

# 現代数理統計学: 学習ノート

Masayuki Sakai

2022/12 -

## 目次

<b>1 確率と1次元の確率変数</b>	<b>3</b>
1.1 確率と確率変数	3
1.1.1 より厳密な確率変数の定義	3
1.1.2 累積分布関数・分布関数	4
1.1.3 連続型確率変数: Continuous random variable	4
1.1.4 確率変数と分布の関係	5
1.2 確率変数の期待値と分布の特性値	5
1.2.1 期待値と線形性	6
1.2.2 定義関数とその期待値	7
1.2.3 期待値の存在性	7
1.2.4 パーセント点	7
1.2.5 一般の場合の定義	8
1.3 母関数 Generating function	10
1.3.1 確率母関数 Probability generating function	10
1.3.2 確率母関数とモーメント	11
1.3.3 積率母関数 Moment generating function	11
1.3.4 母関数の諸問題と特性関数	12
1.4 主な一次元分布	12
1.4.1 2項分布	12
1.4.2 2項分布の確率母関数・期待値・分散	13
1.4.3 ポアソン分布	13
1.4.4 負の2項分布	14
1.4.5 幾何分布	14
1.4.6 負の二項分布の確率母関数・平均・分散	14
1.4.7 超幾何分布: Hypergeometric distribution	15
1.4.8 超幾何分布の平均と分散	16

1.4.9	一様分布 . . . . .	18
1.4.10	正規分布 . . . . .	19
2	Description of the Principle	20
3	Creation of Cosmos as Objectivation	21
4	More on Cosmic Subjectivation	22
5	More on Multiplicity	23
6	Curvature	24
7	Action-Reaction Principle	25
8	Creation of the Physical Universe	26
9	An Example from the Later Stages	26

# 1 確率と1次元の確率変数

## 1.1 確率と確率変数

確率的に発生する二つの事象  $A, B$  について  $P(A) = P(B) = 1/2$  であるとする。この時,

$$X(e) = \begin{cases} 1 & e = A \\ 0 & e = B \end{cases}$$

とすると,  $X$  が取りうる値は確率的に変動する。このような変数を**確率変数** (random variable) と呼ぶ。実際に確率変数が取った値を**実現値**という。記法上は確率変数と実現値をアルファベットの太文字・小文字で区別する慣習もある。例えば

$$P(X = x) = \frac{1}{2}, \quad x = 0, 1 \quad (1.1)$$

となる。

### 1.1.1 より厳密な確率変数の定義

$X$  が**離散型確率変数** (discrete random variable: d-rv) であるとは,  $X$  のとり得る値が有限個または可算無限個の場合であり,  $X$  が**連続型確率変数** (continuous random variable: c-rv) であるとは, 以下で定義する密度関数を持つ場合をいう。

d-rv  $X$  の場合 (1.1) 式のように各  $x$  での確率を考えて, これを関数と見做したものを確率変数  $X$  の**確率関数** (probability function) という。

$$p(x) = P(X = x)$$

が確率関数であり, 上の例では

$$p(x) = \begin{cases} \frac{1}{2} & x = 1, 2 \\ 0 & \text{else} \end{cases}$$

である。

#### 性質 1.1 (確率関数の性質)

確率関数は次の性質を持つ。

$$p(x) \geq 0, \forall x$$

$$\sum_x p(x) = 1$$

### 1.1.2 累積分布関数・分布関数

確率関数の累積和を取ったものを**累積分布関数** (cumulative distribution function: cdf), あるいは**分布関数** (distribution function) という. すなわち

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$$

である. cdf は「右側連続」となる.

cdf から pf を求めるには

$$p(x) = F(x) - \lim_{x_n \rightarrow x-} F(x_n) = F(x) - F(x-)$$

とすればよく, すなわち差を取れば良いことがわかる. ここで  $\lim_{x_n \rightarrow x-}$  は負の方向から単調に増加した時の収束先を意味している.

### 1.1.3 連続型確率変数: Continuous random variable

連続型確率分布 (c-rv) は取りうる実現値が連続値であるような場合をいう. 例えば,  $X \in [0, 1]$  でどの値も同様に確からしく取るもの, などが挙げられる. c-rv については, 特定の点の確率を考えることは意味がなく, **確率密度関数** (probability density function: pdf)  $f(x)$  を用いて特徴づける.

**定義 1.1** (確率密度関数: pdf)

$$f(x) = \lim_{\varepsilon \rightarrow 0} \frac{P(x \leq X \leq x + \varepsilon)}{\varepsilon} \quad (1.2)$$

(1.2) では任意の  $x$  で右辺の極限が存在することを仮定している. 厳密に測度論を用いれば, 「ほとんどすべての  $x$  について」存在すれば良い.

c-rv が従う分布を連続分布と良い, 連続分布について cdf は次のように定義される.

**定義 1.2** (連続分布の累積分布関数)

$f(x)$  を連続分布の確率密度関数 (pdf) として, その累積分布関数 (cdf) は

$$f(x) = \lim_{\varepsilon \rightarrow 0} \frac{F(x + \varepsilon) - F(x)}{\varepsilon} = F'(x)$$

と定義される.

すなわち, 連続分布においては, pdf は cdf の微分として, cdf は pdf の積分として得られる. そこで,

$$P(a < X \leq b) = F(b) - (a) = \int_a^b f(x) dx$$

を得る。これは、 $f(x)$  における  $[a, b]$  と  $x$  軸に囲まれた領域の面積と対応する。これが確率をなすためには

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (1.3)$$

とならなければならない。例えば、pdf  $f(x)$  が特定の関数  $h(x)$  と定数  $c$  によって

$$f(x) = c \times h(x)$$

と表される時、(1.3) より、

$$c = \frac{1}{\int_{-\infty}^{\infty} h(x) dx}$$

と定めることができる。この時  $c$  を**基準化定数** (normalizing constant), 正規化定数などと呼ぶ。

#### 1.1.4 確率変数と分布の関係

確率変数  $X$  の cdf  $F(x)$  であるとき、 $X$  は**分布  $F$  に従う** ( $X$  is distributed according to  $F$ ) などと言い、

$$X \sim F$$

と表すことが多い。

## 1.2 確率変数の期待値と分布の特性値

**定義 1.3** (期待値 (expected value))

$X$  が  $d-rv$  の時、 $X$  の期待値  $\mu = E[X]$  は次のように定義される。

$$E[X] = \sum_x xp(x)$$

また、 $X$  が  $c-rv$  の時、 $X$  の期待値は次のように定義される。

$$E[X] = \int xf(x)dx$$

期待値は平均とも呼ばれるが、標本平均と明確に区別するために**母平均** (population mean) と呼ぶこともある。期待値のように分布の位置を表す母数を**位置母数** (location parameter) という。

**定義 1.4** (分散 (variance, population variance))

$X$  が  $d-rv$  の時、 $X$  の分散  $\sigma^2 = Var[X]$  は次のように定義される。

$$E[X] = \sum_x (x - \mu)^2 p(x)$$

また,  $X$  が  $c-rv$  の時,  $X$  の分散は次のように定義される.

$$E[X] = \int (x - \mu)^2 f(x) dx$$

分散の正の平方根  $\sigma = \sqrt{\sigma^2}$  を母標準偏差 (population standard deviation) という. これは実現値の散らばり具合を表し, **尺度母数** (scale parameter) と呼ばれる.

一般に  $g(X)$  を  $X$  の関数として,  $g(X)$  の期待値を

$$E[g(X)] = \begin{cases} \sum g(x)p(x) \\ \int g(x)f(x)dx \end{cases} \quad (1.4)$$

と表す.

### 定義 1.5 (モーメント)

$g(X) = (X - \mu)^2$  とおくと  $Var[X] = E[(X - \mu)^2] = E[g(X)]$  と表せる.  $g(X)$  を  $X$  の冪乗としたとき  $E[g(X)]$  を **モーメント** (積率, moment) という. 特に, 原点周りの  $k$  次のモーメント ( $k = 0, 1, 2, \dots$ ) は

$$\mu'_k = E[X^k]$$

で定義される. また平均周りの  $k$  次のモーメント ( $k = 0, 1, 2, \dots$ ) は

$$\mu_k = E[(X - \mu)^k]$$

で定義される.

#### 1.2.1 期待値と線形性

期待値の計算においては, 次のような線形性を持つ.

$$E[ag(X) + bh(X)] = aE[g(X)] + bE[h(X)]$$

ここで  $a, b$  は定数,  $g, h$  は  $X$  の関数である. これより, 分散は

$$\begin{aligned} \sigma^2 &= E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - \mu^2 \end{aligned}$$

と表現することができる. これは, 二乗の平均から平均の二乗を引いたものである. また, 分散については以下が成り立つ (容易に確認できる).

$$Var[a + bX] = b^2 Var[X]$$

### 1.2.2 定義関数とその期待値

#### 定義 1.6

$g(x)$  が 0, 1 どちらかのみを取る場合を考える. いま,  $g(x) = 1$  となる  $x$  の集合を  $A = \{x | g(x) = 1\}$  とする. 一般に集合  $B$  の**定義関数**あるいは**指示関数** (*indicator function*) を

$$I_B(x) = \begin{cases} 1, & \text{if } x \in B \\ 0, & \text{otherwise} \end{cases}$$

と定義する.

定義より明らかに,  $g(x) = I_A(x)$  である. これより, 集合  $A$  の定義関数の期待値は  $X$  が  $A$  に属する確率に一致する. つまり,

$$E[I_A(X)] = P(X \in A)$$

となる.

ここで,  $Y = g(X)$  を新たな確率変数として  $Y$  の分布に基づいて定義 1.3 で定義したものが (1.4) に一致するかを確かめる必要があるが, 厳密には測度論が必要になる.

特定の条件下における証明は練習問題として取っておく.

### 1.2.3 期待値の存在性

期待値が無級数や積分を含むことを踏まえれば, 収束について議論する必要がある. 有名な例としては, ペテルスブルグのパラドックスがある. これは確率変数  $X$  の取りうる値を  $2^k, k = 1, 2, \dots$  として,

$$P(X = 2^k) = \frac{1}{2^k}$$

とする. このとき

$$E[X] = \sum_{k=1}^{\infty} 2^k = 1 + 1 + \dots = \infty$$

となる. これは歪みのないコインを投げて  $k$  回目に初めて表が出たら  $2^k$  円もらえるようなくじの結果とみなせて, このくじの平均的な収益が無級大であるというのは直感にそぐわない. このことからパラドックスと呼ばれている.

期待値が収束しない場合は注意する必要がある. 多くの場合は  $E[|g(X)|] < \infty$  となる場合を考えることになる. この絶対値は負の方向への発散も除外するためである.

### 1.2.4 パーセント点

1次元分布のパーセント点あるいはパーセンタイル (percentile) とは, ある値  $x$  がその分布において昇順に  $100\alpha\%$  となるようなとき,  $x$  は  $100\alpha\%$  点であるという. 特に 50% 点のこと

を**中央値** (median) などと呼ぶ。逆に降順に考える場合もあるが、その場合明確に区別するために**上側  $100\alpha\%$  点**, **下側  $100\alpha\%$  点**, などという。

もし,

$$P(X \leq -x) = P(X \leq x), \quad \forall x$$

を満たすのであれば  $X$  の分布は原点に関して対象であるという。このとき、上側  $\alpha$  点と下側  $\alpha$  点は符号のみが異なるだけなので、**両側  $2\alpha$  点**と呼ぶこともある。

累積分布関数  $F$  が連続かつ狭義の単調増加の場合は、 $F$  の逆関数  $F^{-1}(u), 0 < u < 1$  が一意に定まり、 $x_u = F^{-1}(u)$  は  $P(X \leq x_u) = u, P(X \geq x_u) = 1 - u$  を満たす。すなわち、 $x_u$  は  $F$  の下側  $u$  点である。このような  $F^{-1}$  を**分位点関数** (quantile function) という。

### 1.2.5 一般の場合の定義

$F$  の不連続点や平坦な部分があった場合 (単調関数ではない場合) は  $F^{-1}$  を自明に定義できない。そこでまず上側確率と下側確率をそれぞれ考えていこう。

まず下側確率に対応して  $F_L^{-1}(u)$  を以下のように定義する。

$$\begin{aligned} F_L^{-1}(u) &= \inf\{x | F(x) \geq u\} \\ &= \min\{x | F(x) \geq u\} \end{aligned} \tag{1.5}$$

*Proof.* (1.5) 式における  $\inf, \min$  の等号成立。まず,

$$\begin{aligned} x_0 &= \inf\{x | F(x) \geq u\} \\ A &= \{x | F(x) \geq u\} \\ x_n &\rightarrow x_0+ \end{aligned}$$

とする。

$F$  の右連続性より、 $F(x_0) = \lim_{n \rightarrow \infty} F(x_n)$  である。いま定めかたより  $\forall x_n \in A$  で  $F(x_n) \geq u$  があるので、もちろん  $F(x_0) \geq u$  である。これより、 $x_0 \in A$  かつ  $x_0 = \min A$  となることがわかる。□

*Proof.*  $F_L^{-1}$  の左連続性。  $F_L^{-1}$  は左連続である。つまり  $u_n \rightarrow u-$  のとき、 $\lim_{n \rightarrow \infty} F_L^{-1}(u_n) = F_L^{-1}(u)$  となる。まず  $x_n = F_L^{-1}(u_n), x = F_L^{-1}(u)$  とおく。  $u_n \rightarrow u-$  つまり負の方向から近づけることを考えれば、明らかに  $x_n$  は単調非減少数列を為す。故に  $x^* = \lim_{n \rightarrow \infty} x_n$  が存在する。これより  $\forall x_n \leq x$  であり、同様に  $x^* \leq x$  となる。(1.5) 式より、 $\forall n$  について  $F(x_n) \geq u_n$  である。さて、 $x^* \geq x_n$  より  $F(x^*) \geq F(x_n)$  で、 $n \rightarrow \infty$  のとき  $F(x^*) \geq u$  である。従って、 $x = F_L^{-1}(u)$  という定義より、 $x^* \geq x$  となる。以上より、 $\lim_{n \rightarrow \infty} x_n = x^* = x$  を得る。よって  $F_L^{-1}$  は左連続である。□

次に上側確率に対応して  $F_R^{-1}$  を次のように定義する。

$$\begin{aligned} F_R^{-1} &= \sup\{x | P(X \geq x) \geq 1 - u\} \\ &= \max\{x | P(X \geq x) \geq 1 - u\} \end{aligned}$$



$F_R^{-1}$  は  $F_L^{-1}$  と対称的に定義されていることから、右連続となる。

さて、ここで  $x_L = F_L^{-1}(u)$  として、 $\forall \varepsilon > 0$  に対して  $F(x_L - \varepsilon) < u$  となる。これより  $P(X \geq x_L - \varepsilon) \geq P(X > x_L - \varepsilon) = 1 - F(x_L - \varepsilon) > 1 - u$  を得る。従って  $x_L - \varepsilon \leq x_R$  が成り立つ。 $\varepsilon > 0$  は任意なので  $x_L \leq x_R$  が示された。従って

$$F_R^{-1}(u) \geq F_L^{-1}(u)$$

が成り立つ。

また、 $I_u = \{x | P(X \leq x) \geq u, P(X \geq x) \geq 1 - u\}$  とおけば、 $I_u$  という区間は  $I_u = [F_L^{-1}(u), F_R^{-1}(u)] = [x_L, x_R]$  と表せることが示せる。これより、 $x_L = x_R$  であれば、 $x$  以下に  $u$  以上の確率を含み、 $x$  以上に  $1 - u$  以上の確率を含む点  $x = x_L = x_R$  の唯1点に定まることを意味している。また、仮に  $x_L < x_R$  であれば、複数個の点が存在しており、その最小値が  $x_L$  で最大値が  $x_R$  となっている。この場合には下側  $u$  点としてこの区間の中点をとるのが自然な定義に思える。

以上の議論より cdf  $F$  を持つ分布の下側  $\alpha$  点  $x(\alpha)$  を次のように定義しよう。

$$x(\alpha) = \frac{F_L^{-1}(\alpha) + F_R^{-1}(\alpha)}{2}, \quad 0 < \alpha < 1$$

$F_L^{-1}$  の定義より、

$$x \geq F_L^{-1}(u) \iff F(x) \geq u \quad (1.6)$$

が成り立つ。 $u$  点以上になる  $x$  であれば、 $x$  における累積分布関数  $F$  の値は  $u$  以上である。

ここで、 $U \sim \text{Unif}(0, 1)$  ( $0, 1$  の間の一様分布) を考えてみる。(1.6) 式が成り立つ確率を評価すると

$$P(F_L^{-1}(U) \leq x) = P(U \leq F(x)) = F(x)$$

となる。従って、 $F_L^{-1}(U)$  は分布  $F$  に従うことがわかる。

これは、一様分布に従う確率変数を  $F_L^{-1}$  で変換すると、 $F$  に従う確率変数が得られることを示している。

標本に対しては後に紹介する経験分布関数を用いることができる。また累積分布関数の逆関数としては  $F_L^{-1}, F_R^{-1}$  のいずれを用いても良く、これらには右、左連続以外に本質的な差はない。

$\alpha$  が 0 または 1 に収束する時の  $x(\alpha)$  の  $\pm\infty$  への発散の速さ、あるいは  $x \rightarrow \pm\infty$  の時の  $F(x)$  の 0、あるいは 1 への収束の速さを、**分布  $F$  の裾の重さ**と表現することがある。

例えば、 $x$  が非常に大きくなっても  $F(x)$  が 1 に近づかない場合、分布の右裾に大きい確率が残っていることになり、このような状態を右裾が重たい、などという。左裾も同様。

期待値や標準偏差は積分を用いて定義されるため分布の裾の大きさの影響を受けやすい。極端な場合にはこれらが有限の値を持たない（存在しない）こともある。コーシー分布などは代表的な例である。

そこで、分布の裾の重さの影響を受けないような指標**頑健な** (robust) 指標として中央値  $x(0.5)$  や 4 分位範囲 (inter-quantile range)  $Q = x(0.75) - x(0.25)$  が用いられることもある。

## 1.3 母関数 Generating function

確率分布やそのモーメントを調べる際に母関数を用いると便利ことがある。母関数と名のつくものには**確立母関数**、**積率母関数**、**特性母関数**があるが、これらは概念的には同じものである。特性関数は応用上最も一般性を持つが、複素積分を必要とするため数学的に難易度が高い（厳密な議論は確率論を必要とする）。

ここでは、確立母関数・積率母関数・特性関数の順に紹介する。

### 1.3.1 確率母関数 Probability generating function

**確率母関数** (probability generating function) は主に皮膚の整数値を取る離散確率変数の分布を調べるのに用いられる。  $|s| \leq 1$  として、  $s$  を変数とする確率母関数を

$$G(s) = E[s^X] = \sum_{x=0}^{\infty} s^x p(x) \quad (1.7)$$

と定義する。  $G(1) = \sum p(x) = 1$  である。 また (1.7) 式の収束は  $p(x) \geq 0, \sum p(x) = 1$  より  $|s| \leq 1$  の範囲で保証される。

母関数と確率分布は 1 対 1 に対応することが知られており、確率分布について知りたい場合、その母関数を対象にすることができる。非負整数上の確率分布  $p(x)$  が与えられれば、(1.7) により確率母関数  $G(s)$  が定まる。逆に、 $G(s)$  が与えられた時、 $G(s)$  を  $k$  回微分して  $s = 0$  とすると、 $G^{(k)}(0)/k! = p(k)$  となる。すなわち、母関数と確率分布が 1 対 1 に対応していることがわかる。

例えば、1 階微分を考えれば、

$$G'(s) = \sum_{x=0}^{\infty} x s^{x-1} p(x) \quad (1.8)$$

であり、 $s = 0$  のとき

$$G'(0) = p(1)$$

となる。これは  $x = 1$  の時のみ  $s^{x-1}$  の項が 1 となり、他の場合では 0 となるためである。さらに 2 階微分を考えれば

$$G^{(2)}(s) = \sum_{x=0}^{\infty} x(x-1) s^{x-2} p(x)$$

となる。同様に  $s = 0$  のとき

$$G^{(2)}(0) = 2p(2)$$

となる. この時出てくる  $k!$  の項をキャンセルするために,  $G^{(k)}(0)/k!$  としておけば, 欲しかった  $p(x)$  の項が残ることがわかる. ただし,  $G(s)$  の微分は収束が保証される  $|s| \leq 1$  で考えるものとする.

### 1.3.2 確率母関数とモーメント

確率母関数はモーメントを求めるためにも用いることができる. いま  $G(s)$  の 1 階微分は (1.8) 式で与えられていて, もし  $E[X]$  が存在するならば  $s = 1$  の時

$$G'(1) = \sum_{x=0}^{\infty} xp(x) = E[X]$$

であることがわかる. 同様に  $X$  の  $k$  次のモーメントが存在するならば

$$G^{(k)}(1) = E[X(X-1)\cdots(X-k+1)]$$

となり,  $k$  次のモーメントと一致する. これを**階乗モーメント** (factorial moment) と呼ぶ.

### 1.3.3 積率母関数 Moment generating function

確率母関数において,  $s = e^\theta, \theta \in \mathbb{R}$  と置いたものが**積率母関数** (moment generating function) であり,

$$\phi(\theta) = E[e^{\theta X}] \quad (1.9)$$

で定義される. 積率母関数は一般の (連続) 確率変数について用いられることに注意されたい.  $\theta \leq 0$  とおくと  $s = e^\theta \leq 1$  となり非負整数を考える場合は, 確率母関数と同等となる.

名の通り, 積率母関数はモーメントを計算するのに便利な関数である. **積分と微分の交換が保証されているとして**,

$$\left. \frac{d^k \phi(\theta)}{d\theta^k} \right|_{\theta=0} = E[X^k e^{\theta X}]|_{\theta=0} = E[X^k] = \mu'_k \quad (1.10)$$

であり,  $\phi(\theta)$  の 0 における高次の微分係数が  $X$  の原点周りのモーメントを表している.

従って, 平均  $\mu$  まわりのモーメントを求めるには,  $e^{-\mu\theta}\phi(\theta) = E[e^{\theta(X-\mu)}]$  に注意して,  $\phi(\theta)$  のかわりに  $\phi(\theta)e^{-\mu\theta}$  を積率母関数として用いれば良い.

積率母関数について問題になるのは (1.9) 式の積分の存在である.  $\theta > 0$  とすれば,  $\forall k > 0, \lim_{x \rightarrow \infty} x^k / e^{\theta x} = 0$  であるから,  $\theta > 0$  に対しては積率母関数が存在すれば, 全ての正数  $k$  について  $\sum_{x \geq 0} x^k p(x)$  あるいは  $\int_{x \geq 0} x^k f(x) dx$  が収束しなければならない.

ちなみに,  $\forall k > 0, \lim_{x \rightarrow \infty} x^k / e^{\theta x} = 0$  については高校数学の範囲で証明が可能である. 例えば  $\lim_{x \rightarrow \infty} e^x / x^r = \infty$  を示せば良い.  $x > r + 1$  で  $x$  が整数の場合を考え, 二項定理を用いて証明することもできる.

同様に,  $\theta < 0$  について積率母関数が存在すれば  $\sum_{x \leq 0} |x|^k p(x)$  あるいは  $\int_{x \leq 0} |x|^k f(x) dx$  が全ての正数  $k$  に対して存在しなければならない. 実際, 積率母関数が有用なのは 0 をない点として含むある区間  $(-a, b)$  で積率母関数が存在する場合で, そのときは全ての次数のモーメントが存在し, また (1.10) 式の微分と積分の交換が保証される. またこの場合原点まわりで

$$\begin{aligned} \phi(\theta) &= E[e^{\theta X}] = E\left[1 + \theta X + \frac{\theta^2 X^2}{2!} + \cdots\right] \\ &= 1 + \theta \mu'_1 + \frac{\theta^2}{2!} \mu'_2 + \cdots = \sum_{k=0}^{\infty} \frac{\theta^k}{k!} \mu'_k \end{aligned}$$

の形の無限級数表現が可能であることが知られている. 積率母関数についても, 積率母関数が存在すれば積率母関数は分布を一意的に決めるという点が重要である.

### 1.3.4 母関数の諸問題と特性関数

積率母関数には積分の存在の問題があり, 平均や分散の存在しない分布には用いることはできない. 対して特性関数は全ての分布に関して存在するので, 最も一般的な概念と言える.

特性関数は積率母関数において  $\theta$  を純虚数  $it$  とおくもので

$$\phi(t) = E[e^{itX}] = E[\cos(tX) + i \sin(tX)]$$

で定義される.

## 1.4 主な一次元分布

### 1.4.1 2項分布

コイン投げのように結果が2通りしかないような確率的な試行を**ベルヌーイ試行** (Bernoulli trial) という. また  $X_i$  のように 0, 1 のみをとる確率変数を**ベルヌーイ変数**という. 1が出る確率  $p$  を**成功確率**などと呼んだりする.

ここで確率変数  $X$  を

$$X = X_1 + \cdots + X_n$$

と定義する. これはベルヌーイ試行を  $n$  回行ったとき 1 が出た回数を表している.  $X$  の分布をパラメータ  $n, p$  の**2項分布** (Binomial distribution) といい,  $\text{Bin}(n, p)$  と表すことにする.  $\text{Bin}(n, p)$  の確率関数は以下で与えられる.

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

ここで  $\binom{n}{k} = {}_n C_k$  である.

### 1.4.2 2項分布の確率母関数・期待値・分散

$X$  の確率母関数は2項定理を用いて

$$\begin{aligned} G(s) &= E[s^X] = \sum_{x=0}^n s^x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^n \binom{n}{x} (sp)^x (1-p)^{n-x} \\ &= (ps + 1 - p)^n = (1 + p(s-1))^n \end{aligned}$$

とできる. これを用いれば,  $G'(s) = np(1 + p(s-1))^{n-1}$  より  $G'(1) = E[X] = np$ . 同様に  $G''(s) = n(n-1)p^2(1 + p(s-1))^{n-2}$  より,  $G''(1) = E[X(X-1)] = n(n-1)p^2$  を得る. これより,

$$\begin{aligned} \text{Var}[X] &= E[(X - E[X])^2] = E[X^2] - E[X]^2 \\ &= E[X(X-1)] + E[X] - E[X]^2 \\ &= n(n-1)p^2 + np - (np)^2 \\ &= np(n-1)p + np - (np)^2 \\ &= np(np - p + 1 - np) \\ &= np(1-p) \end{aligned}$$

を得る. 2項分布と呼ぶ限りは, パラメータ  $n, p$  の組み合わせの数だけ分布が存在するので, 分布の集合と言って良い. このような確率分布の集合のことを**分布族** (family of distributions) と呼ぶ. ただし2項分布のことをわざわざ2項分布族などと呼ぶことはない. この意味で, 一般的に知られている確率分布というのは基本的に分布族を指している. 分布族には確率分布を定めるためのパラメータ (あるいは母数ともいう) を持っている.

### 1.4.3 ポアソン分布

ポアソン分布は2項分布において,  $n$  を大,  $p$  を小とした時の極限として定義される.

### 1.4.4 負の2項分布

負の2項分布は、 $X_i = 1$ となる確率が $p$ であるようなベルヌーイ試行に関連して定義される。いま、確率変数 $X$ を $r$ 回の1が出るまでに0が出た回数とする。 $X = k$ とは $r$ 回1を得るまでに $k$ 回0が出たことを意味する。従って、この場合ベルヌーイ試行は $r + k$ 回行われていることになる。 $r + k - 1$ 回までに0が $k - 1$ 出て、 $r + k$ 回目は1で固定される。このような確率は

$$P(X = k) = \binom{r+k-1}{k} (1-p)^k p^{r-1} \times p = \binom{r+k-1}{k} (1-p)^k p^r \quad (1.11)$$

で表現できる。このような確率関数を持つ分布をパラメータ $r, p$ の負の二項分布といい、 $NB(r, p)$ と表す。

(1.11) 式が確率関数となっているのか(和が1となっているか)を確認しておこう。いま、 $0 < p \leq 1, q = 1 - p$ として、 $p^{-r} = (1-p)^{-r}$ を $q$ の無限級数にテーラー展開すれば

$$(1-q)^{-r} = 1 + rq + \frac{r(r+1)}{2!}q^2 + \cdots = \sum_{k=1}^{\infty} \binom{r+k-1}{k} q^k$$

ここで両辺に $p^r$ をかければ和が1になることが確かめられる。

### 1.4.5 幾何分布

負の二項分布において、 $r = 1$ のとき、

$$P(X = k) = p(1-p)^{k-1}$$

となり、確率関数が幾何級数となることから、これを幾何分布 (geometric distribution) という。幾何分布は最初に1が出るまでに0が出る回数の分布として解釈できる。

### 1.4.6 負の二項分布の確率母関数・平均・分散

負の二項分布の確率母関数は以下のようになる。

$$\begin{aligned}
G(s) &= E[s^X] = \sum_{x=0}^{\infty} s^x \binom{r+x-1}{x} (1-p)^x p^r \\
&= p^r \sum_{x=0}^{\infty} \binom{r+x-1}{x} (s(1-p))^x \\
&= \frac{p^r}{(1-(s(1-p)))^r} \sum_{x=0}^{\infty} \binom{r+x-1}{x} (s(1-p))^x (1-s(1-p))^r \\
&= \left( \frac{p}{1-sq} \right)^r \sum_{x=0}^{\infty} P(Y=x) \\
&= \left( \frac{p}{1-sq} \right)^r
\end{aligned}$$

途中の式変形で、生起確率  $s(1-p)$  の負の二項分布に従う確率変数  $Y$  の確率分布の総和の形を作り、これが 1 となることを利用している。

あとはこの形を利用して  $G'(s), G''(s)$  を求め、 $s=1$  の場合を用いて  $E[X], \text{Var}[X]$  を求めれば良い。

$$\begin{aligned}
G'(s) &= p^r r q (1-q s)^{-(r+1)} \\
G''(s) &= p^r r(r+1) q^2 (1-q s)^{-(r+2)}
\end{aligned}$$

より、

$$\begin{aligned}
G'(1) &= E[X] = \frac{r(1-p)}{p} \\
G''(2) &= E[X(X-1)] = \frac{r(r+1)(1-p)^2}{p^2}
\end{aligned}$$

を得る。従って  $E[X] = r(1-p)/p$  で  $\text{Var}[X]$  については分散の分解の性質を利用して

$$\begin{aligned}
\text{Var}[X] &= E[X^2] - E[X]^2 = E[X(X-1)] + E[X] - E[X]^2 \\
&= \frac{r(r+1)(1-p)^2}{p^2} + \frac{r(1-p)}{p} - \left( \frac{r(1-p)}{p} \right)^2 \\
&= \frac{r(1-p)}{p^2}
\end{aligned}$$

を得る。

#### 1.4.7 超幾何分布: Hypergeometric distribution

超幾何分布は、壺からの球を取り出す例えば分かりやすい。いま、壺に  $M$  個の赤玉と  $N-M$  個の白玉が入っているとする (合計は  $N-M+M=N$  個)。この壺から無作為に  $n$  個の玉を

取り出した時の赤玉の数を  $X$  として  $P(X = k)$  を考えよう.  $N$  個のものから  $n$  個のものを取り出す場合は  $\binom{N}{n}$  で取り出し方は同様に確からしい. ところで,  $k$  個の赤玉が取り出されるには  $M$  個の中から  $k$  個, 一方白玉を  $N - M$  個の中から  $n - k$  個取り出していることになる. これを考慮すれば, 取り出し方の総数は  $\binom{M}{k} \times \binom{N-M}{n-k}$  となる ( $k \leq M, n - k \leq N - M$ ). これを踏まえて,  $X$  の確率関数は

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$$

$$\max\{0, n + M - N\} \leq k \leq \min\{n, M\}$$

となる. この確率関数の和が1となることは,  $(1+x)^N = (1+x)^M (1+x)^{N-M}$  の両辺で  $x^n$  の係数を比較するとよい. 二項展開  $(a+b)^N = \sum_{n=0}^N N a^n b^{N-n}$  を用いれば,

$$(1+x)^N = (1+x)^M (1+x)^{N-M}$$

$$\sum_{n=0}^N \binom{N}{n} x^n = \left( \sum_{m=0}^M \binom{M}{m} x^m \right) \left( \sum_{k=0}^{N-M} \binom{N-M}{k} x^k \right)$$

が成り立つ. ここで  $x^n$  の係数を両辺で比較すると, 右辺で  $n$  を固定した時, 左辺では  $m, k$  の組み合わせで丁度  $n$  になっているものがあって,

$$\binom{N}{n} = \sum_{x=0}^n \binom{M}{x} \binom{N-M}{n-x}$$

となっていることがわかる. これより, 超幾何分布の確率関数の和を考えると,

$$\begin{aligned} \sum_{k=0}^n P(X = k) &= \sum_{k=0}^n \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}} \\ &= \sum_{k=0}^n \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{M}{k} \binom{N-M}{n-k}} \\ &= 1 \end{aligned}$$

となり, 確かに和が1になることが確かめられた.

#### 1.4.8 超幾何分布の平均と分散

超幾何分布の確率母関数は簡単に表せないので, 平均と分散は定義に従って計算を進めていく.



**補題 1.1**

$N, n \in \mathbb{N}, N, n \neq 0, N > n$  とする. このとき, 以下が成り立つ.

$$\binom{N}{n} = \frac{N}{n} \binom{N-1}{n-1}$$

*Proof.*

$$\begin{aligned} \binom{N}{n} &= \frac{N(N-1)!}{n(n-1)!((N-1)-(n-1))!} \\ &= \frac{N}{n} \frac{(N-1)!}{(n-1)!((N-1)-(n-1))} \\ &= \frac{N}{n} \binom{N-1}{n-1} \end{aligned}$$

□

超幾何分布の平均  $E[X]$  については, 補題 1.1 より  $k \binom{M}{k} = M \binom{M-1}{k-1}$ , また  $x = 0$  の時は値が 0 になることに注意すると

$$\begin{aligned} E[X] &= \sum_{x=0}^n x \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}} \\ &= n \frac{M}{N} \sum_{x=1}^{n-1} \frac{\binom{M-1}{x-1} \binom{(N-1)-(M-1)}{(n-1)-(x-1)}}{\binom{N-1}{n-1}} \\ &= n \frac{M}{N} \sum_{x'=0}^{n'} \frac{\binom{M'}{x'} \binom{N'-M'}{n'-x'}}{\binom{N'}{n'}} \\ &= n \frac{M}{N} \end{aligned}$$

となる. ここで最後に  $n' = n - 1, x' = x - 1, M' = M - 1, N' = N - 1$  とおいた. 次に分散を計算する. まず, 後に,  $V[X] = E[X(X-1)] + E[X] - E[X]^2$  を利用するために  $E[X(X-1)]$  を計算しておく. これは平均の計算で行った手続きを繰り返し適用すれば,

$$\begin{aligned} E[X(X-1)] &= \sum_{x=0}^n \frac{\binom{M}{x} \binom{M-N}{n-x}}{\binom{N}{n}} \\ &= n(n-1) \frac{M(M-1)}{N(N-1)} \end{aligned}$$

とできる. 次に分散を計算すると

$$\begin{aligned}
V[X] &= E[X(X-1)] + E[X] - E[X]^2 \\
&= n(n-1) \frac{M(M-1)}{N(N-1)} + n \frac{M}{N} - n^2 \frac{M^2}{N^2} \\
&= n \frac{M}{N^2} \left( \frac{N(M-1)(n-1) + N(N-1) - nM(N-1)}{N-1} \right) \\
&= n \frac{M}{N^2} \left( \frac{N^2 - NM - Nn + nM}{N-1} \right) = n \frac{M}{n^2} \frac{(N-n)(N-M)}{N-1} \\
&= n \frac{M}{N} \frac{N-n}{N-1} \left( 1 - \frac{M}{N} \right)
\end{aligned}$$

となる．ここで， $p = M/N, q = 1 - p$  とおくと  $V[X] = \frac{N-n}{N-1} npq$  と表せる．

なお，超幾何分布においては， $M, n$  を入れ替えても変わらないため， $M$  と  $n$  について対称性がある．

#### 1.4.9 一様分布

区間  $[a, b]$  上の一様分布 (uniform distribution) を  $U(a, b)$  と表す．このとき pdf および cdf はそれぞれ

$$\begin{aligned}
f(x) &= \begin{cases} \frac{1}{b-a}, & \text{if } a \leq x \leq b \\ 0, & \text{otherwise} \end{cases} \\
F(x) &= \begin{cases} 0, & \text{if } x < a \\ \frac{x-a}{b-a}, & \text{if } a \leq x \leq b \\ 1, & \text{if } x > b \end{cases}
\end{aligned}$$

平均と分散は以下の通りである．

$$\begin{aligned}
E[X] &= \int_a^b \frac{x}{b-a} dx \\
&= \frac{b+a}{2}
\end{aligned}$$

$$\begin{aligned}
E[X^2] &= \int_a^b \frac{x^2}{b-a} dx \\
&= \frac{b^2 + ba + a^2}{3}
\end{aligned}$$

$$\begin{aligned}
V[X] &= E[X^2] - E[X]^2 \\
&= \frac{(b-a)^2}{12}
\end{aligned}$$

### 1.4.10 正規分布

正規分布 (normal distribution, Gaussian distribution) はパラメーター  $\mu, \sigma$  を持つ分布族で, その密度関数は

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

と表される. さらに, その中でも  $\mu = 0, \sigma = 1$  とした分布を標準正規分布といい, この時密度関数  $f(x)$  は

$$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

となる.

## 2 Description of the Principle

aaa

### 3 Creation of Cosmos as Objectivation

## 4 More on Cosmic Subjectivation

## 5 More on Multiplicity

## 6 Curvature



## 7 Action-Reaction Principle

## 8 Creation of the Physical Universe

## 9 An Example from the Later Stages

## 参考文献

- [1] Books of Shrii Shrii Anandamurti (Prabhat Ranjan Sarkar):  
<http://shop.anandamarga.org/>
- [2] Avtk. Ananda Mitra Ac., *The Spiritual Philosophy of Shrii Shrii Anandamurti: A Commentary on Ananda Sutram*, Ananda Marga Publications (1991)  
ISBN: 81-7252-119-7