講義ノート
経済学のための確率論入門

増山　幸一
明治学院大学経済学部

2006 年 11 月

始めに

近年、経済学の主要な専門分野、例えば、情報の経済学あるいは金融経済学などでは、モデル分析に際して、確率論の知識が必要不可欠となってきている．この講義ノートは、確率論あるいは統計学の基礎知識を持たない学部学生を対象に、確率論の基礎知識をコンパクトな形で導入することを目的としている．

第 1 節で、確率概念を定義し、確率空間の概念を数学的に導入している．条件付確率と独立性の概念も例を用いて詳しく説明している．第 2、3 節で、確率変数の概念を導入して、確率変数を用いてランダム現象を定式化する方法を導入している．第 2 節では、確率変数の値域が加算個の集合であるような場合、つまり確率変数が離散的なケースを取り上げている．第 3 節では、確率変数の値が連続的に変化するケースを取り扱っている．確率計算で必須の概念である期待値 (平均値) と分散の概念を定義し、それらの計算方法を説明している．第 3 節で、母関数と特性関数という概念を導入している．母関数と特性関数は確率分布を見出し、期待値や分散などを計算する際には非常に有用な道具となることを説明している．

1　確率と確率空間

1.1　確率と確率空間

最初に例を用いて確率と確率空間について説明する[1]．

例 1.1
箱の中に s 個のボールがあり、各ボールは番号が $1, 2, 3, ...$ と付いているとする．さらに、最初の r 個のボールは赤で塗られ、残りのボールは黒で塗られているとする．ボールは良くかき混ぜられている．この箱からボールを 1 個ランダムに取り出し、取り出したボールの番号を紙に書いて、もとの箱にボールを戻すことを繰り返す．

例 1.1 の試行を n 回繰り返すとき、k という数字が書かれたボールが取り出された回数を $N_n(k)$ と表記しよう．例えば、$s = 3$、$n = 20$ のケースを考えると、その結果は

$$\{1, 1, 3, 2, 1, 2, 2, 3, 2, 3, 1, 2, 1, 2, 3, 3, 1, 3, 2, 2\}$$

[1]確率論の最も基本的な教科書は古典とも言えるフェラーの入門書である．William Feller(1957), *An Introduction to Probability Theory and its Applications*, vol.1(河田監訳『確率論とその応用 I(上・下)』紀伊国屋書店．

のようになるときがある．この場合、

$$N_{20}(1)=6, N_{20}(2)=8, N_{20}(3)=6$$

である．したがって、番号 $1,2,3$ が取り出された相対頻度は

$$\frac{N_{20}(1)}{20}=0.30, \frac{N_{20}(2)}{20}=0.40, \frac{N_{20}(3)}{20}=0.30$$

である．試行回数 n が無限に多くなるとき、これらの相対頻度はある一定の値に収束すると予想される．相対頻度

$$\frac{N_n(k)}{n}\ (k=1,2,3,...,s)$$

の極限値を p_k と表記することにしよう．直観的にも、

$$p_i=\frac{1}{s},\ i=1,2,...,s$$

となることは明らかである． 通常、この相対頻度の極限値 $p_i(i=1,2,3,...,s)$ が、ボールをランダムに取り出すときに番号 i のボールが取り出される確率である．

例 1.1 の試行実験に数学的定式化を与えるにはどうすれば良いでしょうか．このためには、まず、箱の中に入っているボールの番号に一対一に対応した集合を考える必要がある．つまり、抽象的には、s 個の要素を持つ集合を対応させればよい．この集合を Ω と表記しよう．ボールの番号 k に対応する集合 Ω の要素を ω_k と表記する．$\Omega=\{\omega_1,\omega_2,...,\omega_s\}$　である．だから、集合 Ω から要素 ω_k が選択される確率は $p_k=\frac{1}{s}$ ということになる．明らかに、

$$1\geq p_k\geq 0,$$

$$p_1+p_2+\cdots+p_s=1,$$

が成立している．

次に、n 回の試行のうちで赤く塗られたボールが取り出される確率について考えよう．ボールの番号が 1 から r までが赤く塗られているから、合計で赤いボールが r 個あり、残りの $s-r$ 個が黒いボールである．赤いボールが取り出される相対頻度は、番号が 1 から r までのいずれかのボールが取り出される相対頻度だから、

$$\frac{N_n(1)}{n}+\frac{N_n(2)}{n}+...+\frac{N_n(r)}{n}$$

となる．この相対頻度は n が無限大に近づくと、ある一定値に収束する．これが赤く塗られたボールが取り出される確率に他ならない．明らかに、この確率は $\frac{r}{s}$ である．

赤いボールが取り出される確率を数学的に定式化しよう．赤いボールは集合 Ω に属する要素の中で、番号 1 から番号 r に対応するすべての要素から構成される．これを、例えば、集合 A としよう．$A=\{\omega_1,\omega_2,...,\omega_r\}$ である．集合 A は集合 Ω の部分集合であり、r 個の要素を持つ．集合 A は一つの「事象」(赤いボールが取り出される事象) に対応している．赤いボールが取り出される確率は事象 A が起こる確率として定式化されることになる．より一般的に言えば、集合 Ω は今考えているランダム現象の全体集合であり、集合 Ω のすべての部分集合はそれぞれある事象を表現することになる．

A と B を全体集合の中のある二つの部分集合（事象）であるとする．このとき、A と B の和集合と共通集合で表現される事象の確率を考える必要が生じる．例えば、集合 A を偶数の番号のボールが取り出される事象、B を赤いボールが取り出される事象とする．番号が偶数で赤いボールが取り出される事象は、事象 A と事象 B の共通集合であるから、集合 $A\cap B$ で表現される．偶数番号のボールもしくは赤いボールが

取り出される事象は集合 A と B の和集合だから、$A \cup B$ となる．番号は偶数または奇数のいずれかであるから、奇数番号のボールが取り出される事象 C は、集合 A の補集合 A^c となる．だから、奇数番号のボールが取り出される確率は事象 A^c が起きる確率である．

例 1.1 のように箱からボールをランダムに 1 個取り出すとき、任意の番号 $i(i = 1, 2, ...s)$ のボールが取り出される確率は $1/s$ であった．つまり、集合 Ω の要素一つからなる集合 (事象)$D = \{\omega_i\}$ が生起する確率は $1/s$ である．だから、A を赤いボールが取り出される事象とするならば、A の確率は A の要素の数 ÷ 全数 $=r/s$ になることは明らかである．事象 A の確率を $P(A)$ と表現するので、$P(A) = r/s$ である．言い換えると、事象 B が j 個の要素を有している集合のとき、事象 B が生起する確率は $P(B) = j/s$ となる．一般的に言えば、

$$P(B) = \sum_{\omega_k \in B} p_k$$

が成立している．実は、この関係は事象 B の各要素の生起確率が独立であるという事実から成立する．一般に、以下の関係式が成立する．

$$P(\emptyset) = 0, P(\Omega) = 1.$$

もし集合 A と B が共通部分を持たなければ、

$$P(A \cap B) = 0,\ P(A \cup B) = P(A) + P(B) \tag{1}$$

である．

今まで、集合 Ω は離散的であるケースを例で取り上げてきたが、Ω が可算無限あるいは非可算無限の場合でも上記の性質は成立する．ただこの場合、Ω の各要素の生起確率はゼロとなる．

上の例で行ったように、考えているランダム現象に対応した全体集合を Ω とする．この全体集合は数学的にはある任意の集合である．全体集合 Ω の部分集合 A はある事象を意味する．事象 A が生起することは、試行の結果、A に属するいずれかの要素が起きたことと同じことである．Ω のすべての部分集合の集まりを $\mathscr{F}$ とすると、$\mathscr{F}$ は事象の集まりを表現する．集合の集まりを集合族という．集合族 $\mathscr{F}$ に属するすべての集合はある事象を表現している．言い換える、集合族 $\mathscr{F}$ は試行で考えられるすべての事象の集合である．上で行ったように、集合族 $\mathscr{F}$ に属するすべての集合 (事象) に確率を付与することができる．集合族 $\mathscr{F}$ に属する集合 A の確率を $P(A)$ と表現する．このようにして、考察しようとしているランダム現象に対して、$(\Omega, \mathscr{F}, P)$ という組を定義することができる．これを確率空間 (probability space) という．

確率計算上で数学的な整合性を持つためには、この集合族は確率の定義域として望ましい条件を満たさなければいけない．正確には、集合族 $\mathscr{F}$ は以下のように定義される σ 代数族でなければいけない．集合 Ω の部分集合の集まりである集合族 $\mathscr{F}$ が、以下の性質 (1)、(2) を満たすとき、σ 代数族 (σ-algebra) であると言われる[2]．

(1)．$A \in \mathscr{F} \Rightarrow A^c \in \mathscr{F}$

(2)．$A_n \in \mathscr{F}, n = 1, 2, ... \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathscr{F}$

確率 P はこの σ 代数族 (ボレル集合族) に対して定義されるもので、以下の性質を持つものである[3]．

(i)．$P(\Omega) = 1$,

[2] σ 代数族およびボレル集合族についての詳細な説明は、H.L.Royden(1968), *Real Analysis*,Macmillan. を参照ください．σ 代数族は完全加法族とも言われる．

[3] 実数値空間におけるすべての開集合を含む最小の σ 代数族をボレル集合族という．

(ii). すべての $A \in \mathscr{F}$ に対して, $P(A) \geq 0$,

(iii). もし $A_n, n = 1, 2, 3, ...$ が互いに共通部分を持たないならば、

$$P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n).$$

したがって、確率 P(probability measure) は、σ 代数族 (ボレル集合族)$\mathscr{F}$ を定義域とし、非負の実数値を値域とする集合関数である．以上の定義から、確率は以下のような性質を有することが分かる．
$A \cup A^c = \Omega$、$B = \Omega \cap B$ だから、

$$B = (A \cup A^c) \cap B = (A \cap B) \cup (A^c \cap B)$$

である．よって、

$$P(B) = P(A \cap B) + P(A^c \cap B). \tag{2}$$

$B = \Omega$ と置けば、$P(A^c) = 1 - P(A)$ ．$B = \emptyset$ とすれば、$P(\emptyset) = 0$．が成立する．また、A と B が共通部分を持つときは一般的に、

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{3}$$

が成り立つことも容易に分かる．

1.2 条件付確率

A と B を二つの事象とする．ただし、$P(A) > 0$ である．このとき、A が起きたとき B が生起する確率を、A を与えたときの B の条件付確率 (conditional probability) といい、$P(B|A)$ と表現する．条件付確率 $P(B|A)$ は、

$$P(B|A) = \frac{P(B \cap A)}{P(A)} \tag{4}$$

と定義される．$P(A) = 0$ の場合は、条件付確率は定義できないが、$P(B|A) = 0$ とする．

例 1.2
ある大学の学生のうち 40 %が男性で、60 %が女性である．このうち近視の学生は男性の 50 %、女性の 30 %である．近視者が男性である確率を求めなさい．

全体集合 Ω は、この大学の学生すべてを要素とする *(*各学生と一対一に対応した要素からなる*)* 集合である．男性が取り出されるという事象を M と表記し、女性が取り出される事象を F とする．近視者が取り出される事象を S、非近視者が取り出される事象を N とする．例のデータから、$P(M) = 0.4, P(F) = 0.6$、M を与えるときの S の条件付確率 $P(S|M) = 0.5$、同様に $P(S|F) = 0.3$ である．条件付確率の定義から、

$$P(M|S) = \frac{P(M \cap S)}{P(S)}$$

である．$P(M \cap S)$ と $P(S)$ の値を計算する必要がある．$P(S|M) = P(S \cap M)/P(M)$ から

$$P(M \cap S) = P(S|M)P(M) = 0.5 \cdot 0.4 = 0.20$$

また、$S=S\cap M+S\cap F$ だから、

$$P(S)=P(S\cap M)+P(S\cap F).$$

$P(S\cap F)$ を計算すると、

$$P(S\cap F)=P(S|F)P(F)=0.3\cdot 0.6=0.18$$

である．したがって、

$$P(S)=0.20+0.18=0.38.$$

それ故、

$$P(M|S)=\frac{0.20}{0.38}\approx 0.53$$

となる．

例 1.3 (ポリアの壺 Polya's urn)
壺に赤いボールが r 個、黒いボールが b 個入っている．壺からランダムにボールを取り出し、そのボールの色を記録する．取り出されたボールに同じ色のボールを $c>0$ 個加えて、壺に戻される．このランダムな取り出しを n 回行う．

取り出された j 番目のボールが赤である事象を $R_j(j=1,2,...,n)$ と表記し、j 番目に黒のボールが取り出される事象を $B_j(j=1,2,...,n)$ とする．当然、任意の j に対して、R_j と B_j は独立な事象である．k 回目の試行を行う時点で、壺の中には、$b+r+(k-1)c$ 個のボールが入っている．取り出し方は完全にランダムであるとすると、ある特定のボールが取り出される確率は

$$\frac{1}{b+r+(k-1)c}$$

である．最初に、確率 $P(R_1\cap R_2)$ を計算してみよう．条件付確率の定義から、

$$P(R_1\cap R_2)=P(R_1)P(R_2|R_1).$$

最初の試行で赤いボールが取り出される確率は

$$P(R_1)=\frac{r}{b+r}$$

１回目のボールが赤であるとき、２回目のボールが赤である条件付確率は

$$P(R_2|R_1)=\frac{r+c}{b+r+c}$$

と計算できる．したがって、

$$P(R_1\cap R_2)=(\frac{r}{b+r})(\frac{r+c}{b+r+c}).$$

同様にして $P(B_1\cap R_2)$ も計算できて、

$$P(B_1\cap R_2)=(\frac{b}{b+r})(\frac{r}{b+r+c}).$$

$R_2=R_1\cap R_2+B_1\cap R_2$ が成立するので、

$$P(R_2)=P(R_1\cap R_2)+P(B_1\cap R_2)=(\frac{r}{b+r})(\frac{r+c}{b+r+c})+(\frac{b}{b+r})(\frac{r}{b+r+c})=\frac{r}{b+r}.$$

それ故、$P(R_1)=P(R_2)$ である．同様の計算で、$P(B_1)=P(B_2)$ という関係も確かめられる．

ここで、条件付確率の計算に便利なベイズの計算式を説明する．$\Omega = \bigcup_{k=1}^{n} A_k$ であり、$A_1, A_2, ...A_n$ は互いに共通部分を持たないとする．さらに、B がある事象で、$P(B) > 0$ であるとする．このとき、

$$B = B \cap (\bigcup_{k=1}^{n} A_k) = \bigcup_{k=1}^{n} (B \cap A_k)$$

であるから、

$$P(B) = \sum_{k=1}^{n} P(B \cap A_k).$$

また、条件付確率の定義から

$$P(B \cap A_k) = P(A_k)P(B|A_k)$$

であるから、

$$P(B) = \sum_{k=1}^{n} P(A_k)P(B|A_k)$$

が成立する．よって、

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^{n} P(A_k)P(B|A_k)} \tag{5}$$

という関係式が成り立つ．これがベイズの規則（Bayes' rule）と呼ばれる関係式である．

1.3 独立性

サイコロを振って出る目の数を記録する．$\Omega = \{1, 2, 3, ..., 6\}$．サイコロが理想的に作られているとすれば、さいころの各目は $1/6$ の確率で生起する．$i \in \Omega$ に対して、$P(\{i\}) = 1/6$ である．A と B を二つの事象とする．例えば、

$$A = \{1, 2\}, B = \{1\}$$

とする．このとき、

$$P(A) = 1/3, P(B) = 1/6, P(A \cap B) = 1/6$$

である．だから、

$$P(B|A) = P(A \cap B)/P(A) = (1/6)/(1/3) = 1/2$$

となり、$P(B|A)$ は $P(B)$ よりも大きい．別の例として、$A = \{1, 2, 3, 4, 5\}, B = \{1, 2, 3, 4, 6\}$ としよう．この場合は、$P(A) = 5/6, P(B) = 5/6, P(A \cap B) = 4/6$ である．よって、$P(B|A) = (4/6)/(5/6) = 4/5$ となり、$P(B|A)$ は $P(B)$ よりも小さい．

上の例で見たとおり、A が起きたという情報は B が生起する確率に影響を与える．しかし、興味のあるケースは A が生起したという情報が B が生起する確率に全く影響を与えない場合である．例えば、$A = \{1, 2, 3\}, B = \{1, 2, 4, 5\}$ であるとしよう．このとき、$P(A) = 1/2, P(B) = 2/3, P(A \cap B) = 1/3$ である．A が起きたときの B の条件付確率は $P(B|A) = (1/3)/(1/2) = 2/3$ となり、$P(B)$ と同じ値となる．このようなケースでは、事象 B は事象 A と独立である (independent) という．

一般的に、A と B をある確率空間の二つの事象とする．$P(A) \neq 0$ とする．

$$P(B|A) = P(B)$$

が成立するとき、A と B は独立であるという．$P(B|A) = P(B \cap A)/P(A)$ であるから、A と B が独立ならば、

$$P(A \cap B) = P(A)P(B). \tag{6}$$

正確には、上の等式が成立するとき、そしてそのときにのみ、二つの事象 A と B は独立であると定義される．

サイコロを振って、1 または 6 が出たときを成功、それ以外の数が出たときを失敗と呼ぶことにする．このとき、成功する確率を p とすると、失敗する確率は $1-p$ である．サイコロを n 回振ったときのランダム現象について、確率の計算方法を考える．i 回目のサイコロ振りの結果が、その後のサイコロの目の出る確率に影響を与えないことは明らかである．サイコロを n 回振ったとき、成功と失敗の可能な組み合わせは 2^n 通りである．つまり、毎回の結果は成功か失敗の 2 通りしかないから、考えられる組合せは、2 通り× n 回である．この n 回のサイコロ振りの結果は、n 個の組

$$(x_1, x_2, ..., x_n), x_i = 0, 1$$

で表現できる．ただし、成功は数字の 1、失敗は数字の 0 に対応する．全体集合 Ω はこれらの n 個の組の可能な組すべてからなる．σ 代数 (ボレル集合族)$\mathscr{F}$ は Ω のすべての部分集合の集まりである．

いま、$A_i(i = 1, 2, ..., n)$ を、i 回目のサイコロ振りで成功の目が出る事象 ($x_i = 1$ となる事象) を表現する．最初の k 回で成功が起きる事象は、

$$(\underbrace{1, 1, ..., 1}_{k}, \underbrace{0, ..., 0}_{n-k})$$

で表現される．よって、

$$(1, 1, ..., 1, 0, ...0) = A_1 \cap \cdots \cap A_k \cap A_{k+1}^c \cap \cdots \cap A_n^c.$$

上の式の右辺にある各事象 $A_1, A_2, ...$ は互いに独立であるから、

$$P(\{(1, 1, ..., 1, 0, ..., 0)\}) = P(A_1) \cdots P(A_k)P(A_{k+1}^c) \cdots P(A_n^c) = p^k(1-p)^{n-k}$$

となる．この確率は、成功の回数が k である事象の確率であり、成功が何回目に生起したかには依存していない．

B_k を n 回の試行うち k 回が成功である事象とする．事象 B_k に属する各要素が生起する確率は上で導出した．n 個の中から k 個 (成功の回数) を取り出す取り出し方は、2 項係数

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

で与えられる．したがって、n 回の試行うち k 回が成功する確率は

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

と計算される．これは 2 項分布といわれる確率分布である．

2 離散型確率分布

2.1 確率密度関数

コインを投げて、表が出たら 1、裏が出たときは、-1 と記録する．表が出る確率が p、裏が出る確率が $1-p$ であるとする．コインを 3 回投げ、記録した数字を合計し、この合計を X の値とする．H が表、T が

裏を表すとき、以下のような結果が記録される[4].

ω	$X(\omega)$	$P\{\omega\}$
HHH	3	p^3
HHT	1	$p^2(1-p)$
HTH	1	$p^2(1-p)$
THH	1	$p^2(1-p)$
HTT	-1	$p(1-p)^2$
THT	-1	$p(1-p)^2$
TTH	-1	$p(1-p)^2$
TTT	-3	$(1-p)^3$

この例で、全体集合は 8 通りの要素 ω_i を持つ.

$$\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

X の値は Ω の各要素に対応して定まっている．ただ、X の値は離散的な値 $\{-1, 1, 3\}$ のみを取りうる．また、各 ω_i の生起確率 $P(\{\omega\})$ は表のように計算できる．さて、Ω の部分集合 (事象) を X の値を用いて定義することができる．例えば、事象 A を

$$A = \{\omega : X(\omega) = 1\}$$

と定義するとき、$A = \{HHT, HTH, THH\}$ である．A は全体集合 Ω の部分集合であり、$A \in \mathscr{F}$ であることは明らかである．事象 A の各要素の生起確率は独立であるから、事象 A の確率は $P(A) = 3p^2(1-p)$ と計算される．この例に見られる変数 X を確率変数 (random variable) という.

より一般化して、確率変数の定義をする．$(\Omega, \mathscr{F}, P)$ を確率空間とし、X を Ω の各要素に実数値を対応させる関数とする．関数 X の値域が有限個、もしくは可算無限個であるとき、X を離散的確率変数 (discrete random variable) という．関数 X の値域は $\{x_1, x_2, ..., x_n, ...\}$ という離散的な値を要素とする集合である．通常、離散的確率変数のとり得る値が整数の場合が多い．このような確率変数を整数値確率変数という．そして、当然のことながら、集合 $\{\omega : X(\omega) = x_i\}$ はボレル集合族 $\mathscr{F}$ に属する確率事象である．こうして確率変数の各値に対応する事象の存在を想定できるので、$\{X = x_i\}$ に対応する事象に確率を与えることができる．事象 $\{\omega : X(\omega) = x_i\}$ の確率を $P(X = x_i)$ と簡単化して表記する.

確率密度関数 (probability density function)f は以下のように定義される．確率変数が取る値 x に対して

$$f(x) = P(X = x)$$

を満たす関数 f を密度関数という．明らかに、この密度関数の値域は $[0, 1]$ である．つまり、

(i). $f(x) \geq 0, x \in R,$ ただし、R は実数の集合を表現する．そして、

(ii). $\{x : f(x) \neq 0\}$ は R の有界個のもしくは可算無限個の要素を持つ部分集合である．この有限個もしくは可算無限個の要素からなる部分集合を $\{x_1, x_2, ...\}$ とするとき、さらに、以下の関係が成り立つことも確認できる.

(iii). $\sum_i f(x_i) = 1$.

[4]以下の例は、Paul G. Hoel, Sidney C. Port, and Charles J. Stone(1971), *Introduction to Probability Theory*, Houghton Mifflin　からの引用である.

確率密度関数は以上の 3 つの条件 (i),(ii),(iii) を満たさなければならない．

例 2.1 (2 項分布 binomial distribution)
コインを投げたときのように、各試行が成功と失敗の 2 種類のみに結果するとき、この試行を n 回行うとする．n 回の試行のうち成功の回数を S_n と表記する．S_n は確率変数である．S_n がとり得る値は、$0, 1, 2, ..., n$ である．だから、S_n は整数値確率変数でもある．各試行で成功する確率を p とするとき、前節で導出したとおり

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 1, 2, ..., n.$$

よって、確率密度関数は

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x},\ x = 1, 2, ..., n \tag{7}$$

となることは自明である．このような確率密度関数を持つ分布を、パラメターが n と p である二項分布という．

例 2.2 (ポアッソン分布 Poisson distribution)
λ を正の実数とする．パラメータ λ を持つポアッソン分布の確率密度は以下のように定義される．

$$f(x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, 2, ..., \\ 0, & \text{その他の場合} \end{cases}$$

明らかに、密度関数の条件 *(i)(ii)* は満たされる．指数関数のテイラー級数展開

$$e^\lambda = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!},$$

を用いれば、条件 *(iii)* も満たされることが確認できる．

例 2.3 (超幾何分布 hypergeometric distribution)
r 個のボールのうち r_1 個が赤で、r_2 個が黒である．ランダムに $n \leq r$ 個のボールを取り出す．取り出されたボールのうち、赤のボールの個数を X とする．X は値域を $\{0, 1, 2 \ldots, n\}$ とする確率変数となる．明らかに、

$$P(X = x) = \frac{C_{r_1,x} C_{r-r_1,n-x}}{C_{r,n}}$$

と計算できる．ここで、組み合わせの記号 C を以下のように定義する：

$$C_{r_1,x} = \binom{r_1}{x}.$$

これを変形すると、

$$\frac{C_{r_1,x} C_{r-r_1,n-x}}{C_{r,n}} = \frac{(r_1)_x (r-r_1)_{n-x}}{x!(n-x)!} \frac{n!}{(r)_n} = C_{n,x} \frac{(r_1)_x (r-r_1)_{n-x}}{(r)_n}$$

が得られる．従って、X の密度関数は

$$f(x) = \begin{cases} \frac{C_{r_1,x} C_{r-r_1,n-x}}{C_{r,n}}, & x = 0, 1, 2, ..., \\ 0, & \text{その他の場合} \end{cases}$$

となる．これを超幾何分布の密度関数という．この式はまた、

$$f(x) = \begin{cases} C_{n,x} \frac{(r_1)_x (r-r_1)_{n-x}}{(r)_n}, & x = 0, 1, 2, ..., \\ 0, & \text{その他の場合} \end{cases}$$

と表現してもよい．

例 2.4 (負の二項分布　negative binomial distribution)
α を任意の正の実数、そして、$0 < p < 1$ とする．以下の密度関数

$$f(x) = \begin{cases} p^{\alpha} \binom{-\alpha}{x} (-1)^x (1-p)^x, & x = 0, 1, 2, ..., \\ 0, & \text{その他の場合} \end{cases}$$

を持つ確率分布を負の二項分布と呼ぶ．この関数が密度関数の条件を満たすことを証明しよう．

$$\begin{aligned} \binom{-\alpha}{x} = \frac{(-\alpha)_x}{x!} = \frac{(-\alpha)(-\alpha-1)\cdots(-\alpha-x+1)}{x!} &= \frac{(-1)^x \alpha(\alpha+1)\cdots(\alpha+x-1)}{x!} \\ &= (-1)\frac{(\alpha+x-1)_x}{x!} = (-1)^x \binom{\alpha+x-1}{x} \end{aligned}$$

なので、

$$p^x \binom{-\alpha}{x} (-1)^x (1-p)^x = p^x \binom{\alpha+x-1}{x} (1-p)^x > 0$$

が成立する．条件 (i) が満たされる．次に、条件 (ii) が満たされることを示そう．

$$(1-t)^{-\alpha} = \sum_{x=0}^{\infty} \binom{-\alpha}{x} (-t)^x$$

が成り立つので、$t = 1 - p$ とおくと、

$$p^{-\alpha} = \sum_{x=0}^{\infty} \binom{-\alpha}{x} (1-p)^x$$

である．よって、$\sum_x f(x) = 1$ が成り立つ．二項分布の密度関数の表現として

$$f(x) = \begin{cases} p^{\alpha} \binom{-\alpha-1}{x} (1-p)^x, & x = 0, 1, 2, ..., \\ 0, & \text{その他の場合} \end{cases}$$

と書くこともできる．

2.2　期待値と分散

X を離散的確率変数とする．変数 X の値域は集合 $\{x_1, x_2, ..., x_n\}$ とする．事象 $\{\omega : X(\omega) = x_i\}$ の確率は $P(X = x_i)$ であり、確率密度関数は $f(x_i) = P(X = x_i)$ となっている．このとき、確率変数 X の期待値 (expected value)EX は

$$EX = \sum_{i=1}^{n} x_i f(x_i), \tag{8}$$

と定義される．期待値は平均値 (mean value) とも言われる．確率変数の値域が無限加算個からなるときは、期待値の定義式は

$$EX = \sum_{i=1}^{\infty} x_i f(x_i)$$

となるが、この値が有限となる保証はない．上の式の右辺が定義されるためには、$\sum_{j=1}^{\infty} |x_j| f(x_j) < \infty$ が要求される．

例 2.5 (2 項分布の期待値)
X をパラメータ n と p を持つ 2 項分布に従う確率変数であるとする．確率密度関数は $f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 1, 2, ..., n$ である．期待値を求めてみよう．

$n = 1$ のときを最初に考える．1 回の試行で成功する ($X = 1$ のとき) 確率は p、成功しない ($X = 0$ のとき) 確率は $1 - p$ である．よって、

$$EX = 0 \cdot f(X = 0) + 1 \cdot f(X = 1) = p,$$

となる．1 回の試行が成功する確率が平均値となる．$n > 1$ のとき、X は値 $\{0, 1, 2, \ldots, n\}$ を取ると仮定する．したがって、

$$EX = \sum_{j=0}^{n} j \binom{n}{j} p^j (1-p)^{n-j}.$$

この計算を行うと、

$$EX = np, \tag{9}$$

が得られる[5]．直接的に計算することもできるが、下の定理を用いるとさらに容易に計算できる．

例 2.6 (ポアッソン分布の期待値)
パラメータ λ を持つポアッソン分布の確率密度は $f(x) = \lambda^x e^{-\lambda}/x!$ である．この分布の期待値は

$$EX = \sum_{j=1}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda} = \sum_{j=1}^{\infty} \frac{\lambda^\lambda}{(j-1)!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = \lambda e^{-\lambda} e^{\lambda} = \lambda,$$

と計算できる．

複数の確率変数の和や積の期待値を計算する必要性も生じる．こうした要請に応えるために、以下のような性質が必要となる．

$\mathbf{X}$ を離散型 r 次元確率ベクトル（r 個の確率変数からなるベクトル) とし、その取り得るベクトル値を $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ とする．$f(x)$ を確率ベクトル $\mathbf{X}$ の密度関数、$\psi(\mathbf{x})$ を R^r 空間上で定義された実数値関数とする．このとき、関数 $\psi(\mathbf{x})$ の期待値は

$$\sum_{\mathbf{x}} \psi(\mathbf{x}) f(\mathbf{x}) = \sum_{j} \psi(\mathbf{x}_j) f(\mathbf{x}_j),$$

と定義される．この定義のもとで、以下の定理が成立する．

[5] この計算の詳細については、Hoel, Port & Stone を参照してください．

定理 2.1
$\mathbf{Z} = \psi(\mathbf{x})$ とする．このとき、

$$E\mathbf{Z} = \sum_{\mathbf{x}} \psi(\mathbf{x}) f(\mathbf{x})$$

である．ただし、

$$\sum_{\mathbf{x}} |\psi(\mathbf{x})| f(\mathbf{x}) < \infty$$

が満たされているとする．

この定理を用いると、更に以下の定理が証明できる．

定理 2.2
X と Y を有界な期待値を持つ二つの確率変数とする．

(i). c が定数であり、$P(X = c) = 1$ であるとき、$EX = c$ である．

(ii). c が定数であるとき、$E(cX) = cEX$ である．

(iii). $E(X + Y) = EX + EY$ が成立する．

(iv). $P(X \geq Y) = 1$ とする．このとき、$EX \geq EY$ であり、さらに、$P(X = Y) = 1$ のときそのときにのみ、$EX = EY$ である．

(v). $E|X| \geq |EX|$ が成立する．

例 2.7 (2 項分布の期待値)
上の定理を用いて 2 項分布の期待値を導出する．$X_1, X_2, \cdots, X_n$ が互いに独立な、成功の *(値 1 を取る)* 確率が p である確率変数であるとする．$S_n = X_1 + X_2 + \cdots + X_n$ と置けば、S_n は 2 項分布を持つ．既に、$EX_i = p, (i = 1, 2, \ldots, n)$ であることは知られている．よって、

$$ES_n = E(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} EX_i = np.$$

次に、確率変数 X のモーメントを定義しよう．$r \geq 0$ を整数とする．このとき、X^r が有界な期待値を持つならば、X は r 次のモーメントを持つといい、それを EX^r、あるいは $E[X^r]$ で表現する．X が r 次のモーメントを持つならば、$X - EX$ の r 次のモーメントを計算できる．これを X の r 次の中心モーメントと言う．つまり,$f(x)$ が確率変数 X の確率密度関数であるとき、

$$EX^r = \sum_x x^r f(x), \tag{10}$$

$$E(X - EX)^r = \sum_x (x - EX)^r f(x). \tag{11}$$

X が有界な 2 次モーメントを持つとき、2 次の中心モーメントを分散 (variance) という．X の分散は、$\mathrm{Var}X$ あるいは $V(X)$ と表記される．

$$\mathrm{Var}X = E[(X - EX)^2] \tag{12}$$

と定義される．変形すると、

$$\mathrm{Var}X = EX^2 - (EX)^2$$

が得られる．$\sigma = \sqrt{\mathrm{Var}X}$ は標準偏差 (standard deviation) と呼ばれる．

二つの確率変数 X, Y の和の分散は

$$\mathrm{Var}(X+Y) = E[\{X+Y-E(X+Y)\}^2$$

と計算される．確率変数の和の分散は分散の和に等しくならない．なぜなら、

$$\mathrm{Var}(X+Y) = \mathrm{Var}X + \mathrm{Var}Y + 2E[X-EX][Y-EY]$$

である．この最後の項を共分散 (covariance) といい、$\mathrm{Cov}(X,Y)$ と表記される．

$$\mathrm{Cov}(X,Y) = E[X-EX][Y-EY] \tag{13}$$

が定義である．したがって、

$$\mathrm{Var}(X+Y) = \mathrm{Var}X + \mathrm{Var}Y + 2\mathrm{Cov}(X,Y)$$

が成立する．さらに、共分散は

$$\mathrm{Cov}(X,Y) = E(XY) - (EX)(EY)$$

と変形できる．X と Y が独立であるときのみ、X と Y の共分散はゼロとなる．よって、X と Y が互いに独立な確率変数であるとき、$\mathrm{Var}(X+Y) = \mathrm{Var}X + \mathrm{Var}Y$ となる．c が定数であるならば、$\mathrm{Var}(X+c) = \mathrm{Var}X + \mathrm{Var}(c) = \mathrm{Var}X$ である．また、変数 $X_i, i = 1, 2, \ldots, n$ が互いに独立で、同一の分散 σ^2 を持つならば、

$$\mathrm{Var}(X_1 + X_2 + \cdots + X_n) = n\mathrm{Var}X_1 = n\sigma^2 \tag{14}$$

が成立する．さらに、

$$\mathrm{Var}(cX) = c^2\mathrm{Var}X \tag{15}$$

が成立することが知られている．この関係の導出は読者の練習問題とする．

例 2.8 (2 項分布の分散)
例 *2.6* で導入した変数を用いる．$ES_n = np$ であることは分かっている．独立な確率変数の和の分散に関する関係式から、

$$VarS_n = n\,VarX_1.$$

$X_1 = 0, 1$ なので、$X_1^2 = X_1$ である．だから、$EX_1^2 = EX_1 = p$ となる．よって、

$$VarX_1 = EX_1^2 - (EX_1)^2 = p - p^2 = p(1-p)$$

が成立する．したがって、*2* 項分布の分散は $VarS_n = np(1-p)$ となる．

例 2.9 (ポアッソン分布の分散)
X がパラメータ λ を持つポアッソン分布であるとき、$EX = \lambda, VarX = \lambda$ になることを計算しなさい．

X と Y がゼロでない有界な分散を持つ確率変数であるとき、X と Y の間の依存関係を相関係数で表現する．相関係数 (correlation coefficient) は

$$\rho = \rho(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{(\mathrm{Var}X)(\mathrm{Var}Y)}} \tag{16}$$

と定義される．もし相関係数 ρ がゼロならば、X と Y は相関関係を持たないという．X と Y が互いに独立な確率変数ならば、明らかに、無相関である．しかし、独立でない変数間でも相関関係を持たないケースが存在する．統計学上で重要な事実は以下の性質である．ρ の値は -1 と $+1$ の間にあること、つまり

$$-1 \leq \rho \leq 1$$

であり、$|\rho| = 1$ となるのは、ある定数 a に対して $P(X = aY) = 1$ が成立するとき, そのときだけである．この性質は Schwarz の不等式を用いて証明できるので、読者の練習問題とする．

この節の最後に、大数の法則を説明しよう．$X_1, X_2, \cdots, X_n$ が互いに独立な、同一の分布を持つ確率変数であるとする．これらの変数は、同じ分布を持つ母集団の中から独立に取り出された試行の結果の集まりと想定できる．母集団の平均値が μ であるとする．このとき、$S_n/n = (X_1 + X_2 + \cdots + X_n)/n$ は、n が非常に大きくなるにしたがって、平均値 μ に近づいていくと予想される．X_i が有界な分散を持つ限り、

$$\mathrm{Var}(S_n/n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

が成立する．ゆえに、$n \to \infty$ となるとき、$\mathrm{Var}(S_n/n) \to 0$ となる．観測された回数が大きくなるにつれて、観測値の平均値は母集団の平均値 μ に収束していく．この性質を厳密に定式化したものが以下の定理である．

定理 2.3 (大数の弱法則 the weak law of large numbers)
任意の実数 $\delta > 0$ に対して、

$$\lim_{n\to\infty} \Pr(|\frac{S_n}{n} - \mu| \geq \delta) = 0$$

が成立する．

定理 2.4 (大数の強法則 the strong law of large numbers)
$X_i,\ i = 1, \cdots, n$ が可測関数ならば、

$$\Pr(\lim_{n\to\infty} \frac{S_n}{n} = \mu) = 1$$

が成立する[6]．

3　連続型確率分布

3.1　分布関数と密度関数

前節までは、確率変数の値域が可算集合であるケースを議論してきた．以下では、確率空間上に定義される確率変数の値域が実数値の集合である一般的なケースを取り上げる．確率変数の値域が実数の区間 (例えば、$(-\infty, \infty)$, $[0, \infty)$, $[a, b]$ など）となるとき、確率変数が特定の値を取る確率はゼロである．つまり、

$$P(\{\omega | X(\omega) = x\}) = 0, -\infty < x < \infty,$$

[6]集合 $A = \{x : X_n(x) < a\}$ がボレル集合族に属するならば、関数 X_n は可測であるという．

を満たす．このような性質をもつ確率変数 X を連続確率変数 (continuous random Variable) という．

確率空間 $(\Omega, \mathscr{F}, P)$ 上で定義される実数値関数 $X(\omega), \omega \in \Omega$ は、集合 $\{\omega | X(\omega) \leq x\}, -\infty < x < \infty$ が確率事象になるように定められる．このとき、確率変数 X の分布関数 (distribution function)F は

$$F(x) = P(X(\omega) \leq x), -\infty < x < \infty$$

と定義される．分布関数は以下の性質を満たすことが知られている．

(i). $0 \leq F(x) \leq 1, -\infty < x < \infty.$

(ii). F は x の非減少関数である．

(iii). $F(-\infty) = 0, F(+\infty) = 1$.

(iv). $F(x+) = F(x), -\infty < x < \infty$.

関数 F が上記の 4 条件を満たすならば、F を分布関数とするような確率空間とその確率空間上で定義される確率変数を導くことができる[7].

条件 (i) から () までは説明を要しないと思われる[8]．$F(x+) = P(X \leq x)$ であり、$F(x-) = P(X < x)$ であるから、条件 () は関数 F が右連続であることを意味している．ここで

$$F(x+) - F(x-) = P(X = x)$$

が成立することも分かる．だから、もし $P(X = x) > 0$ ならば、F は x でジャンプしており、その大きさは $P(X = x)$ である．したがって、もし確率変数が連続的な値域を持つならば、分布関数は連続である．言い換える、確率変数 X が $P(X = x) = 0$ を満たすならば、X は連続的な確率変数であるという．X が連続確率変数であることと、分布関数が連続であることとは等価である．

以下の条件を満たす非負の関数 f を密度関数 (probability density function) という．

$$\int_{-\infty}^{\infty} f(x)dt = 1$$

f が密度関数ならば、

$$F(x) = \int_{-\infty}^{x} f(y)dy, -\infty < x < \infty$$

によって定義される関数 F は条件 () から () を満たす連続な分布関数である．

X を密度関数 f をもつ連続確率変数とする．もし

$$\int_{-\infty}^{\infty} |x| f(x)dx < \infty,$$

ならば、X は有限な期待値をもち、その期待値は

$$EX = \int_{-\infty}^{\infty} x f(x)dx$$

[7]分布関数の一般的な取り扱いについては、例えば、Chung(1964) が参考になる．Kailai Chung(1964), *A Course in Probability Theory*, Second Edition, Academic Press.

[8]正の実数 h に対して、$\lim_{h \to 0} f(x+h)$ を点 x での右側極限といい、$f(x+)$ で表現する．同様に、$\lim_{h \to 0} f(x-h)$ を点 x での左側極限といい、$f(x-)$ で表現する

で与えられる．

もし X が有限な m 次モーメントをもつならば、

$$E(X^m) = \int_{-\infty}^{\infty} x^m f(x)dx, \tag{17}$$

$$E[(X-\mu)^m] = \int_{-\infty}^{\infty} (x-\mu)^m f(x)dx, \tag{18}$$

である．ただし、μ は X の期待値である．

以下に連続確率変数の主要な例をあげる．

例 3.1 (正規分布 normal distribution)
平均値が μ、標準偏差が σ であるような正規分布の密度関数は

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\}, -\infty < \mu < \infty, \sigma > 0$$

で与えられる．確率変数 V の自然対数が正規分布をしているとき、変数 V は対数正規分布 *(lognormal distribution)* をしているという．つまり、確率変数 X が平均値 μ、標準偏差 σ をもつ正規分布をしているならば、$V = e^X$ は対数正規分布をしている．このとき、V の密度関数は

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma v} \exp\{-\frac{(\ln v - \mu)^2}{2\sigma^2}\},\ v \geq 0$$

となる．平均値と分散は

$$EV = \exp\{\mu + \sigma^2/2\},\ \mathrm{Var}V = \exp\{2(\mu + \sigma^2/2)\}[\exp\{\sigma^2\} - 1]$$

である．

例 3.2 (指数分布 exponential distribution)
密度関数が

$$f(x) = \lambda e^{-\lambda x}, x \geq 0; \quad f(x) = 0, x < 0$$

であるとき、非負の確率変数 x はパラメター λ の指数分布をしているという．対応する分布関数は

$$F(x) = 1 - e^{-\lambda x}, x \geq 0$$

となる．平均値と分散は

$$EX = 1/\lambda,\ \mathit{Var}X = 1/\lambda^2$$

で与えられる．

例 3.3 (ガンマ分布)
パラメター α と λ を持つガンマ密度関数 $\Gamma(x;\alpha,\lambda)$ は

$$\Gamma(x;\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, x > 0,$$

で与えられる．ただし、$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ である．

以下の性質が知られている．

1. $X_1, \ldots, X_n$ が独立な n 個の確率変数で、各変数 X_i がガンマ密度関数 $\Gamma(\alpha_i, \lambda)$ を持つとする．このとき、確率変数の和 $X_1 + \cdots + X_n$ はガンマ密度関数 $\Gamma(\alpha, \lambda)$ を持つ．ただし、$\alpha = \alpha_1 + \cdots + \alpha_n$ である．

2. ガンマ密度関数 $\Gamma(1, \lambda)$ はパラメター λ をもつ指数分布の密度関数と同じである．従って、パラメター λ の指数密度関数を持つ n 個の確率変数の和はガンマ密度関数 $\Gamma(n, \lambda)$ を持つ．

3. また、$X_1, \ldots, X_n$ が独立な正規分布 $N(\mu_i, \sigma_i^2)$ であるならば、n 個の正規分布の和は平均値 $\mu = \mu_1 + \cdots + \mu_n$ と分散 $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2$ を持つ正規分布となる．

3.2 結合分布と条件付密度関数

X, Y を同一の確率空間上に定義された確率変数とする．この二つの確率変数の結合分布関数 (joint distribution function) は

$$F(x, y) = P(X \leq x, Y \leq y), x, y \in R$$

と定義される．X と Y それぞれの分布関数は、

$$F_X(x) = P(X \leq x); \quad F_Y(y) = P(Y \leq y)$$

と定義され、X と Y それぞれの周辺分布関数と言われる．そして、

$$F_X(x) = F(x, \infty) = \lim_{y \to \infty} F(x, y),$$
$$F_Y(y) = F(\infty, y) = \lim_{x \to \infty} F(x, y),$$

が成立している．以下の性質を満たす非負の関数 f を結合密度関数と言う．

$$F(x, y) = \int_{-\infty}^{x} (\int_{-\infty}^{y} f(u, v) dv) du, \quad x, y \in R.$$

ある例外を除けば、

$$f(x, y) = \frac{\partial}{\partial x \partial y} F(x, y)$$

が成立する．また、確率変数 X の周辺密度関数が

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

となることは容易に理解できる．もし以下の条件

$$P(X \leq x, Y \leq y) = P(X \leq x) P(Y \leq y)$$

が成立するならば、X と Y は独立であるといわれる．従って、X と Y が独立であるための必要十分条件は

$$F(x, y) = F_X(x) F_Y(y)$$

となる．

X を与えたときの、Y の条件付密度を

$$f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}, 0 < f_X(x) < \infty; \quad f_{Y|X}(y|x) = 0, f_X(x) = 0 \tag{19}$$

で定義する．よって、$X = x$ を与えたときの Y の条件付期待値 $E[Y|X = x]$ は

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy = \frac{\int_{-\infty}^{\infty} y f(x, y) dy}{f_X(x)}$$

で計算される．ただし、$0 < f_X(x) < \infty$ を仮定する．

$$f(x, y) = f_X(x) f_{Y|X}(y|x)$$

だから、

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx = \int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x) dx.$$

この式を (19) 式に代入すると、

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y|x) dx}. \tag{20}$$

この関係式はベイズの規則の連続確率変数版になっている．

3.3 中心極限定理

$X_1, X_2, \cdots, X_n$ を互いに独立な、同一の分布を持つ確率変数であるとする．この確率分布の平均値は μ であり、分散は σ^2 であるとする．新たな確率変数 S_n を

$$S_n = X_1 + \cdots + X_n$$

と定義すると、S_n の平均値は $n\mu$、分散は $n\sigma^2$ である．このとき、以下の定理が成立することが知られている．

定理 3.1 (中心極限定理 central limit theorem)

$$\lim_{n \to \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x), \quad x \in R$$

が成立する．ただし、Φ は標準化された正規分布関数であり、

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$$

と表現される．

中心極限定理は、n が大きいならば、近似式

$$P(S_n \leq x) \approx \Phi\left(\frac{x - n\mu}{\sigma\sqrt{n}}\right) = \Phi\left(\frac{x - ES_n}{\sqrt{\mathrm{Var} S_n}}\right)$$

が有効であることを意味する．この近似式を実際に使用しても誤差が問題にならないようになるには、n がどの程度大きくなればよいのでしょうか．典型的な例では、$n = 25$ を超えれば、近似式が十分有効性を発揮すると言われている．

例 3.4
ある電球の寿命は平均寿命が *10* 日の指数分布をしている．*1* 年以内に電球が *50* 個以上必要となる確率はどれほどか．X_i を i 番目に設置された電球の寿命とすると、n 個目の電球が切れる時間は $S_n = X_1 + \cdots + X_n$ で表現できる．各変数は独立で、パラメター $\lambda = 1/10$ の指数分布をしている．S_n の平均値は $50\lambda^{-1} = 500$、分散は $50\lambda^{-2} = 5000$ である．中心極限定理によれば、近似式

$$P(S_{50} \leq 365) \approx \Phi(\frac{365 - 500}{\sqrt{5000}}) = \Phi(-1.91) = 0.028$$

が成り立つ．よって、*1* 年に *50* 個以上の電球が必要となる確率は 0.028 となる．

4　母関数と特性関数

4.1　母関数

確率変数 X のモーメント母関数 (moment generating function)$M_X(t)$ は

$$M_X(t) = E[e^{tX}],$$

と定義される．ただし、$M_X(t)$ の定義域はすべての実数で、その値は有界であると仮定する．
X が離散確率変数のケースを考えよう．$X = 0, 1, 2, \cdots, n, \cdots, \infty$ なので、上の定義式から

$$M_X(t) = \sum_{n=0}^{\infty} e^{tn} P(X = n)$$

となる．また、確率母関数 (probability generating function) が

$$\phi_X(s) = \sum_{n=0}^{\infty} P(X = n) s^n,$$

と定義される．ただし、$1 \geq s \geq -1$ である．モーメント母関数と確率母関数とは

$$M_X(t) = \phi_X(e^t)$$

の関係式によって関連付けられる．離散型確率分布で母関数というときは、後者の確率母関数のことを指す場合が多い．ここではこれに従う．

例 4.1 (ポアッソン分布の母関数)
ポアッソン分布は

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

なので、

$$\phi_X(s) = \sum_{k=0}^{\infty} P(X = k) s^k = \sum_{k=0}^{\infty} s^k \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda s)^k}{k!}.$$

これを計算すると、

$$\phi_X(s) = e^{-\lambda(1-s)}.$$

例 4.2 (正規分布の母関数)
X は平均値 μ、分散 σ^2 をもつ正規分布に従うとする．

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)^2/2\sigma^2]} dx \quad .$$

これを計算すると[9]、

$$M_X(t) = e^{\mu t} e^{\sigma^2 t^2/2}, t \in R.$$

母関数と確率分布関数は一対一に対応しているので、母関数を用いてモーメントなどを計算する方法は非常に有効な手法である．ここで主要な関係式を説明しよう．
X と Y が独立な確率変数であるとき、e^{tX} と e^{tY} は独立である．よって、

$$M_{X+Y}(t) = Ee^{t(X+Y)} = Ee^{tX}e^{tY} = Ee^{tX}Ee^{tY} = M_X(t)M_Y(t).$$

これを一般化すれば、$X_1, \cdots, X_n$ が独立な確率変数であるとき、

$$M_{X_1+\cdots+X_n}(t) = M_{X_1}(t) \cdots M_{X_n}(t) \tag{21}$$

が成り立つことは容易に類推できる．離散型確率変数の母関数に対してもこの関係は成立する．$\xi_1, \xi_2, \ldots, \xi_n$ を独立な離散型確率変数とし、それぞれ母関数 $\phi_1, \phi_2, \cdots, \phi_n$ をもつとする．このとき、確率変数の和 $X = \xi_1 + \xi_2 + \cdots + \xi_n$ の母関数は各母関数の積になる、

$$\phi_X(s) = \phi_1(s)\phi_2(s) \cdots \phi_n(s). \tag{22}$$

これらの関係式は、独立な確率変数の和からなる分岐過程のような確率過程を取扱うとき非常に有用なものである．
離散型確率変数 ξ の母関数を

$$\phi(s) = \sum_{k=0}^{\infty} p_k s^k$$

とする．この母関数を 1 階微分すると、

$$\frac{d\phi(s)}{ds} = p_1 + 2p_2 s + 3p_3 s^2 + \cdots,$$

となるので、

$$\left.\frac{d\phi(s)}{ds}\right|_{s=1} = p_1 + 2p_2 + 3p_3 + \cdots = E[\xi],$$

となる．同様に、２回微分を行うと、

$$\left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} = E[\xi^2] - E[\xi],$$

が得られる．だから、

$$E[\xi^2] = \left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} + \left.\frac{d\phi(s)}{ds}\right|_{s=1}$$

となる．したがって、

$$\mathrm{Var}[\xi] = E[\xi^2] - (E[\xi])^2 = \left.\frac{d^2\phi(s)}{ds^2}\right|_{s=1} + \left.\frac{d\phi(s)}{ds}\right|_{s=1} - \{\left.\frac{d\phi(s)}{ds}\right|_{s=1}\}^2 \tag{23}$$

が成立する．

[9]計算の詳細は、Hoel, Port & Stone,197 頁を参照してください．

例 4.3 (ポアッソン分布の平均値と分散)
ポアッソン分布の母関数は $\phi_X(t) = e^{\lambda(t-1)}$ である．よって

$$\frac{d\phi(t)}{dt} = \lambda e^{\lambda(t-1)}, \quad \frac{d^2\phi(t)}{dt^2} = \lambda^2 e^{\lambda(t-1)}.$$

これより、

$$E\xi = \lambda, \quad \mathrm{Var}\xi = \lambda^2 + \lambda - \lambda^2 = \lambda$$

が得られる．

連続型確率変数に対して同様の関係式が成立する．母関数は、

$$M_X(t) = Ee^{tX} = E\sum_{n=0}^{\infty}\frac{t^n X^n}{n!} = \sum_{n=0}^{\infty}\frac{EX^n}{n!}t^n. \tag{24}$$

母関数を $t=0$ の周りでテイラー級数展開すると、

$$M_X(t) = M_X(t)\Big|_{t=0} + t\frac{d}{dt}M_X(t)\Big|_{t=0} + \frac{t^2}{2!}\frac{d^2}{dt^2}M_X(t)\Big|_{t=0} + \cdots \tag{25}$$

となる．式 (24) と式 (25) の各項を比較すると

$$EX^n = \frac{d^n}{dt^n}M_X(t)\Big|_{t=0} \tag{26}$$

が得られる．

例 4.4 (正規分布のモーメント)
平均値ゼロの正規分布の母関数は $M_X(t) = e^{\sigma^2 t^2/2}$ である．よって、

$$\frac{d}{dt}M_X(t) = \sigma^2 t e^{\sigma^2 t^2/2}$$

より、$EX = 0$ ．また

$$\frac{d^2}{dt^2}M_X(t) = (\sigma^2 + \sigma^4 t^2)e^{\sigma^2 t^2/2}$$

なので、$VarX = \sigma^2$ となる．

4.2 特性関数

確率変数 X の特性関数 (characteristic function)$\varphi_X(t)$ は

$$\varphi_X(t) = E[e^{itX}]$$

と定義される．ただし、$i = \sqrt{-1}$ である．母関数は $M_X(t) = Ee^{tX}$ であるから、

$$\varphi_X(t) = M_X(it)$$

が成り立つ．

X と Y が独立な確率変数であるとき、e^{tX} と e^{tY} も独立である．よって、

$$\varphi_{X+Y}(t) = Ee^{it(X+Y)} = Ee^{itX}e^{itY} = Ee^{itX}Ee^{itY} = \varphi_X(t)\varphi_Y(t).$$

独立な確率変数の有限個の和の特性関数は各確率変数の特性関数の積になる．

離散型確率変数 ξ の特性関数は

$$\varphi_X(t) = \sum_{-\infty}^{\infty} e^{ikt} f_X(k)$$

で与えられる．この特性関数の逆変換は

$$f_X(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} \varphi_X(t) dt$$

となることが知られている．この関係を用いると、特性関数 (母関数) から確率分布関数を計算することができる．さらに以下の定理が知られている．

定理 4.1
二つの確率変数が同じ特性関数を持つならば、これらの二つの確率変数は同一の分布関数を持つ．

この唯一性定理を用いると、二つの独立な正規分布をする確率変数の和はそれ自身正規分布をするという事実を証明できる．例えば、X が平均値 μ_1、分散 σ_1^2 を持つ正規分布、Y が平均値 μ_2、分散 σ_2^2 を持つ正規分布であるとする．それぞれの特性関数は

$$\varphi_X(t) = e^{i\mu_1 t} e^{-\sigma_1^2 t^2/2}, \quad \varphi_Y(t) = e^{i\mu_2 t} e^{-\sigma_2^2 t^2/2}$$

となるので、

$$\varphi_{X+Y}(t) = e^{i(\mu_1+\mu_2)t} e^{-(\sigma_1^2+\sigma_2^2)t^2/2}$$

である．かくして、確率変数 X と Y の和は平均値 $\mu_1 + \mu_2$、分散 $\sigma_1^2 + \sigma_2^2$ を持つ正規分布となることが分かる．

練習問題

1. サイコロを 2 回投げて出た目の平均を z、第 1 回目に出た目を x、第 2 回目に出た目を y とする．サイコロ投げの 1 回目と 2 回目は独立である．1 回投げたときの出た目の標準偏差と、2 回投げた場合の平均 z の標準偏差を比較しなさい．

2. 二つの株式が存在して、第 1 の株式の期待収益率は 0.12、収益率の標準偏差は 0.20 であり、第 2 の株式の期待収益率は 0.15、標準偏差は 0.18 で、両株式の収益率間の共分散は 0.01 である．第 1 株式の保有比率が 0.25 で第 2 株式の保有比率が 0.75 のポートフォリオがある．このポートフォリオの期待収益率と標準偏差の大きさを求めなさい．

3. M 社の部門別収益のデータが以下の表のように与えられている．

経済状態	A 部門	B 部門	確率
好況	50 億円	30 億円	0.40
普通	30 億円	20 億円	0.40
不況	-10 億円	5 億円	0.20

A 部門と B 部門の期待利益および標準偏差を求めなさい．また、A 部門と B 部門の間の利益における共分散および相関係数を求めなさい．

4. 株式市場における価格変動について考える．株価は1日のうちに非常に頻繁に変化を繰り返している．1日で n 回の価格変化が起こると想定し、各瞬時の価格変化の大きさを ξ_j とする．1日における価格変動の大きさは $X = \xi_1 + \xi_2 + \cdots + \xi_n$ となる．株価は1日内で頻繁に変化するので、n は十分大きい．中心極限定理に従うと、価格変化 X の分布は正規分布をするはずである．現実の市場ではこのようなことが観察できるでしょうか．調べてみてください．

5. $\xi_1, \xi_2, \ldots$ が平均値 μ、分散 σ^2 を持つ互いに独立な確率変数である．また、N が平均値 ν、分散 τ^2 を持つ離散的な確率変数で、$\xi_1, \xi_2, \ldots$ と N は互いに独立であるとする．確率変数の和 $X = \xi_1 + \cdots + \xi_N$ の平均値は $\mu\nu$、分散は $\nu\sigma^2 + \mu^2\tau^2$ となること示しなさい．(ヒント：確率変数 X の密度関数が $f_X(x) = \sum_{n=1}^{\infty} f_{X|N}(x|n) f_N(n)$ となることを用いる.)