項係数で表現できる。

$$P(Y_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \cdots, n$$

これをパラメータ n, p の 2 項分布という。2 項分布の平均 μ と分散 σ^2 は

$$\mu = np, \quad \sigma^2 = np(1-p)$$

とくに $n = 1$ はパラメータ p のベルヌイ分布とよばれる。ベルヌイ分布や 2 項分布から生じるさまざまな現象は，ランダムウオーク（酔歩），正規分布やポアソン分布などいろいろな分布の解析に結びつく。

4　ポアソン分布

実数のパラメータ $\lambda > 0$ をもつポアソン分布とは

$$P(X = k) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad k = 0, 1, 2, \cdots$$

とする。2 項分布から派生される分布のひとつで，稀な（起こりにくい）現象を表す。パラメータ n, p の 2 項分布において， n を大きくし，p をゼロに近づけ，ただしその積が一定の値 λ であるようなものが，ポアソン分布である。

ポアソン分布の平均 μ と分散 σ^2 は

$$\mu = \lambda, \quad \sigma^2 = \lambda$$

一般に 2 項分布の確率計算は n が大きいとやっかいであるが，もし p がそう大きくなければ，ポアソン分布で近似できる。条件 $n > 50, np \leq 5, n(1-p) \leq 5$ を満たせば，かなり正確な値で求められる。

5　一様分布（離散型と連続型）

実験，結果のとり得る値がいくつかあることとき，これらが同じ確かしさをもつときには，一様分布にしたがう確率変数で結果が表される。とり得る可能な値，結果の集合が n 点の点集合のときには，離散型といい，ある有限な区間内の実数の点であれば，連続型とよばれる。点集合 $\{1, 2, \cdots, n\}$ 上の離散型密度関数は

$$P(X = k) = \frac{1}{n}, \quad k = 1, 2, \cdots, n$$

この平均は $\dfrac{n+1}{2}$, 分散は $\dfrac{n^2-1}{12}$ である。一方，区間 $[a, b]$ 上の一様分布とは

$$f_X(x) = \frac{1}{b-a}, \quad a \leq x \leq b$$

矩形分布とのよばれ，平均は $\dfrac{b-a}{2}$, 分散は $\dfrac{(b-a)^2}{12}$ である。

6　幾何分布

コイン投げを何回か繰り返すと，表（成功）○ と裏（失敗）● がそれぞれ何回か続けて起こるときがある。このように 2 通りの結果をもつ実験で，その成功が続けておこる回数を調べる。○ の確率が p, ● の確率が $1-p$ であれば，$\overbrace{\bullet\bullet\cdots\bullet}^{k-1}\overset{k}{\circ}$ の確率は $p(1-p)^{k-1}$ であるから，確率変数 X を始めて成功するまでに要した回数とすると

$$P(X = k) = p(1-p)^{k-1}, \quad k = 1, 2, \cdots$$

で，これをパラメータ p の幾何分布とよぶ。

7　超幾何分布

つぼの中からボールを取り出すとき，取り出されたボールを元のつぼに戻すかどうかで次に取り出すボールの結果に影響を与える。復元抽出と非復元抽出である。ボールの総個数が有限個しかないならば，復元抽出は同じ状況の繰り返しであるが，非復元の場合には取り出しの度毎に変わっていく。条件付きの確率を考えることになる。ボールの総個数を N とし，2 種類のボール，たとえば赤 r と黒 $N-r$ のボールがあるとき n 個のボールを非復元抽出するとき，この中に赤ボールが X 個含まれる，つまり黒ボールが $n-X$ 個となる確率，

$$P(X = k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \cdots, r$$

をパラメータ N, n, r の超幾何分布という。ボールの総数 N がかなり大きいときには，ボールをもとに戻してもつぎのボールの取出しにはほとんど影響ない。すなわち，同じ状況の繰り返しであるから，これはパラメータ n, p の 2 項分布になる。復元抽出（繰り返しのばあい）には 2 項分布であるが，非復元抽出（もとに戻さない取り出し）では，超幾何分布である。ここで $\lim_N r/N = p, \lim_N (N-r)/N = 1-p$ それぞれ赤ボール，黒ボールの比率を表す。上の超幾何分布の確率は $N \to \infty$ とすれば，2 項分布の確率に近づく。

8　指数分布

連続型分布のひとつに，放射性粒子の崩壊時刻，ある部品が壊れるまでの寿命分布，新しい客が到着するまでの待ち時間分布はどはつぎの指数分布にしたがうとされる。重要な性質（メモリーレスとよばれる）として

$P(X > a+b|X > a) = P(X > b)$, $a, b \geq 0$. この性質は時刻 a までは故障しなかった条件での，時刻 $a+b$ まで故障しない（条件付き）確率は，はじめから b まで故障しない確率に等しいことを意味する。パラメータ $\lambda > 0$ の指数密度とは

$$f_X(x) = \lambda e^{-\lambda x},\ x \geq 0$$

平均は $\dfrac{1}{\lambda}$, 分散は $\dfrac{1}{\lambda^2}$ このメモリーレスという性質は前に述べた幾何分布にも当てはまり，幾何分布が離散時間パラメータにおけるばあいで, 指数分布が連続時間パラメータのばあいである。

9　ガンマ分布と負の 2 項分布

連続時間パラメータでの指数分布，離散時間パラメータでの幾何分布にそれぞれ対応して，ガンマ分布，負の 2 項分布が知られている。両者ともはじめて起こるという事象の回数を繰り返し，複数回にしたばあいである。すなわち，指数分布の和がガンマ分布，幾何分布の和が負の 2 項分布にしたがう。

練習問題

$\boxed{1}$ (1) 拡張された 2 項係数で $\dbinom{-1}{n}, n = 1, 2, \cdots$ を計算しなさい。(2) また $\dbinom{2n}{n}(\frac{1}{2})^{2n} = (-1)^n \dbinom{-1/2}{n}, n = 1, 2, \cdots$ を示しなさい。

$\boxed{2}$ パラメータ p, n の 2 項分布にしたがう X の平均と分散の計算をつぎの方法でおこなえ。（1）平均については直接 $E[X]$ を計算する。（2） $E[X(X-1)]$ を求めてから, 分散 $V(X) = E(X^2) - (EX)^2$ を計算する。

$\boxed{3}$ パラメータ $n = 10, p = 0.2$ の 2 項分布 X における確率が

k	0	1	2	3	4	5	6	7	8	$\cdots$
$P(X = k)$	0.107	0.268	0.302	0.201	0.088	0.026	0.006	0.001	0.000	0.000

となることを計算し，ポアソン分布と比較せよ。

（答え）パラメータ $\lambda = 2$ のポアソン分布 Y は

k	0	1	2	3	4	5	6	7	8	$\cdots$
$P(Y = k)$	0.135	0.271	0.271	0.180	0.090	0.036	0.012	0.003	0.001	0.000

$\boxed{4}$ パラメータ n, p の 2 項分布とパラメータ m, p の 2 項分布との和はパラメータ $n+m, p$ の 2 項分布になることを示しなさい。

$\boxed{5}$ 指数分布の平均と分散を計算しなさい。また $P(X > a+b\,|\,X > a) = P(X > b)$, $a, b \geq 0$ が成り立つことを確かめなさい。