数理統計学 講義ノート（2019 年の 2 年用，担当：原隆）

（このノートは 2019 年 6 月現在の暫定版で，漸く，確率論の部分をだいたい書きました．講義ノートの章立ては教科書とは少し異なりますが，大体の対応を各節の最初に脚注として書いています.）

1 確率論の基礎

（教科書の第 2 章から入ります.）まずは確率論の基礎（枠組み）から考えて行く．

1.1 確率論の舞台 — 事象と標本空間[1]

現実の問題の「確からしさ」を議論するのはなかなか大変である．そこで，数学ではまず，現実から少し切り離した形で，考えやすい舞台を設定する.（確率そのものはもう少し後で導入）．以下のような「実験」[2]を行うことを考える．

例 1： コインを一回だけ投げる．

例 2： コインを２回投げる.（この場合，２回続けて投げたものを一回の「実験」と考える.）

例 3： さいころを一回だけ投げる．

例 4： さいころを２回投げる．

例 5： ５２枚あるトランプから一枚取り出す．

このような例では，まず，上の「実験」の結果は何通りかある．一回「実験」をやった場合にその結果が何になるかは分からないが —— だからこそ「確率論」がでてくる ——，少なくとも**可能な結果の全体**はわかっている．そこで，以下の定義を行おう．

> **定義 1.1.1** 「実験」をやる場合，**可能な結果の全体**からなる集合を**標本空間**（sample space）$\mathcal{S}$ と言う．標本空間の元（つまり，一回の「実験」の結果になりうるもの）を**標本点**または**根元事象**と言う．

- 例 1 では $\mathcal{S} = \{H, T\}$．ここで H は表が出ること，T は裏が出ることで，根元事象は T と H.
- 例 2 では $\mathcal{S} = \{(H,H),(H,T),(T,H),(T,T)\}$．ここで例えば (T,H) は一回目に表，２回目に裏がでること．
- 例 3 では $\mathcal{S} = \{1,2,3,4,5,6\}$．ここで i はさいころの i の面が出ること（$i = 1,2,\ldots,6$）
- 例 4 では $\mathcal{S} = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),\ldots,(6,5),(6,6)\} = \{(i,j)\,\big|\,i,j = 1,2,\ldots,6\}$. ここで (i,j) は一回目に i の面，２回目に j の面が出ること．
- 例 5 では $\mathcal{S} = \{$ ハートのエース，ハートの２，ハートの３$,\ldots\}$ と全部で５２個の要素からなる集合.

以下では有限な標本空間，および有限からのアナロジーで考えられる場合のみを考察する[3]．

さて，我々は根元事象のみに興味があるわけではない．たとえば例２で,「一回目に表が出ること」を知りたかったり，例３で「さいころで偶数の目が出ること」を知りたかったり，例５で「ハートが出ること（数字は問わない)」を知りたかったりする．このような問いに答えるため，**事象**と言う概念を導入する．

> **定義 1.1.2** **事象**とは実験の結果が持っている性質のこと．数学的に厳密に言うと，**事象**とは単に**標本空間の部分集合**，つまり「根元事象の集まり」のことである．なお，事象には空集合（起こり得ないこと)，および標本空間全体も含めて考える．

「部分集合」と言うと大げさだが，普通に我々の言っている「出来事」に相当していることを，下の例で納得されたい．

---
[1] 教科書の 2.1 節前半
[2] 「実験」と言っているが,「観測」などと思った方が良い場合も含める
[3] 有限でない場合はいろいろとややこしい（＝数学的に面白い）ことが起こるが，この講義ではすべて略

- 例１では可能な事象は $\varnothing$ （起こり得ない）, $\{H\}$ （「表が出た」）$\{T\}$ （「裏が出た」）, $\mathcal{S}=\{H,T\}$ （「表または裏が出た」）．
- 例２での事象の例は（根元事象で無いものを書くと） $\{(H,H),(H,T)\}$ （「一回目に表が出た（２回目は何でも良い）」）， $\{(H,T),(T,T)\}$ （「２回目に裏が出た（１回目は何でも良い）」）， $\{(H,H),(T,T)\}$ （「２回とも同じ目が出た」）など．
- 例３では $\{1,3,5\}$ （「奇数の目が出た」）， $\{1,2,3,4\}$ （「４以下の目が出た」）など．
- 例４では $\{(1,j)\,\big|\,j=1,2,\ldots,6\}$ （「１回目に１が出た」）， $\{(i,j)\,\big|\,i+j=\text{偶数}\}$ （「１回目と２回目の数字を足すと偶数」）など．
- 例５では｛ハートのエース，ハートの２，ハートの３, . . . ,ハートの１３｝（「ハートが出た」），とか｛ハートの３，スペードの３，ダイヤの３，クローバーの３｝（「３が出た」）など．

事象を標本空間の部分集合として定義するのは，以下の事象の演算ともあっている．まず，２つの事象 E,F に対して，その**和事象**を集合としての和集合 $E\cup F$ として，またその**積事象**を集合としての交わり $E\cap F$ として定義する（事象の場合，$E\cap F$ を EF と略記することが多い）．日常言語に直せば，$E\cup F$ とは E または F の**どちらかが起こること**，$E\cap F=EF$ とは E と F の**両方が起こること**を意味する．更に，E^c を $\mathcal{S}\backslash E$ （E の補集合）をして定義し，E の **余事象**と言う．これは日常言語では「事象 E が起こらないこと」に相当する．

- 例１で，$E=\{H\},F=\{F\}$ とすると，$E\cap F=\varnothing$．これは「表と裏が同時に起こることは無理」という直感にあっている．$E^c=\{T\}$ であるが，裏が出るというのは「表が出ない」ことでもあるから，これも余事象の定義にあっている．また，$E\cup F=\mathcal{S}$ であるが，これは「表または裏が出る」と言うのは要するに可能性全部だから．
- 例２で，$E=\{(H,H),(H,T)\},F=\{(H,T)\},G=\{(T,H)\},D=\{(T,T)\}$ とすると，$E\cap F=\{(H,T)\}$，$E\cap G=\varnothing$，$E\cup G=\{(H,H),(H,T),(T,H)\}$ などとなる．また，$D^c=E\cup G$ であるが，確かに「『２回とも裏』と言うことはない」という事象になっている．

なお，$A\cap B=\varnothing$ の時,「A と B は互いに背反」という．

1.2 数学における確率[4]

今までは単に確率をやる舞台を設定したにすぎない．これからいよいよ,「確率」を割り振っていこう．

数学ではある意味で「天下りに」確率を定める．本当のところを言うと，確率の定め方そのものは数学の仕事ではなく，実験の行い方に即して物理学・化学・心理学... などに基づいて決めるべきものだ．しかし，通常は確率を定めるところから始めることになる．

ただし，ここでどのような p_j を選ぶか，は個々の問題に応じてうまく決めてやる必要がある．

- 例１で，コインが裏表同じように出やすいのなら，$P(H)=P(T)=1/2$ とするのが良いだろう．
- 例３で，さいころのどの目も同じように出やすいのなら， $P(j)=1/6$ とすべし．しかし，イカサマさいころで 6 が出やすく，1 が出にくい，のなら，例えば $P(1)=\frac{1}{12},P(6)=\frac{3}{12},P(2)=P(3)=P(4)=P(5)=\frac{1}{6}$ ととるのが良いかも知れない．

今までの話を，標本空間が $\mathcal{S}=\{e_1,e_2,\ldots,e_N\}$ になる実験について一般化しておく（e_j が根元事象）．上で見たように，数学的に確率を決めるというのは，それぞれの根元事象の確率（起こり易さ） p_j （$j=1,2,\ldots,N$） を与えることである．それでこの根元事象の起こり易さ（確率）は現実をできるだけ反映するように決めるのだった．

しかし，この根元事象の確率 p_j はいくつかの性質を満たすべきである．まず，これは確率だから 0 と 1 の間にないといけない．更に，$\mathcal{S}$ そのものというのは全事象だから（いつでも起こる）この確率は 1 であるべし．要するに

$$0\le p_j\le 1,\qquad \sum_{j=1}^{N}p_j=1 \tag{1.2.1}$$

[4]教科書の 2.1 節の後半

であればよい，ということになる．そして，根元でない事象 $E = \{e_1, e_2, e_3, \ldots, e_m\}$ については，

$$(E\text{ の確率}) = \sum_{j=1}^{m} p_j \tag{1.2.2}$$

となるはずである．と言うのも，E とは「e_1 か，e_2 か，…，e_m の**どれか**が起こる」事象だから，それぞれの事象の確率の和になるのが自然．

これが数学での確率論の出発点である．要するに

- 標本空間 $\mathcal{S}$ 上に根元事象の確率 p_j を (1.2.1) を満たす形で与え，
- 根元事象でない一般の事象 E の確率を (1.2.2) で計算する．

それで，**このルールを満たすものを全て確率と認める**のである．(しつこいが，どのように p_j を選ぶか，は個々の問題に応じてうまく決める.)

さて，上のように決めた「それぞれの事象の確率」はどんな性質を満たしているだろうか？上では根元事象から確率を決めたが，そうでない場合 —— つまり，根元事象の和事象である色々な事象の確率から決めた方が楽な場合 —— も（後でたくさん）出てくる．そのために，(根元事象から出発しない場合にもなりたつ）抽象的な確率の性質を公理としてまとめておく．

定義 1.2.1 (確率の公理) 標本空間 $\mathcal{S}$ が与えられたとき，$\mathcal{S}$ 上の**確率**（または確率測度）とは，以下を満たす関数（数の組） P のこと：$\mathcal{S}$ の部分集合（事象） E のそれぞれについて値 $P[E]$ が定まり，かつ

1. 全ての $E \subset \mathcal{S}$ に対して $0 \le P[E] \le 1$ （確率は E を超えない）
2. $P(\mathcal{S}) = 1$ （全確率は E）
3. E_1, E_2 が**排反**，つまり「$E_1 \cap E_2 = \varnothing$」，のとき，$P\bigl[E_1 \cup E_2\bigr] = P[E_1] + P[E_2]$

なお，標本空間 $\mathcal{S}$ とその上の確率測度 P をあわせて**確率空間**と言う．

上の性質を満たしている P なら何でも確率と認めてしまおう，と言うのが数学の立場である．しつこいけども，実際にどのような P を採用するかは考えている具体的問題によって，適当に（適切に）決める．

命題 1.2.2 確率について，以下が成り立つ（ベン図を書いて意味を確認しよう）．

$$P[E^c] = 1 - P[E] \qquad (E^c\text{ は }E\text{ が起こらない事象のこと}) \tag{1.2.3}$$

$$E \subset F \quad \Longrightarrow \quad P[E] \le P[F] \tag{1.2.4}$$

$$P[E \cup F] = P[E] + P[F] - P[EF] \tag{1.2.5}$$

根元事象から考えるよりも，他の事象から考えた方が確率を割り振りやすい例として，２枚のイカサマコインを投げる場合を考えよう．2 枚のコインがあり，１枚目は表が p，裏が $1-p$ の確率で出る．２枚目は表が q，裏が $1-q$ の確率で出る，としよう．

このとき標本空間は $\{(H,H), (H,T), (T,H), (T,T)\}$ である．さて，この４つの根元事象にどのように確率を割るふるべきか，だが：１枚目と２枚目の出方は無関係と思うのが良いだろう（数学的には「独立」という；後述）．すると，

$$P[\,1\text{ 枚目が表}] = p, \qquad P[\,2\text{ 枚目が表}] = q \tag{1.2.6}$$

ととるのが良いのでは？これは根元事象の言葉では

$$P[\{(H,H),(H,T)\}] = p, \qquad P[\{(H,H),(T,H)\}] = q \tag{1.2.7}$$

と言うことになるね．後，基本的性質から

$$P[\{(T,H),(T,T)\}] = 1 - p, \qquad P[\{(H,T),(T,T)\}] = 1 - q \tag{1.2.8}$$

も言えているわけだ．でもこれだけでは４つの根元事象の確率は決まらない．実際，

$$P[\{(H,H)\}]=a,\quad P[\{(H,T)\}]=b,\quad P[\{(T,H)\}]=c,\quad P[\{(T,T)\}]=d \tag{1.2.9}$$

と書くと，上のは

$$a+b=p,\quad a+c=q,\quad c+d=1-p,\quad b+d=1-q \tag{1.2.10}$$

となって，不定方程式になる．でも，この場合はやはり余分な仮定をおくのが良いだろう．１枚目と２枚目が「独立」なのなら，

$$P[\{(H,H)\}]=P[\,1\text{ 枚目が表，}2\text{ 枚目も表}]=P[\,1\text{ 枚目が表}]\times P[\,2\text{ 枚目が表}]=pq \tag{1.2.11}$$

と考えるのがよいだろう．その他も同様に考えると，

$$P[\{(H,T)\}]=P[\,1\text{ 枚目が表，}2\text{ 枚目は裏}]=P[\,1\text{ 枚目が表}]\times P[\,2\text{ 枚目が裏}]=p(1-q) \tag{1.2.12}$$

$$P[\{(T,H)\}]=P[\,1\text{ 枚目が裏}]\times P[\,2\text{ 枚目が表}]=(1-p)q \tag{1.2.13}$$

$$P[\{(T,T)\}]=P[\,1\text{ 枚目が裏}]\times P[\,2\text{ 枚目が裏}]=(1-p)(1-q) \tag{1.2.14}$$

となる．

1.3 数の数え方の復習（高校の復習；流し読みで良い）

（始めに）以下のようなことは頭から覚え込むのではなく，自分で納得して理解するようにすべし．まず記号を導入する．

定義 1.3.1
- $n>0$ に対して， $n! := n\cdot(n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1$，また $0!=1$ と定義する．
- $0\le k\le n$ に対して，$\displaystyle\binom{n}{k} := \frac{n!}{k!(n-k)!}$ と定義し，「二項係数」と呼ぶ．
- $0\le n_i\ (i=1,2,\ldots,r)$, $\displaystyle\sum_{i=1}^{r} n_i = n$ のとき，$\displaystyle\binom{n}{n_1\,n_2\,n_3\,\cdots\,n_r} := \frac{n!}{n_1!\,n_2!\,n_3!\,\cdots\,n_r!}$ を**多項係数**と言う．

さて，上の記号は何に使うかというと：1 から n までの数字を書いた n 枚のカードがあって，これから k 枚を取り出す場合を考える．取り出し方（戻し方）に応じて，大体３とおりある．

Case 1: n 枚のカードから繰り返しを許して k 枚とり，その結果を並べる場合．この場合の結果は $(a_1,a_2,\ldots,a_k)$ と言う列になる（a_j は j 番目に出たカードの目）．ここでそれぞれの a_j は勝手に 1 から n の値をとれるので，結果の総数（場合の数）は

$$n\cdot n\cdot n\cdots n = n^k \tag{1.3.1}$$

となる．

Case 2: n 枚のカードから繰り返しを許さないで k 枚とり，その結果を並べる場合．やはり結果は $(a_1,a_2,\ldots,a_k)$ の形になるが，今回は a_j は全て別のものにならざるを得ない．a_1 は n 通り，a_2 は a_1 をよけるから $(n-1)$ 通り，と考えて行くと，結果は

$$n\cdot(n-1)\cdot(n-2)\cdots(n-k+1)=\frac{n!}{(n-k)!} \tag{1.3.2}$$

となる．高校ではこの数を ${}_nP_k$ と書いた．

Case 3: n 枚のカードから繰り返しを許さないで k 枚とるが，その順序は気にしない場合．やはり結果は case 2 のように $(a_1,a_2,\ldots,a_k)$ の形になるが，今は a_j の順序を気にしない（順序が異なっても同じものと見なす）．従って場合の数は *Case* 2 のものを 「k 個の数字を並べる並べ方」$k!$ で割ったものになる：

$$\frac{n!}{(n-k)!}\times\frac{1}{k!}=\binom{n}{k}={}_nC_k \tag{1.3.3}$$

１つだけ，これらの応用例を挙げておく．この証明は帰納法でもできるし， Case 3 の数え方を使う方法もある．

命題 1.3.2 (二項定理，高校でやったかな) $1 \le n$ では，$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k \, y^{n-k}$.

Case 4. なお，補足的に Case 3 の一般化を考えておく．n 枚のカードを，それぞれ $n_1, n_2, \ldots, n_r$ 枚のカードからなる r 個のグループに分ける場合（$\sum_{i=1}^{r} n_i = n$）．この場合はまず n 枚から n_1 枚を取り出し，次に $n - n_1$ 枚から n_2 枚を取り出し，次に $n - n_1 - n_2$ 枚から n_3 枚を取り出し... と考えて

$$\binom{n}{n_1} \times \binom{n-n_1}{n_2} \times \binom{n-n_1-n_2}{n_3} \times \cdots \times 1 = \frac{n!}{n_1!\, n_2!\, n_3! \cdots n_r!} = \binom{n}{n_1\, n_2\, n_3 \cdots n_r} \tag{1.3.4}$$

となることがわかる．

1.4 確率変数[5]

今まではランダムな事象を考えてきた（例：このクラスの学生から一人選んだら男であった，とか）．事象はそれが起こるか起こらないかの２通りしかない．しかし，実際には選ばれた標本の数値的な性質を問題にすることも多い（例：選んだ学生の身長はいくらか）．

このような問題では（我々の注目する）実験の結果が数値で表されている．つまり，実験の結果として**ランダムな数値**が出てくるわけだ．そこで，このようにランダムに値がきまる数値のことを**確率変数**と呼ぶ（ちょっとえーかげん）．

確率変数には「離散的な確率変数」と「連続な確率変数」がある．まずは簡単な「離散的」なものから考える．

離散的な確率変数とはとびとびの（有限個の）値しかとらないもので[6]，例は以下の通り．

例 1.4.A： サイコロを一回振る実験を考える．X を出た目の数とすると，X のとりうる値は $1, 2, 3, 4, 5, 6$ の６通り．また，それぞれの値をとる確率は（マトモなサイコロなら）

$$P[X=1] = P[X=2] = \ldots = P[X=6] = \frac{1}{6} \tag{1.4.1}$$

と考えるのが自然だろう．また，Y を「出た目が４以下なら 0，出た目が５以上なら 10」である確率変数とすると，Y のとりうる値は $0, 10$ で，その確率は

$$P[Y=0] = \frac{4}{6} = \frac{2}{3}, \qquad P[Y=10] = \frac{2}{6} = \frac{1}{3} \tag{1.4.2}$$

例 1.4.B： サイコロを２個振る実験を考える．Z を出た目の和とすると，Z のとりうる値は $2, 3, 4, \ldots, 12$ の１１通り．また，それぞれの値をとる確率は（マトモなサイコロなら）

$$P[Z=2] = \frac{1}{36}, \qquad P[Z=3] = \frac{2}{36} = \frac{1}{18}, \qquad \text{（場合が多すぎて書ききれない）} \tag{1.4.3}$$

などとなる．

上の例でもわかるように，離散的な確率変数を記述するには「確率変数のとりうる値」と「それぞれの値をとる確率」を全て与えれば良い．つまり，確率変数 X が $x_1, x_2, \ldots, x_n$ の値をとりうる場合，X がそれぞれの x_i をとる確率，つまり $P[X = x_i]$ （$i = 1, 2, \ldots, n$）を与えればよいわけだ．

連続的な確率変数とは文字通り，連続な値をとりうる確率変数だ．例を見るのが良いだろう．

例 1.4.C： X は区間 $[0, 1]$ 内の全ての値を，同じ確率でとりうる確率変数である．

例 1.4.D： Y はこのクラスの学生を一人選んだ場合の学生の身長である（ただし，身長はいくらでも細かく測るものとする）．

[5] 教科書の 2.2 節
[6] とびとびの値しかとらないけど，全体としては無限個の値をとりうる例もある．が，話を簡単にするため，ここはごまかした

例 1.4.E： Z は学研都市の駅で，福岡方面の地下鉄に乗る場合の待ち時間（ただし，時間を計る場合にいくらでも細かく測定するものとする）である．

例 1.4.C では，X のとりうる値は連続無限個あり，これらの確率は同じと仮定しているから，X が特定の値（例：$X = \frac{1}{2}$）をとる確率はゼロだ．(ゼロでなかったら，全確率が無限大になってしまう！)

このように，連続な確率変数を記述するには，離散的な確率変数のような $P[X = x_i]$ を与えるやり方は使えない．仕方がないので， $P[X = x_i]$ に相当するものとして，

$$P[a \le X \le b] = \int_a^b f(x)dx \tag{1.4.4}$$

のように，**確率密度関数** $f(x)$ を用いて積分の形で表すことにする（より詳しくは後述）．

例 1.4.C の場合は $f(x) = 1$ である．例 1.4.D や例 1.4.E の分布関数は厳密にはわかりそうにないが，大体の感じは書けそうだ．

1.5 確率分布[7]

さて，上で導入した確率変数を特徴付ける（定義する）最も基本的な量（函数）として，**確率分布**の概念を導入する．

離散型確率変数の場合がわかりやすいので，ここから始めよう．

定義 1.5.1 (離散型確率変数の確率函数) ある離散型確率変数 X が n 個の値 $x_1, x_2, \ldots, x_n$ を取りえて，それぞれの値をとる確率が

$$P[X = x_i] = p_i \qquad (i = 1, 2, 3, \ldots, n) \tag{1.5.1}$$

と与えられているとする．この時，各 x_i にその確率 p_i を対応させる函数 f

$$f(x_i) = p_i \qquad (i = 1, 2, 3, \ldots, n) \tag{1.5.2}$$

のことを**確率函数**という．また，x_i と p_i の組み（$i = 1, 2, 3, \ldots, n$）を**確率分布**という．

先のサイコロの例 1.4.A なら，

$$f(1) = \frac{1}{6}, \quad f(2) = \frac{1}{6}, \quad \ldots, \quad f(6) = \frac{1}{6} \tag{1.5.3}$$

ということになる．

次に，連続型の確率変数を考えよう．

定義 1.5.2 (連続型確率変数の確率密度函数) ある連続型確率変数 X に対して，以下の (1), (2), (3) を満たす函数 $f(x)$ が存在すると仮定する：

(1) 任意の x に対して $f(x) \ge 0$

(2) $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$

(3) $a \le b$ なる任意の a, b に対して，

$$P[a \le X \le b] = \int_a^b f(x)dx \tag{1.5.4}$$

この時，f を，「確率変数 X の**確率密度函数**」という．また，確率密度函数を**確率分布**ということもある．

変数変換の式（教科書 p.81）には少し注意．

[7]教科書の 2.3 節

1.6 （累積）分布函数[8]

世の中には，離散型，連続型に分類できないような確率変数もある．また，そもそも，上の定義にあった確率密度函数が存在しないような確率変数もありうる．これらを統一的に扱うために，以下の定義を導入する．

定義 1.6.1 (（累積）分布函数) 確率変数 X に対して，

$$F(x) := P[X \le x] \tag{1.6.1}$$

により定義される函数 $F(x)$ を，X の**累積分布函数**または単に**分布函数**という．
また，$F(a)$ を a における累積確率という．

実のところ，累積分布函数の方が，先に導入した確率（密度）函数よりも，基本的な量である．ただ，累積分布関数は直感的にわかりにくいかもしれないので，先に確率（密度）函数を導入した．

（例 1.4.A 続き）サイコロの場合，

- $x < 1$ では $P[X < x] = 0$ だから $F(x) = 0$
- $1 \le x < 2$ では $P[X < x] = P[X = 1] = 1/6$ なので，$F(x) = 1/6$
- $2 \le x < 3$ では $P[X < x] = P[X = 1 \text{ または } x = 2] = 2/6 = 1/3$ なので，$F(x) = 1/3$．以下同様に，
- $3 \le x < 4$ では $F(x) = 3/6$，$4 \le x < 5$ では $F(x) = 4/6$，$5 \le x < 6$ では $F(x) = 5/6$，
- 最後に，$6 \le x$ では全確率になるので $F(x) = 1$．

（例）連続型確率変数 X の確率密度函数を $f(x)$ とすると

$$F(x) = \int_{-\infty}^{x} f(y)\, dy \tag{1.6.2}$$

である．

命題 1.6.2 累積分布函数について，以下が成り立つ：
(1) 任意の実数 x に対して $0 \le F(x) \le 1$
(2) F は広義単調増加，つまり $x < y$ ならば $F(x) \le F(y)$
(3) F は右連続，つまり $\displaystyle\lim_{y \to x+0} F(y) = F(x)$
(4) $F(-\infty) = 0$ かつ $F(\infty) = 1$

1.7 期待値と分散[9]

確率変数が与えられたとき，この確率変数の分布をどのように特徴づけたらよいだろうか？もちろん，完全に特徴づけるには，確率分布や累積分布函数を考えれば良い．しかしこれは一般に大変すぎるし，そもそも，このようにすべてを知ったとして，分布の特徴がつかめるとは限らない．そうではなくて，**もっと少ない情報量で分布の特徴を捉える**ことを考えたい．その代表的なものが「期待値（平均値）」と「分散」「標準偏差」であり（この節）,「メジアン」「モード」（次節）である．

定義 1.7.1 離散的な確率変数 X が $x_1, x_2, \ldots, x_n$ の値をとり，その確率が

$$P[X = x_i] = p_i \qquad \Bigl(\text{もちろん}, \sum_{i=1}^{n} p_i = 1\Bigr) \tag{1.7.1}$$

[8] 教科書の 2.4 節
[9] 教科書の 2.5 節

と与えられているとする．このとき，X **の期待値**（平均値）を

$$E[X] := \langle X \rangle := \sum_{i=1}^{n} x_i \, p_i \tag{1.7.2}$$

により定義する．(数学では $E[X]$ の記号を，物理などでは $\langle X \rangle$ の記号を用いることが多い．）また，X **の分散**を

$$\mathrm{Var}[X] := E\Big[\big(X - E[X]\big)^2\Big] = E\big[X^2\big] - E[X]^2 = \big\langle X^2 \big\rangle - \langle X \rangle^2 = \Big\langle \big(X - \langle X \rangle\big)^2 \Big\rangle \tag{1.7.3}$$

により定義する．その平方根

$$\sigma := \sqrt{\mathrm{Var}[X]} \qquad （これによると \quad \mathrm{Var}[X] = \sigma^2 \quad となる）$$

を X の**標準偏差**と呼ぶ．

期待値とは，要するに平均値（ただし，p_i の重みを用いた加重平均）のことであり，確率変数の分布の「中心」を表す（どのような意味で中心かは要注意）．
分散とは平均からのズレ（の２乗）の平均だから，分散の平方根（標準偏差）が分布の「拡がり」を表す．

（少し脱線）事象 G の確率を期待値の形で書くことができる．すなわち，関数 $I[G]$ を

$$I[G] := \begin{cases} 1 & (G\,が起こるとき) \\ 0 & (G\,が起こらないとき) \end{cases} \tag{1.7.4}$$

として定義すると，

$$P[G] = E[\,I[G]\,] = \langle I[G] \rangle \tag{1.7.5}$$

となる．つまり，<u>F の起こる確率は関数 $I[G]$ の期待値</u>なのである．

離散的な場合と同じく，連続な確率変数に対しても期待値や分散を定義する．

定義 1.7.2 連続な確率変数 X（その確率密度関数は $f(x)$）に対しては，(1.7.2) の代わりに X **の期待値**を

$$E[X] := \langle X \rangle := \int_{-\infty}^{\infty} x \, f(x) \, dx \tag{1.7.6}$$

により定義する．また，X **の分散**を

$$\mathrm{Var}[X] := E\Big[\big(X - E[X]\big)^2\Big] = E\big[X^2\big] - E[X]^2 = \big\langle X^2 \big\rangle - \langle X \rangle^2 = \Big\langle \big(X - \langle X \rangle\big)^2 \Big\rangle \tag{1.7.7}$$

により定義する．その平方根

$$\sigma[X] := \sqrt{\mathrm{Var}[X]} \qquad （これによると \quad \mathrm{Var}[X] = \sigma^2 \quad となる）$$

を X の**標準偏差**と呼ぶ．

命題 1.7.3 確率変数 X の期待値と分散は以下の関係を満たす（a, b は任意の実数）：

$$E[aX + b] := a \, E[X] + b \tag{1.7.8}$$

$$\mathrm{Var}[aX + b] = a^2 \, \mathrm{Var}[X], \qquad \sigma[aX + b] = |a| \, \sigma[X] \tag{1.7.9}$$

命題 1.7.4 (Markov の不等式と Chebyshev の不等式) 確率変数 X に対して以下が成り立つ．
(1) X が非負の値しかとらないとき，任意の $a>0$ に対して

$$P\Big[\,X\geq a\,\Big]\leq\frac{E[X]}{a}\tag{1.7.10}$$

(2) 任意の確率変数（ただし，その期待値 $\mu[X]$ と分散 Var は有限とする）に対し，任意の $a>0$ に対して

$$P\Big[\,\big|\,X-\mu[X]\,\big|\geq a\,\Big]\leq\frac{\mathrm{Var}[X]}{a^2}\tag{1.7.11}$$

これらの不等式は,「平均や分散の値」から「その確率変数が（平均から）大きく外れた値をとる確率」を見積もる不等式である．大抵，かなり損をした評価にはなるが，最初の出発点としては役に立つ．

1.8 メジアンとモード（簡単に）[10]

定義 1.8.1 確率変数 X のメジアン m とは,

$$P[X\leq m]\geq\frac{1}{2}\quad\text{かつ}\quad P[X\geq m]\geq\frac{1}{2}\tag{1.8.1}$$

となるような数 m のことである．
また，確率変数 X のモード m とは，確率函数や確率密度関数の値が最大になる値 m のことである．

（注意）メジアンやモードは複数存在することもある（特に離散分布の場合）．

メジアンやモードは，確率変数 X の「外れ値」（分布の端の方の値）にあまり影響を受けず,「分布の真ん中付近」や「確率の一番大きいところ」を表すのに適している．

2 多次元確率分布

2.1 2 次元確率分布[11]

さて，確率変数が 2 つある場合を考えよう[12]．まずは離散的な場合から始める．今，確率変数 X が値 $x_1,x_2,\ldots,x_n$ をとり，確率変数 Y が値 $y_1,y_2,\ldots,y_m$ をとるとする．これらがそれぞれの値をとる確率は

$$P[X=x_i\text{ かつ }Y=y_j]=p_{ij}\tag{2.1.1}$$

であるとしよう．このとき

定義 2.1.1 上の確率 p_{ij} に対して

$$f(x,y)=p_{ij}\qquad(x=x_i\text{ かつ }y=y_j\text{ の時})\tag{2.1.2}$$

となる函数 f を，2 次元確率変数 (X,Y) の**同時確率函数**という．また，(x_i,y_j) と p_{ij} の組みを 2 次元確率変数 (X,Y) の確率分布という．

このとき，Y の値は気にしないで，X のみの分布に着目すると，

$$P[X=x_i]=\sum_{j=1}^{m}P[X=x_i\text{ かつ }Y=y_j]=\sum_{j=1}^{m}p_{ij}\tag{2.1.3}$$

[10] 教科書の 2.6 節
[11] 教科書 3.1 節
[12] 3 つ以上ある時も同様に話ができるが，これは教科書に従って，後で扱う

となる．これを X の**周辺分布**という．同様に，Y のみの分布は

$$P[Y=y_j]=\sum_{i=1}^{n}P[X=x_i \text{ かつ } Y=y_j]=\sum_{i=1}^{n}p_{ij} \tag{2.1.4}$$

で与えられる．

X,Y が連続分布の場合は，上の定義などは以下のようになる．

定義 2.1.2 連続的確率変数 X,Y に対して，以下を満たす函数 $f(x,y)$ が存在するとする：
(1) $f(x,y)\geq 0$
(2) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\,dx\,dy=1$
(3) $a\leq b$ かつ $c\leq d$ なる a,b,c,d に対して

$$P[a\leq X \text{ かつ } c\leq Y\leq d]=\int_a^b dx\int_c^d f(x,y) \tag{2.1.5}$$

このとき，$f(x,y)$ を，X,Y の**同時確率密度函数**という．また，

$$f_X(x):=\int_{-\infty}^{\infty}dy\,f(x,y) \qquad f_Y(x):=\int_{-\infty}^{\infty}dx\,f(x,y) \tag{2.1.6}$$

をそれぞれ，X の**周辺分布密度函数**，Y の**周辺分布密度函数**という．

変数変換については，教科書の定理 3.1 を参照．

2.2 独立性と条件付き確率[13]

ここでは「独立性」「条件付き確率」の概念を導入する．教科書と少し順序が前後するが，言ってることは同じである．

（あ）まず，事象の独立性と条件付き確率について

定義 2.2.1 (独立な事象) 確率空間 $(\mathcal{S},P)$ 中の事象 E,F が，

$$P[E\cap F]=P[E]\,P[F] \qquad \text{(E と F が起こる確率は E, F それぞれが起こる確率の積)} \tag{2.2.1}$$

を満たすとき，F と E は独立な事象であると言う．

日常言語で言えば，E と F が独立とは，E と F の**起こり方が無関係**（F が起こっても起こらなくても，E の起こり方には影響がない）と言う場合にあたる（この事情は以下の「条件付き確率」を考えた方がわかりやすいかも）．

E,F が独立でない場合は F の起こり方が E の起こり方に影響しているわけだ．影響の度合いを測るため，「条件付き確率」を導入する．

定義 2.2.2 (条件付き確率) 確率空間 $(\mathcal{S},P)$ 中の事象 E,F を考える．$P[F]\neq 0$ の場合に，

$$P[\,E\,|F\,]:=\frac{P[E\cap F]}{P[F]} \tag{2.2.2}$$

をF の下で E が起こる条件付き確率と言う．（ベン図で感じをつかもう！）

[13] 教科書 3.2 節

註 2.2.3 E と F が独立の場合はもちろん，$P[E|F] = P[E]$ となる．これがまさに，E と F が独立なら,「F が起こっても起こらなくても E の起こる確率は変わらない」という意味である．

さて，$P[E]$ そのものよりも $P[E|F]$ と $P[F]$ の方が良くわかる場合が往々にしてある．この場合（条件付き確率の定義からすぐに出てくる式）

$$P[E] = P[E|F]\,P[F] + P[E|F^c]\,P[F^c] \tag{2.2.3}$$

を用いて $P[E]$ を計算することができる．条件付き確率そのものに興味がある場合もあるが，このような計算や後述のベイズ推定において，**条件付き確率を計算の中間段階として利用する**場合も非常に多い．

例 2.A： 袋の中に赤玉が１０個，白玉が３個，黒玉が４個入っている．目をつぶって１つ取り出すとき：

1. 白が出る確率は？
2. 「出た玉は赤ではない」ことがわかった場合，取り出した玉が白である確率は？

例 2.B： 男と女の生まれる確率は $\dfrac{1}{2}$ ずつとする．A さんちには子供が二人いる.（まあ，探偵がこの家のことをいろいろと調べていると思って下さい.）

1. 二人とも男の子である確率は？
2. 「少なくとも一人が男の子だとわかっている」場合，二人とも男の子である確率は？

例 2.C： 袋の中に赤サイコロが１個，白のサイコロが２個入っている．白の方は普通の１〜６が書かれたサイコロだが，赤の方は１，２，３が２つずつ書かれている変態サイコロである．この袋から目をつぶってサイコロを一つ取り出して転がした．１の目が出る確率を求めよ．

例 2.D： （これはあくまで例．深読みはしないように）．僕はある大学で 200 人の学生に物理を教えているが，そのうちの４割は高校で物理を履修しており，残りの６割は未履修である．過去の経験から，僕の物理の講義に受かる確率は,「高校での物理既習者では 0.9，物理未修者では 0.3」と予測される．以上から，僕の物理の講義に受かる学生は 200 人中何人くらいと考えられるか？

例 2.E： 2 個のサイコロ（6 つの面が 1/6 の確率ででるものとする）を一回ずつ転がすことを考える．2 つのサイコロの目が異なる場合，少なくとも一方が 6 をだした確率はいくらか？

(い) 続いて，確率変数の独立性と条件付き確率について

定義 2.2.4 (独立な確率変数) 確率変数 X と Y が任意の $A, B \subset \mathbb{R}$ に対して

$$P[X \in A \text{ かつ } Y \in B] = P[X \in A]\,P[Y \in B] \tag{2.2.4}$$

を満たすとき， X と Y は**独立**な確率変数と言う．

定義 2.2.5 (条件付き確率分布) 離散型確率変数 X と Y がそれぞれ $x_1, x_2, \ldots, x_m$ と $y_1, y_2, \ldots, y_n$ の値を取るものとする．この時，$Y = y_j$ の条件の下で $X = x_i$ となる確率分布を

$$P[X = x_i \,|\, Y = y_j] = \frac{P[X = x_i \text{ かつ } Y = y_j]}{P[Y = y_j]} \tag{2.2.5}$$

とし，条件付き確率分布と呼ぶ．
また，X, Y が連続型確率分布の場合には，$Y = y$ の下での X の条件付き確率密度函数を

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} \tag{2.2.6}$$

として定義する．

問 2.2.6 さいころを続けて n 回投げることを考える．この n 回のうちに出る**異なった目**の数を N_n としよう．N_n の期待値はいくらか？(注：例えば 5 回投げたとき，$(1,3,2,1,1)$ とでたら，異なった目は $1,2,3$ なので，$N_5=3$ と言うこと.)

問 2.2.7 駅の切符売り場や銀行での行列の作り方を考える．窓口は M 個あり，全体で N 人のお客が並んでいる．このとき，

1. 一列待ち：お客を一列に並べておいて，開いた窓口へ誘導していく
2. M 列待ち：お客を勝手に，それぞれの窓口に並ばせる

のどちらが良い（苦情が少ない）だろうか．待ち時間の期待値や分散を考えてみよう．

3 つ以上の確率変数がある場合も，同様に議論できるが，一言だけ注意を．確率変数 $X,Y,\ldots,Z$ が**独立**であるとは，これらの確率変数の分布が，それぞれの確率変数の周辺分布の積に分解することをいう．つまり，離散の場合に書けば，

$$P[X=x_i,Y=y_j,\ldots,Z=z_k]=P[X=X_i]\,P[Y=y_j]\,\ldots\,P[Z=z_k] \tag{2.2.7}$$

となることをいう．

2.3 ベイズの公式と推定[14]

ここでは条件付き確率の，今までとは少し違った解釈を考えよう．これまでの解釈では $P[F\,|E]$ は「E が起こったという条件の下で F が起こる確率」だったが，新しい解釈として「E が起こったという情報を知った後で F の確率をどのように設定する（見積もる）のがよいか」を示す式とも考えられる．この節では，このような解釈に基づく推論を考える．

まずは，この節の議論の元になる公式を述べよう．

命題 2.3.1 (Bayes の公式) 確率空間 $(\mathcal{S},P)$ を考える．すると，$E,F\subset\mathcal{S}$ に対して

$$P[F\,|\,E]=\frac{P[F\cap E]}{P[E]}=\frac{P[E\,|F]\,P[F]}{P[E\,|F]\,P[F]\,+\,P[E\,|F^c]\,P[F^c]} \tag{2.3.1}$$

が成立する．事象が 3 つ以上の場合に一般化すると，事象 F_i （$i=1,2,\ldots,k$）が互いに排反（$F_i\cap F_j=\varnothing$ for $i\neq j$），かつ $\displaystyle\bigcup_{i=1}^{k}F_i=\mathcal{S}$ を満たすときは，

$$P[F_j\,|E]=\frac{P[F_j\cap E]}{P[E]}=\frac{P[E\,|F_j]\,P[F_j]}{\displaystyle\sum_{i=1}^{k}P[E\,|F_i]\,P[F_i]} \tag{2.3.2}$$

が成立する．

上の式は単に条件付き確率の定義

$$P[F\,|E]=\frac{P[F\cap E]}{P[E]} \tag{2.3.3}$$

と (2.2.3) の一般化

$$P[E]=\sum_{i=1}^{k}P[E\,|F_i]\,P[F_i] \tag{2.3.4}$$

[14] 教科書の 3.3 節

を組み合わせた**だけ**のものであるから**無理に暗記しない**方がよい．$P[E]$ の計算に (2.3.4) が不可欠な事例が多々あるから，応用上は非常に役立つ．また，解釈としても，左辺は E で条件づけているのに，右辺は F_i で条件付けていて，**条件付けの立場が逆転**しているように見えるのも面白い．

残念ながら，時間の関係から，ベイズの公式を用いた面白い問題については詳しく述べることはできない．以下に過去の講義で用いた例題をいくつか挙げるにとどめる．

まずは条件付き確率を使った全確率の計算

問 2.3.2 僕はある大学で 200 人の学生に物理を教えている．学生の

- ４割（$= r_1$）は高校で物理 I, II を履修
- ２割（$= r_2$）は高校で物理 I のみを履修
- 残りの４割（$= r_0$）は物理を未履修

である．過去の経験から，僕の物理の講義に受かる確率は，

- 物理 I, II の既習者では 0.9（$= p_1$），
- 物理 I のみの既習者では 0.6（$= p_2$），
- 未修者では 0.3（$= p_0$）

と予測される．以上から，僕の物理の講義に受かる学生は 200 人中何人くらいと考えられるか？

つづいてベイズ型の推定について

問 2.3.3 上の例 2.D や上の問 2.3.2 と同じ状況を考える．僕のクラスの A 君は健闘むなしく，僕の物理の単位が取れなかった．A 君は高校で物理（I まで，II まで？）を履修してきたのだろうか？(物理 II まで履修して来た確率はどのくらいと考えるのが妥当か？)

言うまでもないことであるが，上のような問いかけは余りにも安易である．単位が取れる — より正確には講義内容が身につく — かどうかは多分に本人のやる気や努力によるわけで，高校時代にどれくらいやったかで単純に推し量ることはできない．この問では現実的でないくらいの非常な単純化を行っていることには注意されたい．(将来，実際にこのような手法を用いる際にはくれぐれも単純化のしすぎに注意！)

上の２問が典型的な問題である．以下では数学的には同じ構造であるが応用としては異なった場面を述べる．

問 2.3.4 （再録）かなり稀な病気の血液テストを考える．このテストの誤差の入り方は，

- この病気にかかっている人をテストすると $(1-p)$ の確率で「病気だ」と正しく判定するが，残りの p の確率で見逃してしまう
- 健康な人をテストすると $(1-q)$ の確率で「健康だ」と正しく判定するが，残りの q では（健康なのに）「病気だ」と言ってしまう

となっている．さて，独立な疫学的調査から病気の人の割合は r であるだろうとわかっている（p, q, r はすべてゼロに近いがゼロではない）．

僕の検査結果は陽性（病気だ）だった．僕が本当に病気である確率，健康なのに間違って病気と診断された確率，をそれぞれ求めよ．

問 2.3.5 ○○科目の期末試験は（数学ではあり得ないことに）○×式の問題で，各問は m 個の選択肢から一つ正解を選ぶ形になっています．A 君はかなり怠けていたので，実力で（つまり，まぐれ無しで）正しく答えられる確率は各問毎に p であると思われます（$P < 1/2$）．答を正しく知っているときは勿論，A 君はその正解を答えますが，答がわからないときはヤケクソで m 個の答から等確率で 1 個を選びます．さて，

1. ある一問に対して（まぐれであれ何であれ）A 君が正解を答える確率はいくらでしょう？
2. ある一問をテストしてみたところ，A 君は正解を答えました．このとき，A 君が実際に答を知っていた（まぐれ当たりではない）確率はいくらでしょう？
3. 以上の結果を解釈せよ． どのような p, m の値の場合に「マグレ当たり」が多くなるか，考えてみよう．

問 2.3.6 行方不明の飛行機を捜索中である．現在，墜落した可能性のあるのは 1, 2, 3 の 3 地区に限ること，およびこれらの 3 地区に墜ちている確率は等しい（つまり 1/3）こと，までは絞り込んだ．これから捜索に入るが，厳しい気象条件のため，確実に見つけられる保証はない — 実際に i-地区に墜ちていたとしても，確率 p_i で見逃すだろうと思われる（$p_i \ll 1$）．

まず 1-地区を捜索したところ，飛行機は見つからなかった．この事実から，i-地区に墜ちている確率を推定せよ（$i = 1, 2, 3$）．

問 2.3.7 (Laplace) $i = 0, 1, 2, \ldots, k$ と（非常に小さな）印が付けられた $(k + 1)$ 個のコインが壺に入っている．これらは非常にいびつなコインで，i 番目のコインを投げたときに表が出る確率は i/k となるように調節されている．目隠しをしたままこの壺から一枚のコインを選んで実験をする．以下の問いに答えよ．

1. 取り出したコインを一回投げたところ，表が出た．このコインが i 番目のコインである確率はいくらか？($i = 0, 1, 2, \ldots, k$)
2. 取り出したコインを更に投げ続け，合計 n 回投げた．結果は全て表だった．このコインが i 番目のコインである確率はいくらか？($i = 0, 1, 2, \ldots, k$)
3. 取り出したコインを更にもう一回（つまり通算で $(n + 1)$ 回目）投げる事にした．このとき，やはり表が出る確率はいくらか？
4. 上の小問 2, 3 の答はそれほど簡単にならなかったかも知れない．そこでこれらの確率が $k \to \infty$ の極限でどうなるか，求めてみよう．結果は直感と合うだろうか？

（注）この問では，コインは最初に一枚取り出したら，同じ物を使い続ける．コインを何回か投げるとき，一回ごとの結果は独立だとする．また，コインについている印は大変小さいので，取り出したコインがどれかは見ただけではわからないものとする.（そうでないと，小問 2, 3 が面白くない.）

問 2.3.8 3 人の射撃手（1, 2, 3）が 200m 離れた，同じ的を狙う．今までの練習成績から，射撃手 i が一発で的に当てる確率はそれぞれ p_i と考えられる（$i = 1, 2, 3$）．さて，3 人が一発ずつ撃ったところ，的には**丁度一発だけ**当たっていた．この当たった一発が射撃手 i のものである（つまり，他の二人ははずした）確率について，以下の問いに答えよ．

1. まず，計算を始める前に，直感的に答を推定してみよう．
2. では，講義での説明に基づき,「正しく」計算してみよう．
3. 2 の結果は直感とあっているか？例えば，$p_1 = 0.2, p_2 = 0.4, p_3 = 0.6$ として，射撃手 1 が当てた確率はいくらになっているか？(勿論，1, 2 の答が一緒になった人は立派なものである．僕にはこの結果は意外だったけどね.)

2.4 期待値と分散[15]

2 つの確率変数 X, Y がある時，X, Y の勝手な関数 $\varphi(X, Y)$ は確率変数になる．その期待値と分散は以下のように定義する．

[15] 教科書の 3.4 節

定義 2.4.1 (期待値) $\varphi(X,Y)$ の期待値は

$$E[\varphi(X,Y)] = \begin{cases} \displaystyle\sum_{i,j} p_{ij}\varphi(x_i,y_j) & \text{(離散型)} \\ \displaystyle\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy\, f(x,y)\,\varphi(x,y) & \text{(連続型)} \end{cases} \tag{2.4.1}$$

として定義する．

定義 2.4.2 (分散) $\varphi(X,Y)$ の分散は

$$\mathrm{Var}[\varphi(X,Y)] = E\Big[\big\{\varphi(X,Y) - E[\varphi(X,Y)]\big\}^2\Big] \tag{2.4.2}$$

として定義する．

これらには以下の簡単な性質がある．

命題 2.4.3 期待値や分散は以下の性質を満たす（a,b,c は定数）：
(0) $\mathrm{Var}[\varphi(X,Y)] = E\big[\{\varphi(X,Y)\}^2\big] - \big\{E[\varphi(X,Y)]\big\}^2$
(1) $E[aX+bY+c] = a\,E[X] + b\,E[Y] + c$
(2) X,Y が独立の場合，1 変数函数 g,h に対して，$E[g(X)\,h(Y)] = E[g(X)]\,E[h(Y)]$
(3) X,Y が独立の場合，$\mathrm{Var}[aX+bY] = a^2\,\mathrm{Var}[X] + b^2\,\mathrm{Var}[Y]$

さて，2 つの確率変数の関係（独立性など）を特徴付けるには，もちろん，その確率分布（密度）を知るのが一番である．しかし，1 つの確率変数の場合と同じく，より少ない量で,（不完全ながら）特徴づけを行いたいことが多い．そのために以下の「共分散」を定義する．

定義 2.4.4 (共分散) 同時確率変数 X,Y の共分散 $\mathrm{Cov}(X,Y)$ を

$$\mathrm{Cov}(X,Y) = E\Big[\big\{X - E[X]\big\}\big\{Y - E[Y]\big\}\Big] \tag{2.4.3}$$

として定義する．またその**相関係数** $\rho(X,Y)$ を，

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\,\sigma(Y)} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}[x]\,\mathrm{Var}[Y]}} \tag{2.4.4}$$

により，定義する．

命題 2.4.5 共分散などは以下の性質を満たす（a,b,c は定数）：
(1) $\mathrm{Cov}(X,Y) = E[X\,Y] - E[X]\,E[Y]$
(2) $\mathrm{Cov}(X,X) = \mathrm{Var}[X]$
(3) $\mathrm{Cov}(X+a, Y+b) = \mathrm{Cov}(x,Y)$
(4) X,Y が独立の場合，$\mathrm{Cov}(X,Y) = 0$
(5) $\mathrm{Var}[a\,X + b\,Y + c] = a^2\mathrm{Var}[X] + 2ab\,\mathrm{Cov}(x,y) + b^2\mathrm{Var}[Y]$

（注意）「独立なら共分散がゼロ」は正しい（上の (4)）が，逆は一般には成り立たない．

2.5 n この確率変数[16]

3 個以上の確率変数も，2 個の確率変数の場合と同じようにして考える．大抵は 2 この場合の自然な拡張なので，「教科書の 3.5 節」を参照することにし，この講義ノートではほぼ省略したい．ただし,「独立」の概念だけは念を押しておいた方が良いかもしれない．

定義 2.5.1 (確率変数の独立性) n この同時確率変数 $X_1, X_2, \ldots, x_n$ が独立であるとは以下の場合をいう：
(1) 離散型の時，

$$P[X_1 = x_1 \text{ かつ } X_2 = x_2 \text{ かつ } \cdots \text{ かつ } X_n = x_n] = P[X_1 = x_1]\, P[X_2 = x_2] \cdots P[X_n = x_n] \tag{2.5.1}$$

(2) 連続型の時：

$$f(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1)\, f_{X_2}(x_2) \cdots f_{X_n}(x_n) \tag{2.5.2}$$

ここで，$f_{X_j}(x_j)$ とは，確率変数 X_j の周辺分布密度函数である．

3 つ以上の確率変数の場合，その 2 つずつが（ペアで）独立であっても，3 つ全部が（上の定義の意味で）独立とは限らない．

2.6 大数の法則[17]

さて，以上の準備の元に，漸く，この講義前半部のメインテーマ「たくさんの確率変数がある場合の極限定理」の最初の一つを述べることができる．

定理 2.6.1 (大数の弱法則) n この確率変数 $X_1, X_2, \ldots, x_n$ は独立で，さらに，共通の期待値と標準偏差を持つものとする．つまり，

$$E[X_i] = \mu, \qquad \sqrt{\mathrm{Var}[X_i]} = \sigma \qquad (i = 1, 2, 3, \ldots, n) \tag{2.6.1}$$

このとき，任意の $\epsilon > 0$ に対して，

$$\lim_{n\to\infty} P\Bigl[\, \Bigl| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \Bigr| > \epsilon \,\Bigr] = 0 \tag{2.6.2}$$

が成立する．

この結果は，直感的にも納得しやすい．例えば，X_i を「コインを何回も投げる時に，i 回目に面が出たら 1，裏が出たら 0」であるような確率変数としよう．この確率変数 $X_1, X_2, \ldots$ は互いに独立で，確率 1/2 で $X_i = 0, 1$ の値をとると考えるのが自然だ．従って，

$$\mu = \frac{1}{2}, \qquad \sigma = \frac{1}{2} \tag{2.6.3}$$

と考えられる．この場合，n 回投げれば，だいたい $n/2$ 回くらい，面が出るだろうと期待するわけだが，上の定理は

$$\lim_{n\to\infty} P\Bigl[\, \Bigl| \sum_{i=1}^{n} X_i - \frac{n}{2} \Bigr| > \epsilon \,\Bigr] = 0 \tag{2.6.4}$$

を意味しており，この直感を裏付けている．

3 二項分布と正規分布[18]

ここのところ，教科書では 1 章を設けており，この講義ノートでもそれに従うが，あまり細かくやっても仕方ないので，簡単にすませる．

[16] 教科書の 3.5 節
[17] 教科書の 3.6 節
[18] 教科書の 4 節

3.1 二項分布[19]

定義 3.1.1 (二項分布) （パラメーター n, p の）二項分布とは，$X = 0, 1, 2, \ldots, n$ の値を，確率

$$P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \tag{3.1.1}$$

でとるような確率変数 X の確率分布のことである．この分布を $\mathrm{Bin}(n,p)$ と略記することが多い．

（注意）上の確率変数 X の意味づけとしては例えば以下のようなものがある：

> 面が出る確率が p であるコインを n 回投げるが，n 回の結果はもちろん，独立なものとする．この時，「n 回のうちちょうど k 回だけ表」はまさに上の確率 $P[X = k]$ になる．つまり，上の X とは,「コインを n 回投げたうち，面がでた回数」を表す確率変数と解釈できる．

より一般に，別にコインでなくても,「『成功』の確率が p,『失敗』の確立が $(1-p)$，である実験を独立に n 回行う場合の『成功』の回数」を X とすると，これはまさに上の二項分布に従うことになる．世の中にはこのような「2 択問題」を何回も繰り返すことは多いので，二項分布はそのような場合の典型例になっている．

簡単に計算できるけど重要なこと：上の二項分布の期待値は $E[X] = np$，分散は $\mathrm{Var}[X] = np(1-p)$ である．

3.2 正規分布[20]

定義 3.2.1 (正規分布) 正規分布とは一般に（μ を実数，σ は正の数として）

$$P[a \le X \le b] = \int_a^b \frac{1}{\sqrt{2\pi}\,\sigma} \exp\Bigl[-\frac{1}{2}\bigl(\frac{x-\mu}{\sigma}\bigr)^2\Bigr]\,dx \tag{3.2.1}$$

を満たすような確率変数 X の確率分布のことを言う．(これは $N(\mu, \sigma^2)$ と書かれる．) また，上のような分布をもった確率変数 X は正規分布に従う確率変数という．

これから学んでいくように，正規分布は様々なところに顔を出す．特に,「ある確率変数が正規分布とは程遠くても，その（適度にスケールした）平均は，近似的に正規分布に従う」ことが頻繁に起こる（中心極限定理）．そのため，正規分布は，この世の中を統一的に理解するために大変に重要な役割を果たし，統計学の基礎づけに関しても最重要の定理となっている．

簡単に計算できるけど重要なこと：t 上の正規分布の期待値は μ，分散は σ^2，標準偏差は σ である．

特に，$\mu = 0, \sigma = 1$ の正規分布を「標準正規分布」とよぶ．通常

$$\Phi(x) := \int_{-\infty}^x \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \tag{3.2.2}$$

と書く．以下に $1 - \Phi(x) = \displaystyle\int_x^\infty \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy$ のいくつかの値を載せておく：

x	0	1	1.645	1.960	2	2.326	2.576	3	4
$1-\Phi(x)$	$\frac{1}{2}$	0.1587	$\frac{1}{20}$	$\frac{1}{40}$	0.02275	$\frac{1}{100}$	$\frac{1}{200}$	1.350×10^{-3}	3.167×10^{-5}

また，正規分布はその期待値 μ の周りに集まっている分布であるが，その集まり具合は以下のようになっている：

$$P[\,|X-\mu| \le \sigma\,] = 0.6827, \qquad P[\,|X-\mu| \le 2\sigma\,] = 0.9545, \qquad P[\,|X-\mu| \le 3\sigma\,] = 0.9973 \tag{3.2.3}$$

[19] 教科書の 4.2 節
[20] 教科書の 4.3 節

逆にいうと，たとえば $|X-\mu|$ が 2σ より大きくなる確率は $1-0.9545=0.0455$ 程度であって，非常に小さい．

さて，積分の変数変換を用いると，一般の正規分布の分布確率を標準正規分布の分布確率から求めることができる．つまり，X が $N(\mu,\sigma^2)$ に従うときに，新しい確率変数

$$Z := \frac{X-\mu}{\sigma} \tag{3.2.4}$$

を定義すると Z が標準正規分布になることが容易にわかる．もちろん，この場合 X と Z のズレを考慮して

$$P[a \le X \le b] = P\Big[\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\Big] \tag{3.2.5}$$

とやる必要はあるが．

ともかく，このようなわけで，いろいろある正規分布は，標準正規分布になおして計算できる．

3.3 正規分布と二項分布の関係[21]

さて，ある状況の下では，二項分布を正規分布で近似できる．これは次節の「中心極限定理」の例にもなっている．

命題 3.3.1 パラメーター (n,p) の二項分布は，n が大きい時，平均 np，分散 $np(1-p)$ の正規分布で近似できる．より正確には，パラメーター (n,p) に従う確率変数を X とすると，$n\to\infty$ の場合，

$$P[a \le X \le b] \approx \int_a^b \frac{1}{\sqrt{2\pi}\,\sigma} \exp\Big[-\frac{1}{2}\big(\frac{x-\mu}{\sigma}\big)^2\Big]\,dx \qquad \text{ここで} \qquad \mu = np,\ \ \sigma = \sqrt{np(1-p)} \tag{3.3.1}$$

が成立する．

上の定理に出ている np と $np(1-p)$ はそれぞれ，この二項分布の平均と分散である．つまり上の命題は，(n が大きい時)「二項分布は，それと平均と分散が同じ正規分布で近似できる」ことを意味している．

問題 3.3.2 マトモな硬貨（表と裏が同じ確率で出る硬貨）を N 回投げる．表の出る回数が投げた回数の 49% から 51% に入る確率を，中心極限定理を用いて考えたい．$N=100,1000,10000$ に対して，この確率がどのような積分で表されるか，求めよ．(注：積分そのものの値は計算できないと思うので，やらなくて良い．)

問題 3.3.3 問 3.3.2 の続き．今度は「表の出る回数が投げた回数の 49% から 51% にほとんど確実に入る」ような N を求めたい．「ほとんど確実」と言うのはいい加減な書き方だから，具体的に「表の出る回数が投げた回数の 49% から 51% に入る確率が 0.95 以上になる」ような，そんな N を求めよ．

中心極限定理の使い方について．

問 3.3.2. 中心極限定理を使うには，まず Z_N を作らないといけない．i 回目に表が出れば $X_i=1$，裏が出れば $X_i=0$ とすると，$S_N=\sum_{i=1}^N X_i$ と書けるから，今までに考えてきた形である．さて，

$$\langle X_i\rangle = \frac{1}{2}, \qquad \mathrm{Var}X_i = \frac{1}{4}, \qquad \sigma = \frac{1}{2} \tag{3.3.2}$$

であるから，中心極限定理にでてくる Z_N は

$$Z_N = \frac{S_N - \frac{N}{2}}{\sqrt{\frac{1}{4}N}} = \frac{2S_N - N}{\sqrt{N}} \tag{3.3.3}$$

[21]教科書の 4.4 節

となっている．さて，表が 49% から 51% 出る，と言うことは

$$0.49 \le \frac{S_N}{N} \le 0.51 \qquad \Longleftrightarrow \Bigl|\frac{S_N}{N} - \frac{1}{2}\Bigr| \le \frac{1}{100} \qquad \Longleftrightarrow \bigl|Z_N\bigr| \le \frac{\sqrt{N}}{50} \tag{3.3.4}$$

と言うことだ．だから，中心極限定理を少しええ加減に使うと，この確率は

$$P\Bigl[0.49 \le \frac{S_N}{N} \le 0.51\Bigr] = P\Bigl[\bigl|Z_N\bigr| \le \frac{\sqrt{N}}{50}\Bigr] \approx \int_{-\frac{\sqrt{N}}{50}}^{\frac{\sqrt{N}}{50}} e^{-z^2/2}\frac{dz}{\sqrt{2\pi}} \tag{3.3.5}$$

となるわけだ．（ヤヤコシイが，積分の上下は $\frac{\sqrt{N}}{50}$.）N に具体的な数を入れると，

$$N = 100 \text{ なら} \quad \int_{-1/5}^{1/5} e^{-z^2/2}\frac{dz}{\sqrt{2\pi}} \approx 0.1585, \tag{3.3.6}$$

$$N = 1000 \text{ なら} \quad \int_{-\sqrt{10}/5}^{\sqrt{10}/5} e^{-z^2/2}\frac{dz}{\sqrt{2\pi}} \approx 0.4729, \tag{3.3.7}$$

$$N = 10000 \text{ なら} \quad \int_{-2}^{2} e^{-z^2/2}\frac{dz}{\sqrt{2\pi}} \approx 0.9545, \tag{3.3.8}$$

（最後の積分の値は数値的に出したもので，皆さんに対しては要求しない.）

問 3.3.3. 今度は

$$\int_{-\frac{\sqrt{N}}{50}}^{\frac{\sqrt{N}}{50}} e^{-z^2/2}\frac{dz}{\sqrt{2\pi}} \ge 0.95 \tag{3.3.9}$$

となるような N を求めればよい．この積分は手計算ではできないから，標準正規分布の Φ で書き直し，表を使うしかない．定義から

$$\Phi(x) = \int_{-\infty}^{x} e^{-z^2/2}\frac{dz}{\sqrt{2\pi}} \tag{3.3.10}$$

であった．(3.3.9) の積分を上の Φ で表すには，一般に（講義で説明）

$$\int_a^b e^{-z^2/2}\frac{dz}{\sqrt{2\pi}} = \int_{-\infty}^{b} e^{-z^2/2}\frac{dz}{\sqrt{2\pi}} - \int_{-\infty}^{a} e^{-z^2/2}\frac{dz}{\sqrt{2\pi}} = \Phi(b) - \Phi(a) \tag{3.3.11}$$

とするのが良い．特に $a < 0$ の場合は，対称性から

$$\int_{-\infty} ae^{-z^2/2}\frac{dz}{\sqrt{2\pi}} = \int_{-a}^{\infty} e^{-z^2/2}\frac{dz}{\sqrt{2\pi}} = 1 - \Phi(-a) \tag{3.3.12}$$

を使う．結局,

$$\int_{-\frac{\sqrt{N}}{50}}^{\frac{\sqrt{N}}{50}} e^{-z^2/2}\frac{dz}{\sqrt{2\pi}} = \Phi\Bigl(\frac{\sqrt{N}}{50}\Bigr) - \{1 - \Phi(\frac{\sqrt{N}}{50})\} = 2\Phi(\frac{\sqrt{N}}{50}\Bigr) - 1 \tag{3.3.13}$$

となる．よって，(3.3.9) の条件は

$$2\Phi\Bigl(\frac{\sqrt{N}}{50}\Bigr) - 1 \ge 0.95 \qquad \Longleftrightarrow \qquad 1 - \Phi\Bigl(\frac{\sqrt{N}}{50}\Bigr) \le 0.025 = \frac{1}{40} \tag{3.3.14}$$

となる．正規分布表を見ると，こうなるには

$$\frac{\sqrt{N}}{50} \ge 1.960 \qquad \Longrightarrow N \ge (50 \times 1.960)^2 = 9604 \tag{3.3.15}$$

となる．まあ，余り細かいことを言っても仕方ないので，$N \ge 9600$ ぐらい，と言うのが答え.

4 中心極限定理[22]

以上の準備の下に，統計学の基礎となる中心極限定理を学ぶ．その過程で，中心極限定理がなぜ成り立つのかを理解するための数学的道具も（少し）学ぶ．なお，以下の講義ノートは，実際の講義よりも少し丁寧に書いてあり，節番号も少しずれている.

[22]教科書 5 節 $+\alpha$

4.1 中心極限定理[23]

まずは中心極限定理について述べよう．

以前に大数の法則をやった．これは要約すると，

分散が有界な独立・同分布な確率変数 $X_1, X_2, \ldots$ の和を考え（X_i の期待値を μ），
$S_N := \sum_{i=1}^{N} X_i$ とすると，$\lim_{N\to\infty} P\Bigl[\frac{1}{N} S_N \neq \mu\Bigr] = 0$ が成り立つ

と言うものだった．更にその証明（チェビシェフの不等式を使った）によると，S_N がその平均値の周り $\sqrt{N}$ くらいのところに集中していった．

集中していく様子をもっと細かく見たい，と思うのが人情であり．これに答えてくれるのが中心極限定理である．この定理はこれからの検定・推定の議論の基礎になる，非常に重要なものである．

定理 4.1.1 X_j （$j = 1, 2, \ldots$）を独立，かつ同分布な確率変数とし，その平均と，標準偏差をそれぞれ

$$\mu := E[X_j], \qquad \sigma := \sqrt{\mathrm{Var}[X_j]} \tag{4.1.1}$$

とする．このとき，

$$S_N := \sum_{j=1}^{N} X_j, \qquad Z_N := \frac{1}{\sigma\sqrt{N}} \sum_{j=1}^{N} \bigl(X_j - \mu\bigr) = \frac{S_N - \langle S_N \rangle}{\sigma\sqrt{N}} \tag{4.1.2}$$

を定義すると，任意の $a < b$ に対して

$$\lim_{N\to\infty} P\Bigl[a \le Z_N \le b\Bigr] = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}}\, dx \tag{4.1.3}$$

が成り立つ．

上の定理の主張をもう少し述べておく．S_N や $S_N - N\mu$ 自身は N 個のものの和だから，N が大きくなると（普通は）大きくなる．けれども，$S_N - N\mu$ の大きくなり方は N に比例するのではなく，$\sqrt{N}$ に比例する，と言うのが大数の法則（の証明）だった．そこで上の定理では $S_N - N\mu$ を $\sqrt{N}$ で割ることによって Z_N を定義した．こうすることで，$N \to \infty$ でも（大抵は）有限にとどまるような量を定義したわけである．それで，定理は，この Z_N が $N \to \infty$ で「標準正規分布」に近づいていくことを主張している．

（注意）この中心極限定理の仮定の本質は**確率変数の独立性**である．独立性さえあれば，同分布とは言い切れない確率変数の和に対しても中心極限定理（またはその類似）が成り立つことが多い．(独立性が「弱く」破れている場合にも成り立つこともあるが.)[24]そのため，中心極限定理（またはその類似）は非常に広い範囲の現象に対して普遍的に成り立つ．その結果,「統計的推測」が可能になるのである．

中心極限定理の証明を厳密に行うのはなかなか大変だ．しかし，その過程で，知っておくに値する解析技法がいくつか使われる．この講義ノートのこの節の残りでは，そのような手法をいくつか紹介し，中心極限定理の証明の「感じ」を解説する．

4.2 これ以降で扱う問題意識の説明

中心極限定理などを証明したい場合の状況はだいたい，以下のようなものである：

- それ自体はよくわかっている確率変数の列 $X_1, X_2, X_3, \ldots$ がある．

[23]教科書では該当箇所が少ない

[24]興味のある人は，まともな確率論の教科書を参照のこと．例えば，Kai Lai Chung: *A Course in Probability Theory* (Academic Press) や，A. N. Shiryaev: *Probability* (Springer) などがオススメ

- それらの和（をさらにスケールしたもの）として，新しい確率変数 Z_N を定義し，この Z_N の確率分布（の $N \to \infty$ の極限）を求めたい．
- この場合，Z_N の確率分布を直接計算できれば何も問題ないのだが，大抵の場合，これは計算しにくい（計算できない）．

このような状況なので，以下の問いに意味がでてくる：

> （問い：かなり漠然とした書き方）今ひとつよくわからない確率変数 Z がある．また，そのような確率変数の列 Z_1, Z_2, Z_3 があることもある．特に，その確率分布を計算することはほぼ，不可能だ．その場合でも，これら確率変数（やその極限）をうまく特徴付ける方法を考えよ．

次節では，上の問いに対する一つの答え（モーメントの方法）を紹介する．さらに後の節では，もう一つの方法（モーメント母函数および特性関数の方法）を紹介する．最後に見るように，これらの方法は，中心極限定理（および他の類似の定理）の証明に有用である．

4.3 モーメントの方法[25]

まずは「モーメント」を定義しよう．

定義 4.3.1 (確率変数のモーメント) 確率変数 X に対し，その r 次の**モーメント (moment)** を

$$M_r := E[X^r] \qquad (r = 1, 2, 3, \ldots) \tag{4.3.1}$$

によって定義する．

X が有界でない場合，上のモーメントが定義できない場合もあるかもしれない．そのような場合はモーメントは定義しない．**以下の全ての定理などにおいても，モーメントの存在は仮定しておく．**

モーメントを定義する大きな理由は，**「確率分布を計算できなくてもモーメントは計算できる」**ことがよく起こるからだ．つまり，モーメントをまず計算して，それから確率分布を推測していこう，というノリで進むのだ．となると，以下の問いかけは自然になるだろう．

> 確率変数が与えられたらそのモーメントは一意に決まる（期待値が定義できずに「定義できない」場合も含めて）．その逆はどうだろうか？　つまり，モーメントが同じなら，同じ確率変数と言えるだろうか？

有限個のモーメントだけを問題にしているなら，この答えはノーだ．例えば，M_1, M_2 の二つが同じになるような確率変数はいくらでも異なるものを考えつくことができる．しかし，無限次まで，全てのモーメントが同じならどうだろうか？この答えは以下の命題で，肯定的に与えられる[26]

命題 4.3.2 二つの確率変数 X, Y のモーメントが全ての正の整数次数で一致しているとする：

$$E[X^r] = E[Y^r] \qquad (r = 1, 2, 3, \ldots) \tag{4.3.2}$$

さらに，これらのモーメントは以下の Carleman の条件を満たしているとする：

$$\sum_{r=1}^{\infty} \Bigl(E[X^{2r}] \Bigr)^{1/(2r)} = \infty \tag{4.3.3}$$

このとき，X, Y の分布は等しい．

---

[25]教科書の 5.2 節前半

[26]証明，および関連する定理については，適当な確率論の教科書を参照．例えば，Kai Lai Chung: *A Course in Probability Theory* (Academic Press) の p.98 付近

細かい条件はおいとけば，上の命題は「全ての次数のモーメントを知れば，そのモーメントの元になった確率分布を知ることができる」ことを主張している．なかなか良い．

実は，これだけではちょっと足りない．中心極限定理の証明などに使われる状況では，考えたい確率変数 Z_N そのもののモーメントは完全には計算しきれない．計算できるのは,「Z_N のモーメントの $N \to \infty$ での極限」である．であるので，上の命題をこの**極限版**に拡張したくなる．その一例が以下の命題である[27]．

命題 4.3.3 確率変数の列 $Z_1, Z_2, Z_3, \ldots$ と確率変数 Y があり，これらの確率変数のモーメントはすべて，上の命題の Carleman の条件をみたしているとする．さらに，それぞれの次数 r ごとに，Z_N のモーメントが Y のモーメントに収束するものとする：

$$\lim_{N\to\infty} E[(Z_N)^r] = E[Y^r] \qquad (r = 1, 2, 3, \ldots) \tag{4.3.4}$$

このとき，Z_N の分布は，$N \to \infty$ で，Y の分布に収束する．つまり，$a < b$ に対して

$$\lim_{N\to\infty} P[a \le Z_N \le b] = P[z \le Y \le b] \tag{4.3.5}$$

あとで中心極限定理に関して具体例を見るけども，このモーメントの方法は，確率変数の極限を考えるにはかなり強力な方法である．

4.4 モーメント母函数と特性函数[28]

さて，モーメントの方法がなかなかに有効なことを説明したのだが，これは実は「モーメント母函数」や「特性函数」というものを考えるとさらに見通しがよくなる．この節ではこれらを説明する．

定義 4.4.1 (確率変数のモーメント母函数) 確率変数 X に対し，**モーメント母函数 (moment generating function)** を

$$g(s) := E[e^{sX}] \qquad (s \in \mathbb{C}) \tag{4.4.1}$$

によって定義する．

うえの期待値が定義できない場合はモーメント母函数は定義しない．

これに密接に関連した量として，以下も定義する：

定義 4.4.2 (確率変数の特性函数) 確率変数 X に対し，**特性函数 (characteristic function)** を

$$\phi(s) := E[e^{isX}] \qquad (s \in \mathbb{C}) \tag{4.4.2}$$

によって定義する．ここで i とは虚数単位 $\sqrt{-1}$ である．

定義を見ればすぐにわかるように，$g(is) = \phi(s)$ であり，モーメント母函数と特性函数は本質的に同じものである．ただし，普通は，s を実数に取ることが多く，その場合は，特性関数は定義できるが，モーメント母函数は定義できない，ことも起こりうる（指数函数 e^{sX} の期待値が定義できる確率変数はそれほど多くない）．このような事情から，本質的には同じであるが，これらの二つを区別して考えることが多い．**なお，特性函数は，確率変数 X の確率密度関数のフーリエ変換に他ならないが，これについては，この節の最後に注意する．**

モーメント母函数の存在意義は，その名前の通り，モーメントを母函数の微分として計算できることである：

$$E[X^r] = \frac{d^r}{d\,s^r}\, g(s)\Big|_{r=0} \qquad (r = 1, 2, 3, \ldots) \tag{4.4.3}$$

---

[27]適当な確率論の教科書を参照．前掲の Chung の本なら Theorem 4.5.5.

[28]教科書の 5.2 節後半

同じく，特性函数からも計算できる：

$$E[X^r] = \frac{1}{i^r}\frac{d^r}{d\,s^r}\phi(s)\bigg|_{r=0} \qquad (r = 1, 2, 3, \ldots) \tag{4.4.4}$$

実のところ,「モーメントは計算しにくいけども，モーメント母函数または特性函数なら計算できる」例も多く存在する．そのような場合には，モーメントを計算する際に，モーメント母函数が役に立つ．

さて，モーメントの場合と同じく，モーメント母函数や特性函数を知ると，元の確率変数を知ることができる．その一例が以下の命題である．特性函数に関するものの方がスッキリしているので，それを示す[29]．

命題 4.4.3 二つの確率変数 X, Y の特性函数が一致するならば，つまり，

$$E[e^{isX}] = E[e^{isY}] \qquad (\text{すべての実数 } s \text{ に対して}) \tag{4.4.5}$$

ならば，X, Y の分布は等しい．

上の命題は「特性函数を知れば，その元になった確率分布を知ることができる」ことを主張している．大変によろしい．

上の命題 4.4.3 を極限版に拡張した一例が以下の命題である[30]．

命題 4.4.4 確率変数の列 $Z_1, Z_2, Z_3, \ldots$ があり，Z_N の特性函数を $\phi_N(s)$ と書く．もし，

(a) すべての $s \in \mathbb{R}$ に対して $\phi_\infty(s) := \lim_{N\to\infty} \phi_N(s)$ が存在し，

(b) $\phi_\infty(s)$ は $s = 0$ で連続である

ならば，

$\phi_\infty(s)$ はある確率変数 Y の特性函数になっており，Z_N の分布は，$N \to \infty$ で，Y の分布に収束する．つまり

$$\lim_{N\to\infty} P[a \le Z_N \le b] = P[z \le Y \le b] \tag{4.4.6}$$

あとで中心極限定理に関して具体例を見るけども，特性函数の方法は，確率変数の極限を考えるには非常に強力な方法である．ともかくこれで，4.2 節で掲げた問題への一応の答えを与えた．

フーリエ変換による特性函数の理解

皆さんの中には,「フーリエ変換」というものについて習った人も多いのではないかと思う．そこで，特性函数とフーリエ変換の関係（およびそれに基づいた，上の結果の理解）について述べておく．記述を簡単にするため，数学的な厳密性にはこだわらない．

まず，フーリエ変換というのはなんだったかというと，実軸上で定義された（性質の良い）函数 $h(x)$ に対して，そのフーリエ変換を

$$\hat{h}(k) := \int_{-\infty}^{\infty} \frac{dx}{\sqrt{2\pi}}\, e^{-ikx}\, h(x) \tag{4.4.7}$$

として定義したのだった．すると驚くべきことに，h が「良い性質」を持っているならば，この逆変換が

$$h(x) = \int_{-\infty}^{\infty} \frac{dk}{\sqrt{2\pi}}\, e^{ikx}\, \hat{h}(k) \tag{4.4.8}$$

と書けるのだった[31]．「元々の函数 h は計算しにくいけどもそのフーリエ変換 $\hat{h}$ が計算できる場合」は多く存在する．そのような場合に，まずは $\hat{h}$ を計算して，それから上の逆変換を用いて f を計算する，ことが多く行われ，大変に有効であった．

[29] 適当な確率論の教科書を参照．前掲の Chung の本なら Theorem 6.2.2 が該当する

[30] 適当な確率論の教科書を参照．前掲の Chung の本なら Theorem 6.3.2 が該当する

[31] e^{-ikx} でなく e^{ikx} にする流儀や，分母の $\sqrt{2\pi}$ がない流儀など，いくつかあるが，これらはすべて，逆変換とセットにして consistent に決めておればなんの問題もない

さて，我々が考えている特性函数はこのフーリエ変換と深い関係を持っている．わかりやすいように，確率変数 X が連続型で，その確率密度函数が $f(x)$ だとしよう．すると期待値の定義から

$$\phi(s) = E[e^{isX}] = \int_{-\infty}^{\infty} f(x)\, e^{isx}\, dx \tag{4.4.9}$$

となっており，これはうえの「フーリエ逆変換」の式に（$\sqrt{2\pi}$ を除いて）等しいことがわかる（うえの逆変換の式で，h を ϕ，x を s，k を x，$\hat{h}$ を f，と思う）．つまり，特性函数というのは，確率密度函数のフーリエ（逆）変換そのものなのだ．となると，フーリエ変換の理論から，確率密度函数を特性函数で表せば

$$f(x) = \int_{-\infty}^{\infty} \frac{ds}{2\pi}\, e^{-isx}\, \phi(s) \tag{4.4.10}$$

となるはずであることがわかる．つまり，特性函数を用いてもとの確率密度函数で表すことができるのだから，特性函数から確率分布が定まる．これが命題 4.4.3 に他ならない．

4.5 特性函数による中心極限定理の証明

では，中心極限定理の「証明」を行う．使うのは上の命題 4.4.4 である．

中心極限定理で出てきた確率変数

$$Z_N := \frac{1}{\sigma\sqrt{N}} \sum_{j=1}^{N} \bigl(X_j - \mu\bigr) \tag{4.5.1}$$

の特性函数 ϕ_N を考えよう．指数関数の性質を使うと，期待値の中身が積で書けていることがわかる：

$$\phi_N(s) = E\bigl[e^{isZ_N}\bigr] = E\Bigl[\exp\Bigl\{\frac{is}{\sigma\sqrt{N}} \sum_{j=1}^{N} \bigl(X_j - \mu\bigr)\Bigr\}\Bigr] = E\Bigl[\prod_{j=1}^{N} \exp\Bigl\{\frac{is}{\sigma\sqrt{N}} \bigl(X_j - \mu\bigr)\Bigr\}\Bigr] \tag{4.5.2}$$

ここで**独立性**により，上の「積の期待値」は「期待値の積」に化け，さらに**同分布**なのでこれは X_1 に関する期待値の N 乗になる：

$$\phi_N(s) = \prod_{j=1}^{N} E\Bigl[\exp\Bigl\{\frac{is}{\sigma\sqrt{N}} \bigl(X_j - \mu\bigr)\Bigr\}\Bigr] = \Bigl\{E\Bigl[\exp\Bigl\{\frac{is}{\sigma\sqrt{N}} \bigl(X_1 - \mu\bigr)\Bigr\}\Bigr]\Bigr\}^N \tag{4.5.3}$$

後々のことを考えると，この N 乗は計算しにくいので，log をとった形で以下，考える：

$$\log\, \phi_N(s) = N\, \log\Bigl\{E\Bigl[\exp\Bigl\{\frac{is}{\sigma\sqrt{N}} \bigl(X_1 - \mu\bigr)\Bigr\}\Bigr]\Bigr\} \tag{4.5.4}$$

ともかく，上の log の中にある，X_1 に関する期待値を計算すれば良い．式を簡単にするために $a = s/\sigma\sqrt{N}$ と書くと，問題の期待値は

$$E\Bigl[\exp\Bigl\{\frac{is}{\sigma\sqrt{N}} \bigl(X_1 - \mu\bigr)\Bigr\}\Bigr] = E[e^{ia(X_1-\mu)}] \tag{4.5.5}$$

であるが，

$$E[e^{ia(X_1-\mu)}] = E\Bigl[1 + ia(X_1-\mu) - \frac{a^2}{2}(X_1-\mu)^2 + \frac{(ia)^3}{3!}(X_1-\mu)^2 + \frac{a^4}{4!}(X_1-\mu)^4 + \cdots\Bigr] \tag{4.5.6}$$

と展開してみると，$a = O(1/\sqrt{N})$ であることから，また $E[X_1] = \mu$ であることから

$$E[e^{ia(X_1-\mu)}] = 1 - \frac{s^2}{2\sigma^2 N}\,\sigma^2 + O\bigl(N^{-3/2}\bigr) = 1 - \frac{s^2}{2N} + O\bigl(N^{-3/2}\bigr) \tag{4.5.7}$$

が結論できる[32]．

[32]簡単のため，上の級数展開による計算が正当化でき，かつ X_1 の全てのモーメントが存在するとした．これらの条件抜きでも，右辺の $O(N^{-3/2})$ を $o(N^{-1})$ に変えたものを証明することができるが，この講義のレベルを超えているので略

$\log\phi_N(s)$ の表式にもどれば

$$\log\,\phi_N(s) = N\,\log\Bigl\{1 - \frac{s^2}{2\,N} + O\bigl(N^{-3/2}\bigr)\Bigr\} = -\frac{s^2}{2} + O(N^{-1/2}) \tag{4.5.8}$$

となり，特に，

$$\lim_{N\to\infty}\log\,\phi_N(s) = -\frac{s^2}{2} \qquad \text{または} \qquad \lim_{N\to\infty}\phi_N(s) = \exp\Bigl(-\frac{s^2}{2}\Bigr) \tag{4.5.9}$$

が得られた．ところが，右辺は標準正規分布 $N(0,1)$ の特性函数に他ならない．したがって，命題 4.4.4 によって，確率変数 Z_N の分布は，標準正規分布 $N(0,1)$ の分布に，$N\to\infty$ で近づくのである． □

4.6 歪度と尖度，キュムラント[33]

少し本筋から離れたこととして，モーメントの仲間の特殊なものについて言及しておく．なお，この節では，期待値を $E[X]$ だけではなく $\langle X\rangle$ と書く．(その方が，キュムラントの表式がみやすくなると思うので.)

まず，モーメント母函数は

$$g(s) = E[e^{sX}] \qquad \text{で，その微分は} \qquad \frac{d^r}{d\,s^r}\,g(s)\Big|_{r=0} = E[X^r] = \bigl\langle X^r\bigr\rangle \qquad (r=1,2,3,\ldots) \tag{4.6.1}$$

であった．この対数をとったものを考え，以下のような定義を行う．

定義 4.6.1 (キュムラントとキュムラント母函数) 確率変数 X のモーメント母函数 $g(s)$ はすでに定義したが，その対数をとったもの：

$$W(s) = \log\,g(s) = \log\,E[e^{sX}] \tag{4.6.2}$$

を（この X の）**キュムラント母函数 (cumulant generating function)** という．また，この $W(s)$ の $s=0$ における r 階の微係数

$$W^{(r)}(0) = \frac{d^r}{ds^r}W(s)\Big|_{r=0} \tag{4.6.3}$$

を X の **r-次のキュムラント**という．

すぐ下で見るように，$W^{(r)}(0)$ の簡単な一般的表式は，$r\ge 4$ では存在しない．**（この点，6/11（火曜日）の講義では大嘘を言ってしまいました．申し訳ありません．訂正します.）** そのため，上の定義でも，キュムラントは単に上の微係数で定義することとしている．

<u>キュムラントの具体的計算例</u>

$W(s)$ の表紙を s で微分して $s=0$ とおいて見ると，以下のようになる（$E[X]=\mu$ と書いた)：

$$W'(s) = \frac{g'(s)}{g(s)}, \qquad W'(0) = E[X] = \bigl\langle X\bigr\rangle = \mu, \tag{4.6.4}$$

1 次のキュムラントは X の期待値だった．高階の微分は，分数の微分であるのでどんどん厄介になるけどもともかくやると，2 階は

$$W''(s) = \frac{d}{ds}\frac{g'(s)}{g(s)} = \frac{g''(s)}{g(s)} - \Bigl(\frac{g'(s)}{g(s)}\Bigr)^2, \qquad W''(0) = \bigl\langle X^2\bigr\rangle - \bigl\langle X\bigr\rangle^2 = \bigl\langle (X-\mu)^2\bigr\rangle = \mathrm{Var}[X] \tag{4.6.5}$$

となる．2 次のキュムラントは，X の分散だ．

3 階以降はこれほど簡単でない．ともかく 3 階の微分は

$$W^{(3)}(s) = \frac{g'''(s)}{g(s)} - 3\frac{g''(s)g'(s)}{g(s)^2} + 2\Bigl(\frac{g'(s)}{g(s)}\Bigr)^3 \tag{4.6.6}$$

[33] 教科書の 5.1 節 $+\alpha$

となって，３次のキュムラントは

$$W^{(3)}(0) = \langle X^3\rangle - 3\langle X^2\rangle\langle X\rangle + 2\langle X\rangle^3 = \langle (X-\mu)^3\rangle \tag{4.6.7}$$

となる．最後の等式については，右辺から左辺に計算すると納得できる．あまり分かりやすい量ではないが，ともかく，この量は $Y = X - \mu$ という，期待値がゼロになるようにずらした確率変数の３次のモーメントになっている．

4 階はいよいよ大変で（式がややこしくなるので，引数の (s) を略）

$$W^{(4)}(s) = \frac{g^{(4)}}{g} - 4\frac{g^{(3)}g'}{g^2} - 3\frac{g''g''}{g^2} + 12\frac{g''g'g'}{g^3} - 6\Big(\frac{g'}{g}\Big)^4 \tag{4.6.8}$$

となるので，

$$W^{(4)}(0) = \langle X^4\rangle - 4\langle X^3\rangle\langle X\rangle - 3\langle X^2\rangle^2 + 12\langle X^2\rangle\langle X\rangle^2 - 6\langle X\rangle^4 = \langle (X-\mu)^4\rangle - 3\langle (X-\mu)^2\rangle^2 \tag{4.6.9}$$

を得る．ここまでくると，$Y = X - \mu$ を用いて書いても，それほどよくわからない．

これ以降はもっと大変である．5 階までやっておくと，

$$W^{(5)}(s) = \frac{g^{(5)}}{g} - 5\frac{g^{(4)}g'}{g^2} - 10\frac{g^{(3)}g''}{g^2} + 20\frac{g^{(3)}g'g'}{g^3} + 30\frac{g''g''g'}{g^3} - 60\frac{g''g'g'g'}{g^4} + 24\Big(\frac{g'}{g}\Big)^5 \tag{4.6.10}$$

$$\begin{aligned} W^{(5)}(0) &= \langle X^5\rangle - 5\langle X^4\rangle\langle X\rangle - 10\langle X^3\rangle\langle X^2\rangle + 20\langle X^3\rangle\langle X\rangle^2 + 30\langle X^2\rangle^2\langle X\rangle - 60\langle X^2\rangle\langle X\rangle^3 + 24\langle X\rangle^5 \\ &= \langle (X-\mu)^5\rangle - 10\langle (X-\mu)^3\rangle\langle (X-\mu)^2\rangle \end{aligned} \tag{4.6.11}$$

また 6 階の結果だけを書くと

$$W^{(6)}(0) = \langle (X-\mu)^6\rangle - 15\langle (X-\mu)^4\rangle\langle (X-\mu)^2\rangle - 10\langle (X-\mu)^3\rangle^2 + 30\langle (X-\mu)^2\rangle^3 \tag{4.6.12}$$

となっている．既に述べたように，これ以上の簡単な関係式はない（一応「Bell の多項式」というもので書けるのだが，Bell の多項式そのものがこのような組み合わせ問題をうまく表すために定義されたようなものなので，あまり「簡単に書けた」気がしない）．興味のある人は，例えば，英語版の Wikipedia “cumulant” の 6.6 節などを参照[34]．

キュムラントの意味

ともかく，このようにキュムラントを定義した．上の具体的表式で見たように，キュムラントは $Y = X - \langle X\rangle$ という確率変数で見れば，そこそこ綺麗な形をしている．さらに，これから述べるように，キュムラントは，**考えている確率変数が正規分布からどのくらいずれているか**の指標を与えてくれる．正規分布は非常に普遍的なものであることはすでに強調しているが，その普遍的な正規分布からどのくらいずれているか，を考えるのは様々な応用（統計的な考え方を含む）上も大事なので，キュムラントは大変に重要な量なのだ．

命題 4.6.2 確率変数 X が正規分布に従う時，その３次以上のキュムラントは全てゼロである．
逆に，その３次以上のキュムラントが全てゼロである確率変数は正規分布に従う．

（証明）X が正規分布 $N(\mu, \sigma^2)$ に従う場合には，そのモーメント母函数は定義から

$$g(s) = E[e^{sX}] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\Big[-\frac{1}{2}\big(\frac{x-\mu}{\sigma}\big)^2\Big]\, e^{sx}\, dx \tag{4.6.13}$$

と言うことになるが，これを $y = (x-\mu)/\sigma$ と変数変換して計算すると

$$= e^{s\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-y^2/2}\, e^{is\sigma y}\, dy = e^{is\mu}\, e^{-s^2\sigma^2/2} = \exp\Big(-\frac{\sigma^2 s^2}{2} + \mu s\Big) \tag{4.6.14}$$

[34] Wikipedia を鵜呑みにするのは大変に危険なのでオススメしない（過去には，正しい記述が間違った記述に書き直されることも良くあった）．ただ，自分で調べるための取っ掛かりとしては，役に立つことも非常に多く，特に理系の題材における英語版の記述はかなり充実してきている．Wiki でヒントを得た後，信用できる文献で（さらに良いのは自分で計算して）確かめるのが現時点での正しい使い方だと思う

となる．つまりキュムラント母函数は

$$W(s) = -\frac{\sigma^2 s^2}{2} + \mu s \tag{4.6.15}$$

という２次式であるから，３次以上のキュムラントはゼロ．

（逆について）キュムラントがゼロということは，任意の次数のモーメントが決まるということで，さらにこのモーメントは正規分布のものになっている．よって,「モーメントから分布が決まる」ことを主張する命題 4.3.2 を用いれば良い． □

とくに，３次のキュムラントを**歪度 (skewness)**，４次のキュムラントを**尖度 (kurtosis)** という．歪度は分布が左右対称からどのくらいずれているか，の指標になる．一方，尖度は分布が正規分布からどのくらいすれているかの指標になる．