

1 確率論の基礎

ここでは初めて確率論に触れる人でも戸惑わないように、最低限の定義などから始める。ただし、初めから非常に深いことをやるとそれだけで一学期かかってしまうので、深いことは必要に応じて補っていく方針で進む。

1.1 確率論の舞台 — 事象と標本空間

「確率論」とはその名の通り、「確率」を扱う学問である。世の中には不確かなことが色々ある(例:天気予報)。確率論の究極の目的はこの世の中の色々な現象を解き明かす(手助けになる)ことにあると僕は考えるが、初めから世の中の現象を扱うのはなかなか大変である。そのような場合には、まず、目的の現象を数学的に扱いやすい形に変形し(モデル化)、そのモデルを考えるのが良い。モデルが理解できた後で、このモデルと現実の現象がどう対応しているのか(またはモデル化に失敗したために対応していないのか)などについて考えるのである(ただし、数学としての確率論で扱うのは上で述べたプロセスの前半、数学的なモデルの解析が主である。)

さて、確率論をやるには、まずその舞台を設定する必要がある。例として1個のサイコロを一回振る実験を考えよう。サイコロが端や角で立たないものとする、サイコロの6つの面のどれかが出るであろう。そこで以下の定義を行う。

定義 1.1.1 (標本点と標本空間, 有限バージョン) 一回の実験の結果として起こりうるものを根元事象または標本点と呼ぶ。標本点の全体からなる集合を標本空間 (sample space) Ω と言う。

このサイコロの例では、根元事象は $E_1, E_2, E_3, \dots, E_6$ のどれか(ここで E_j はサイコロの j の目が出ると言うこと)であり、標本空間は $\{E_1, E_2, \dots, E_6\}$ である。

標本空間が有限でない場合はいろいろとややこしいことが起こるので、上の定義は根元事象が有限個しかない(つまり、標本空間が有限集合)の場合のものと理解されたい(無限の場合は後述)。この講義では標本空間が有限の場合(および有限からのアナロジーで理解できる場合)から出発し、段々と深いところに入っていくつもりである。話が分かりにくくなったらいつでも有限の場合のアナロジーに戻って考えるのが良からう。

さて、我々は根元事象のみに興味があるわけではない。そのために根元事象の集まりとして、「事象」を考える。

定義 1.1.2 (事象, 有限バージョン) 標本空間が有限集合の時、数学的には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。

サイコロの例で言えば、事象の例としては「2と3の目がでること」「偶数の目が出ること」「6の目が出ないこと」などがある。

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 $\Omega \setminus E$ (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。

なお、以上のをまとめると、以下の「事象の公理」になる。今までは故意に Ω が有限集合の場合を考えてきたが、 Ω が無限の時には以下のように考える。

定義 1.1.3 (事象の公理 = 可測空間, 無限でもいけるバージョン) Sample Space Ω が与えられたとき、 Ω の事象の集まりとは、以下を満たす Ω の部分集合の集まり(部分集合族) \mathcal{F} のことである。

1. $\mathcal{F} \ni \emptyset$
2. $E \in \mathcal{F}$ ならば $E^c \in \mathcal{F}$
3. $E_1, E_2, E_3, \dots \in \mathcal{F}$ に対し, $\bigcup_{i=1}^{\infty} E_i \in \mathcal{F}$

いくつかの注意を列挙する.

- 上の事象の公理を満たす Sample Space にはちゃんと名前が付いている. 数学ではこいつを 可測空間 と言う. この場合の \mathcal{F} とは Ω の σ -field と呼ばれる.
- このバージョンになると, もはや「 Ω の全ての部分集合を事象と認める」とは 言っていない事 に注意. 事象と認めるのは Ω の σ -field \mathcal{F} の元になっているような, 特別な部分集合だけである. このような特別の部分集合にのみ, 確率を割り振るのである (以下参照).
- 無限になると, なぜこんな変なことをするのかと思うだろうが, それは追々, 具体例を通して考える (今までに確率論をちゃんと勉強してきたこの辺りが良くわかっている人は勿論良いが) 何となくモヤモヤしていても, 今のところは余りにしないで有限の場合を念頭に, 次に進んで欲しい.

1.2 数学における確率

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」を割り振っていきこう.

数学ではある意味で「天下りに」確率を定める. 標本空間が有限集合の場合 から始めよう. 標本空間 $\Omega = \{e_1, e_2, \dots, e_N\}$ を考える (e_j が根元事象).

そもそも, 確率とは何だろうか? いろんな事象の「起こり易さ」を表すもののハズである. その「起こり易さ」は根元事象 e_j の「起こり易さ」を決めれば決まるだろう. だから, 要するに, 根元事象の起こり易さ p_j ($j = 1, 2, \dots, N$) をすべて与えれば確率が決まったと言えるのではないか?

では, この根元事象の確率 p_j はどんな性質を満たすべきだろうか? まず, これは確率だから 0 と 1 の間にはないとけない. 更に, Ω そのものというのは全事象だからこの確率は 1 であるべし. 要するに

$$0 \leq p_j \leq 1, \quad \sum_{j=1}^N p_j = 1 \quad (1.2.1)$$

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

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

となるはずである. というのも, $E = \{e_1\} \cup \{e_2\} \cup \{e_3\} \cup \dots \cup \{e_n\}$ であるので, E とは「 e_1 か, e_2 か, \dots , e_n 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 (ただし, 標本空間が有限の場合). 要するに

- sample space Ω 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 (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) 出てくる. 特に, 標本空間が無限の場合は大抵の根元事象の確率はゼロであり (でなければ確率の和が 1 にならない!), 根元事象から出発することはできない. そのために (根元事象から出発しない) 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, 有限バージョン) 有限な標本空間 Ω が与えられたとき, Ω 上の確率 (または確率測度) とは, 以下を満たす Ω 上の関数 P のこと: すなわち, Ω の部分集合 E のそれぞれについて関数の値 $P[E]$ が定まり, かつ

1. 全ての $E \subset \Omega$ に対して $0 \leq P[E] \leq 1$.
2. $P[\Omega] = 1$
3. $E_1, E_2, E_3, \dots \subset \Omega$ が mutually exclusive, つまり 「 $i \neq j$ ならば $E_i \cap E_j = \emptyset$ 」, のとき,

$$P\left[\bigcup_i E_i\right] = \sum_i P[E_i]$$

が成り立つ. なお, 標本空間 Ω とその上の確率測度 P をあわせて確率空間と言い, (Ω, P) と書く.

要するに, 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, というノリである. 勿論, 実際にどのような P を採用するか (どのように p_j を与えるべきか) は考えている具体的問題による (サイコロの問題でも, イカサマサイコロなら 6 つの面に同じ確率を割り振るのは良くないよね.)

標本空間が無限の場合も考え方は全く同じである. ただ, 確率は「事象」にしか割り振らないので, Ω の全ての部分集合について確率が定義できるわけではない. その事情を書き下すと以下ようになる:

定義 1.2.2 (確率の公理, 一般バージョン) 事象の公理を満たす標本空間 Ω と σ -field \mathcal{F} が与えられたとき, すなわち可測空間 (Ω, \mathcal{F}) が与えられた時, (Ω, \mathcal{F}) 上の確率 (測度) とは, 以下を満たす \mathcal{F} 上の関数 P のこと. すなわち, \mathcal{F} の元 E のそれぞれについて関数の値 $P[E]$ が定まり, かつ

1. 全ての $E \in \mathcal{F}$ に対して $0 \leq P[E] \leq 1$.
2. $P[\Omega] = 1$
3. $E_1, E_2, E_3, \dots \in \mathcal{F}$ が mutually exclusive, つまり 「 $i \neq j$ ならば $E_i \cap E_j = \emptyset$ 」, のとき,

$$P\left[\bigcup_{i=1}^{\infty} E_i\right] = \sum_{i=1}^{\infty} P[E_i]$$

が成り立つ. なお, 標本空間 Ω と σ -field \mathcal{F} , その上の確率測度 P をあわせて確率空間と言い, (Ω, \mathcal{F}, P) と書く.

この定義は, 有限の場合とほとんど変わらない. 唯一の違いは確率 $P[E]$ が計算できるもの (つまり事象 E) が Ω の部分集合全てではない可能性があることで, そのために「有限バージョン」では「全ての部分集合 E に対して」となっていたところを「 \mathcal{F} の元である E に対して」と書き直してあるところである.

なお, 有限の場合の σ -field \mathcal{F} は Ω の部分集合全体にとるのが自然であり, 実際, 定義 1.2.1 でもそうした. だから, この場合は \mathcal{F} が自明なので \mathcal{F} を省略して (Ω, P) と書いた. しかし, Ω が無限の場合は \mathcal{F} として色々な可能性がある. そのため, どのような \mathcal{F} を考えているのかを明記する必要があるので, 確率空間として (Ω, \mathcal{F}, P) と書くのである. 以下では Ω が有限の場合でも形式的に (Ω, \mathcal{F}, P) と書くことが多いが, その場合でも (おそらくいつでも) \mathcal{F} は Ω の部分集合全体と解釈する.

この確率の性質については以下が成り立つ.

命題 1.2.3 確率空間 (Ω, \mathcal{F}, P) が与えられたとき, $E, F \in \mathcal{F}$ に対して:

$$P[E^c] = 1 - P[E] \tag{1.2.3}$$

$$E \subset F \implies P[E] \leq P[F] \tag{1.2.4}$$

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

1.3 事象の独立性と条件付き確率

この節の内容は一応「事象」について書いてあるが、実際には「確率変数」に対して同様の概念を使うことが多い。しかし、少なくとも標本空間が有限の場合にはまず「事象」について「独立性」「条件付き」を考える方が直感的であると思うので、ここに載せることにした。

定義 1.3.1 (独立な事象) 確率空間 (Ω, \mathcal{F}, P) 中の事象 $E, F \in \mathcal{F}$ が、

$$P[E \cap F] = P[E] P[F] \quad (1.3.1)$$

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) と言う場合に当たる。

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 1.3.2 (条件付き確率) 確率空間 (Ω, \mathcal{F}, P) 中の事象 $E, F \in \mathcal{F}$ を考える。 $P[F] \neq 0$ の場合に、

$$P[E|F] \equiv \frac{P[E \cap F]}{P[F]} \quad (1.3.2)$$

を F の下で E が起こる条件付き確率 と言う。

註 1.3.3 E と F が独立の場合はもちろん、 $P[E|F] = P[E]$ となる。

場合によっては、 $P[E]$ そのものよりも $P[E|F]$ と $P[F]$ の方が良くわかる場合があり、この場合

$$P[E] = P[E|F] P[F] + P[E|F^c] P[F^c] \quad (1.3.3)$$

として $P[E]$ を計算することもある。条件付き確率そのものに興味がある場合もあるが、このように、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(詳しくは講義で説明していく)。

上では2つの事象の「独立性」を定義したが、勿論、3つ以上の事象についても拡張できる。ただし、注意が必要。

定義 1.3.4 (独立な事象 II, 4月23日の講義後、訂正) 確率空間 (Ω, \mathcal{F}, P) 中の事象 $E_1, E_2, \dots, E_n \in \mathcal{F}$ が任意の k ($2 \leq k \leq n$) と、 i_1, i_2, \dots, i_k ($1 \leq i_j \leq n$, ただし i_j 達は互いに等しくない) に対して

$$P[E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_k}] = \prod_{j=1}^k P[E_{i_j}] \quad (1.3.4)$$

を満たすとき、 E_1, E_2, \dots, E_n は独立な事象 であると言う。

註 1.3.5 E_1, E_2, \dots, E_n がペア毎に独立、つまり

$$P[E_i \cap E_j] = P[E_i] P[E_j] \quad (1 \leq i < j \leq n) \quad (1.3.5)$$

であったとしても、定義 1.3.4 の意味で独立とは限らない(例を作ってみよう)。

4月23日の連絡事項：重要！評価方法の変更の可能性について。
先週、大体の評価方法を予告した。しかし、「知識量調査アンケート」の結果、中間試験・期末試験に予定した期間が教育実習期間と重なることが判明した。公平性を期す意味でも、予定していた期間に中間・期末試験を行うことは良くないだろう。となると、かなり大幅に評価方法を変える必要が生じると思われる。具体策については現在も検討中であり、もっと具体化し次第みなさんに連絡するが、ともかく今は先週に説明した評価方法を大幅に（下手すると跡形もないくらいに）変更する方向でご了解願いたい。

確率論 I, 確率論概論 I 第 1 回レポート問題

問 1： 確率空間 (Ω, \mathcal{F}, P) があって、事象 A, B について以下のことがわかっている：

$$P[A] = \frac{1}{3}, \quad P[A \cup B] = \frac{2}{3}, \quad P[B] = \frac{1}{2}. \quad (1.3.6)$$

このとき、次の 2 つの確率を求めよ。

$$P[A \cap B], \quad P[A \setminus B]. \quad (1.3.7)$$

(注) 集合算において、 $A \setminus B \equiv \{x \in A : x \notin B\}$ 。

問 2： 正 12 面体で出来たサイコロを転がす実験を考える（12 の面のどれも同じ確率で出ると思って良い）。12 の面に 1~12 の数字で互いに異なる番号を振り、これを転がす。転がした結果出た面（一番上になっている面）の数字を Z としよう。次の問に答えよ。

1. 確率変数 Z のとりうる値と、その値をとる確率を求めよ。また、 Z の分布関数を求めよ。
2. 新しい確率変数 X と Y を Z を通して以下のように定める：
 - Z を 3 で割ったときの余りを X とする。
 - Z を 4 で割ったときの余りを Y とする。
 このとき、 X, Y それぞれのとりうる値と、その値をとる確率を求めよ。
3. X と Y は独立な確率変数か？
4. さらに新しい確率変数 $W = X + Y$ を作る。 W のとりうる値と、その値をとる確率を求めよ。
5. W の期待値と分散を求めよ。

締め切りなど：

締め切りは 2002 年 4 月 26 日（金）の 17:00，

提出場所は僕の部屋（理 1-508）の前の封筒かポスト

用紙はできうる限り A4 の紙を用いる（B5 などの小さい紙は紛れてなくなるかも）

とします。

レポートのお約束：

- 友達と相談しても、本を調べても、何をやっても良いから、自分で理解した範囲を書くこと。その際、参考文献や議論した友達の名前も明記すること（友達と議論したり、本を見たからと言って悪い点をつける、などと言うことは絶対にしない。一番大事なのは自分でわかったところを表現することだから、それまでの過程で何をやっても問題ない。）
- なお、問題の番外編として、今までの講義内容・講義形態についての感想、不満、文句、このように改善すべしとの意見などもできるだけ書いてください。お願いします。

1.4 確率変数と期待値

今まで、確率空間とその上の事象のみを相手にしてきた。しかし、ランダムな確率変数に応じてランダムに値の変わる関数を考えると、物事がよく見えることが多い。例えば、10000 個のサイコロを同時に投げるときには、それぞれのサイコロがどのような目を出したかには余り興味がなく、むしろ「1 の目を出したサイコロは何個か」「出た目の合計はいくらか」などに興味のあることが普通であろう。この節では、そのようなランダムな変数について考える。

1.4.1 確率変数とは

確率空間 (Ω, \mathcal{F}, P) (可測空間 (Ω, \mathcal{F}) とその上の確率測度 P) が与えられたとする。 (Ω, \mathcal{F}, P) 上の確率変数とは、大ざっぱには「その値が確率的に (ランダムに) 変動する数」のこと。土台になる確率空間を考えた上での確率変数だから、それぞれの値をとる確率は (原理的に) 計算できる。例えば、

例 1.4.1: サイコロを一回投げる場合、出た目の数を X とすると、 X は 1, 2, 3, 4, 5, 6 のどれかをとる確率変数。 $P[X = i] = 1/6$ というのが自然 ($i = 1, 2, 3, \dots, 6$)。

例 1.4.2: サイコロを 2 つ投げるとき、出た目の合計を Z とすると、 Z は 2 から 12 の値をとる確率変数。 $P[Z = 2] = \frac{1}{36}$, $P[Z = 3] = \frac{1}{18}$, $P[Z = 4] = \frac{1}{12}$ など。

例 1.4.3: 宝くじを一枚買ったとして、それが当たった賞金の額も確率変数 (ハズレは 0 円として)。

概念としては簡単なんだけど、これは実用上、なかなか有用である。そもそも確率変数は、以下の「期待値」や「分散」などを通して、対象とする確率モデルをよりよく理解する (特徴づける) ために使われることが多い。

まあ、こういうものなんだが (標本空間が有限の場合はそれでよいのだが) 一般の場合の厳密な定義を一応、書いておこう。一般には確率変数も実数値をとるとは限らない (もっとややこしい空間内に値をとることもある: 例としてはブラウン運動)。しかし、そんなややこしいことは後にして、普通「確率変数」と言うのは「実確率変数」のことである。これを定義するためにまず、可測関数 の概念を導入する。

定義 1.4.1 (可測関数) 可測空間 (Ω, \mathcal{F}) がある。 Ω 上の実数値関数 X が \mathcal{F} -可測 とは、

$$\text{全ての } A \in \mathcal{B}_1 \text{ に対して } X^{-1}(A) \equiv \{\omega \in \Omega \mid X(\omega) \in A\} \in \mathcal{F} \quad (1.4.1)$$

が成り立つことである。ここで \mathcal{B}_1 とは 1 次元ボレル集合の全体 (= \mathbb{R} の開集合全体を含む最小の σ -field)¹。

この言葉を使うと、実確率変数は以下のように定義される。

¹この講義では \mathcal{B}_1 をいつも 1 次元ボレル集合の全体の意味で用いる。また、 \mathcal{B}_d を d 次元ボレル集合の全体の意味に用いる

定義 1.4.2 (実確率変数) 確率空間 (Ω, \mathcal{F}, P) 上の \mathcal{F} -可測関数を (Ω, \mathcal{F}, P) 上の実確率変数と言う。しつこく書き下すと, 実確率変数とは (Ω, \mathcal{F}, P) 上の実数値関数 $X: \Omega \rightarrow \mathbb{R}$ で,

$$\text{全ての } A \in \mathcal{B}_1 \text{ に対して } X^{-1}(A) \equiv \{\omega \in \Omega \mid X(\omega) \in A\} \in \mathcal{F} \quad (1.4.2)$$

を満たすものことである。

(大ざっぱに言うと) 確率変数とは Ω から \mathbb{R} への関数 X で、「 X が特定の値をとる確率がもとの Ω に戻ったら計算できるもの」なのである。上の定義で $X^{-1}(A)$ と言うところが「 X が A 内の値をとるような, 元々の ω 達」を計算しているわけ。

さて, 上の定義から $X^{-1}(A) = E$ と書くと, E は \mathcal{F} の元であるから, 事象 E の実現確率 $P[E]$ が定義できている。そこでこれを確率変数 X の言葉で $X \in A$ となる事象の確率, と解釈し, $P[X \in A]$ と書く。

(おまけ) 一般の場合の確率変数の定義を載せておく。一般というのは確率変数が一般の可測空間 (S, \mathcal{S}) に値をとる場合である。

定義 1.4.3 (確率変数, 一般バージョン) 確率空間 (Ω, \mathcal{F}, P) と可測空間 (S, \mathcal{S}) が与えられたとき (ランダムな) 関数 $X: \Omega \rightarrow S$ が確率変数とは, X が

$$\text{全ての } A \in \mathcal{S} \text{ に対して } X^{-1}(A) \equiv \{\omega \in \Omega \mid X(\omega) \in A\} \in \mathcal{F} \quad (1.4.3)$$

が成り立つことである。つまり, X が S に値をとる \mathcal{F} -可測関数である, ということ (下の定義 1.4.1 参照)。

定義 1.4.3 で S を実数全体の集合, \mathcal{S} を \mathcal{B}_1 としたものが実確率変数 (定義 1.4.2) に相当する。以下では特に断らない限り, 確率変数と言うときは「実数値確率変数」を指すこととする。

1.4.2 分布と分布関数

確率空間 (Ω, \mathcal{F}, P) 上の実確率変数 X が与えられたとき, X そのものがどのように分布しているのかを問題にしたい。 X の分布の仕方を表すものとして「分布」と「分布関数」を導入する。

定義 1.4.4 (X の分布) まず, \mathcal{B}_1 上の確率測度 μ を $\mu \equiv P \circ X^{-1}$ として定義する。この意味は $A \in \mathcal{B}_1$ に対してその μ の値を

$$\mu(A) \equiv P(X \in A) \quad (1.4.4)$$

と定義することを意味する。この μ を確率変数の分布または法則と言う。

しつこいが, この μ は確率変数 X のばらつきを表す確率測度になっている。

さて, 分布は定義そのものから測度であるが, これをもっとわかりやすくしたい。そのために:

定義 1.4.5 (X の分布関数) 函数

$$F(a) \equiv P[X \leq a] \equiv P[X \in (-\infty, a]] = \mu(-\infty, a] \quad (1.4.5)$$

を, 確率変数 X の (累積) 分布関数 と言う.

読んで字のごとく, 分布関数 $F(a)$ は, 確率変数 X が a 以下の値をとる確率を表す。「累積」と言うのは, X が $-\infty$ から a までの全ての値をとる, と言う (累積している) 感じ.

また, n 個の実数値確率変数 X_1, X_2, \dots, X_n に対しての拡張は自然に行う. すなわち, 函数

$$F(a_1, a_2, \dots, a_n) \equiv P[X_1 \leq a_1, X_2 \leq a_2, \dots, X_n \leq a_n] \quad (1.4.6)$$

を X_1, X_2, \dots, X_n の分布関数という.

1.4.3 確率変数の独立性

確率変数についても, 事象の時と同じく「独立」などの概念を導入することができる (大ざっぱに言うと) 実数値確率変数 X と Y は任意の $A, B \subset \mathcal{B}_1$ に対して

$$P[X \in A \text{ かつ } Y \in B] = P[X \in A] P[Y \in B] \quad (1.4.7)$$

を満たすとき, X と Y は独立な確率変数 と言う. 要するに「 X が集合 A に入る事象」と「 Y が集合 B に入る事象」が独立事象である場合を言うわけだ.

一般化した形で述べると, 以下のようになる.

定義 1.4.6 確率空間 (Ω, \mathcal{F}, P) 上にいくつかの確率変数 X_i ($i = 1, 2, \dots, n$) が与えられている場合を考える. それぞれの X_i は可測空間 (S_i, \mathcal{S}_i) 内に値をとるものとする. このとき, 任意の $A_i \in \mathcal{S}_i$ ($i = 1, 2, \dots, n$) に対して

$$P[X_1 \in A_1, X_2 \in A_2, \dots, X_n \in A_n] = \prod_{i=1}^n P[X_i \in A_i] \quad (1.4.8)$$

となるとき, X_1, X_2, \dots, X_n は独立 と言う.

これを分布関数の言葉で言うと以下のようになる.

定理 1.4.7 X_i が全て実数値確率変数の時, X_1, X_2, \dots, X_n が独立であることは, 分布関数で言う

$$F(a_1, a_2, \dots, a_n) = \prod_{j=1}^n F_j(a_j) \quad (\text{全ての实数 } a_j \text{ に対して}) \quad (1.4.9)$$

が成り立つこと. ただし上の右辺で F_j と言うのは X_j の分布関数である.

1.5 期待値と分散

確率変数に対して, 期待値の概念を導入する. まず, 標本空間が有限の場合から:

定義 1.5.1 (期待値, 有限バージョンその 1) 有限個の根元事象からなる確率空間 (Ω, P) があり, $\Omega = \{e_1, e_2, \dots, e_n\}$, またその確率は

$$P[e_j] = q_j, \quad \left(\sum_{j=1}^n q_j = 1 \right) \quad (1.5.1)$$

であるとする. (Ω, P) 上の実確率変数 X に対して, X の期待値 を

$$E[X] \equiv \langle X \rangle \equiv \sum_{j=1}^n q_j X(e_j) \quad (1.5.2)$$

により定義する.

記号について: 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. この講義では両方の記号を自由に使うことにする.

さて, 上の定義を X の「分布」を使って書き直すと以下ようになる.

定義 1.5.2 (期待値, 有限バージョンその 2) 確率変数 X が x_1, x_2, \dots, x_m の値をとり, その確率が

$$P[X = x_i] = p_i \quad \left(\sum_{i=1}^m p_i = 1 \right) \quad (1.5.3)$$

と与えられているとする. このとき, X の期待値 を

$$E[X] \equiv \langle X \rangle \equiv \sum_{i=1}^m p_i x_i \quad (1.5.4)$$

により定義する.

さて, 標本空間が有限とは限らない場合には, 以下のように定義する.

定義 1.5.3 (期待値, 一般バージョン) 確率空間 (Ω, \mathcal{F}, P) 上の実数値確率変数 X の期待値 を

$$E[X] \equiv \langle X \rangle \equiv \int_{\Omega} X(\omega) P(d\omega) \quad (1.5.5)$$

により定義する.

これもいくつかの例を考えてみて欲しい.

期待値を X の分布 μ を用いて表すことができ,

$$E[X] = \int_{-\infty}^{\infty} x d\mu(x), \quad E[f(X)] = \int_{-\infty}^{\infty} f(x) d\mu(x) \quad (1.5.6)$$

が成り立つ. ここで f は $E[f(X)]$ が存在するような実数値関数.

次に「分散」を定義する. これは以下のような確率変数の期待値として定義される.

定義 1.5.4 (分散) 確率変数 X の分散を

$$\text{Var}[X] \equiv E\left[(X - E[X])^2\right] = E[X^2] - E[X]^2 = \langle X^2 \rangle - \langle X \rangle^2 = \langle (X - \langle X \rangle)^2 \rangle \quad (1.5.7)$$

により定義する.

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

$$I[F] \equiv \begin{cases} 1 & (F \text{ が起こるとき}) \\ 0 & (F \text{ が起こらないとき}) \end{cases} \quad (1.5.8)$$

として定義すると,

$$P[F] = E[I[F]] = \langle I[F] \rangle \quad (1.5.9)$$

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

期待値の重要な性質はその線形性である. すなわち, 確率空間 (S, P) における確率変数 X, Y に対して,

$$E[X + Y] = E[X] + E[Y] \quad (1.5.10)$$

が成り立つ. 他にも類似の式が成り立つので, 以下にまとめておこう.

命題 1.5.5 確率空間 (Ω, P) における確率変数 X, Y と実定数 $a > 0$ に対しては以下が成り立つ:

$$E[X + Y] = E[X] + E[Y], \quad E[aX] = aE[X] \quad (1.5.11)$$

$$\text{Var}[aX] = a^2 \text{Var}[X] \quad (1.5.12)$$

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}(X, Y), \quad \text{Cov}(X, Y) \equiv \langle (X - \langle X \rangle)(Y - \langle Y \rangle) \rangle. \quad (1.5.13)$$

$\text{Cov}(X, Y)$ は X と Y の共分散と言う.

註: これらの結果は X, Y の分布が独立でなくても成り立つ.

証明:

以下ではある意図を持って, 有限の場合の証明をまず載せた(馬鹿にするな! と叱られそうだが). 一般の場合の証明はその後に入れた. 両者を比べて思うところがあると嬉しいのだが...

標本空間が有限の場合

X のとりうる値を x_i ($i = 1, 2, \dots, N$), Y のとりうる値を y_j ($j = 1, 2, \dots, M$), それぞれの値をとる確率を $P[X = x_i \text{ かつ } Y = y_j] = p_{ij}$ とおく. すると,

$$E[X + Y] = \sum_{ij} p_{ij}(x_i + y_j) = \sum_{ij} p_{ij}x_i + \sum_{ij} p_{ij}y_j \quad (1.5.14)$$

であるが, $\sum_{j=1}^M p_{ij} = P[X = x_i \text{ かつ } Y \text{ は何でも良い}] = P[X = x_i]$ であるので,

$$\sum_{ij} p_{ij}x_i = \sum_{i=1}^N x_i \left(\sum_{j=1}^M p_{ij} \right) = \sum_{i=1}^N x_i P[X = x_i] = E[X] \quad (1.5.15)$$

が成り立つ．同様に

$$\sum_{ij} p_{ij} y_j = E[Y] \quad (1.5.16)$$

なので, $E[X + Y] = E[X] + E[Y]$ が証明された．

次に, $E[aX]$ については,

$$E[aX] = \sum_{i=1}^N P[X = x_i](ax_i) = a \sum_{i=1}^N P[X = x_i] x_i = a E[X]. \quad (1.5.17)$$

また, $\text{Var}[aX]$ については

$$E[(aX)^2] = E[a^2 X^2] = a^2 E[X^2] \quad (1.5.18)$$

であることと線形性から

$$\text{Var}[aX] = E[(aX)^2] - (E[aX])^2 = a^2 E[X^2] - (aE[X])^2 = a^2 E[X^2] - a^2 (E[X])^2 = a^2 \text{Var}[X] \quad (1.5.19)$$

標本空間が一般の場合の形式的証明

使うのは積分の線形性だけである．まず期待値の定義から

$$E[X+Y] = \int_{\Omega} [X(\omega)+Y(\omega)] P(d\omega) = \int_{\Omega} X(\omega) P(d\omega) + \int_{\Omega} Y(\omega) P(d\omega) = E[X]+E[Y] \quad (1.5.20)$$

次に, $E[aX]$ については,

$$E[aX] = \int_{\Omega} aX(\omega) P(d\omega) = a \int_{\Omega} X(\omega) P(d\omega) = a E[X]. \quad (1.5.21)$$

また, $\text{Var}[aX]$ についての証明は有限の時と同じ． □

X と Y が独立な場合には,

$$E[XY] = E[X] E[Y], \quad \text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \quad (1.5.22)$$

が成り立つ．

問 1.5.6 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.)

1.6 数の数え方の復習

この節の内容は流石に 4 年生・大学院生なら知っているだろうと思われることであるが, 復習を兼ねて載せておく．講義では触れない(白状すると, ここは一年向けの「数学展望」の講義ノートからそのまま抜いてきた)．もし怪しい人は, 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．

まず記号を導入しておく．

定義 1.6.1

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

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, \dots, 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

$$n \cdot n \cdot n \cdots n = n^k \quad (1.6.1)$$

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, \dots, 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

$$n \cdot (n-1) \cdot (n-2) \cdots (n-k+1) = \frac{n!}{(n-k)!} \quad (1.6.2)$$

となる.

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, \dots, 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:

$$\frac{n!}{(n-k)!} \times \frac{1}{k!} = \binom{n}{k} \quad (1.6.3)$$

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, \dots, 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

$$\binom{n}{n_1} \times \binom{n-n_1}{n_2} \times \binom{n-n_1-n_2}{n_3} \times \cdots \times 1 = \frac{n!}{n_1! n_2! n_3! \cdots n_r!} = \binom{n}{n_1 n_2 n_3 \cdots n_r} \quad (1.6.4)$$

となることがわかる.

4月30日の連絡：恥ずかしながら，一日目に配ったプリントにミスプリを見つけました（汗）．以下に正しいものを載せますので，みてください（僕の web page に載せてある4月30日分の PFD file は訂正済み）．

なお，採点方法はまだ考えています．下手に書くとまた混乱を招きそうなので，今日は口頭で大体の方向を言うに留めます．

（プリント4ページ，定義 1.3.4 を以下のように訂正します）

定義 1.3.4. 独立な事象 II

確率空間 (Ω, \mathcal{F}, P) 中の事象 $E_1, E_2, \dots, E_n \in \mathcal{F}$ が任意の k ($2 \leq k \leq n$) と i_1, i_2, \dots, i_k ($1 \leq i_j \leq n$, ただし i_j 達は互いに等しくない) に対して

$$P[E_{i_1} \cap E_{i_2} \cap \dots \cap E_{i_k}] = \prod_{j=1}^k P[E_{i_j}] \quad (1.6.5)$$

を満たすとき， E_1, E_2, \dots, E_n は独立な事象 であると言う．

（要するに， n 個の積事象の確率のみを見ているのではダメで， n 個より少ない数の積事象も全部見る必要がある，と言うこと．）

確率変数に関する同様の定義はそのままで良い．この理由は，確率変数に関しては n 個の A_i の内のいくつかを X_i （全空間）ととることで，実質的に n 個より少ない確率変数の相関を見ることが出来るからである．

転んでもタダでは起きない問題：

間違っていた「定義 1.3.4」によれば独立と思われるが，上で訂正した定義では独立にならない例を作れ（正しい定義と間違っていた定義の違いを理解する問題）．

2 大数の法則と中心極限定理

2.1 問題設定

この節では，以下のような質問に答えたい．

問 2.1.1 表と裏が確率 $\frac{1}{2}$ ずつで出るようなコインを何回も投げる． N 回投げたとき，表の出た回数は N 回の内の何回くらいだろうか？（勿論，一回ごとのコイン投げの結果は互いに独立だと仮定する．）

上の質問のもう少し複雑なものとして，

問 2.1.2 さいころを何回も投げることを考えよう．一回投げる毎に以下の要領で点数をもらえらるものとする．

- 出た目が 1 または 2 の時は +2 点
- 出た目が 3 から 6 の時は -1 点

毎回出た点数を加算していくとして, さいころを N 回投げたときの得点 S_N はどのように分布しているだろうか?(ここでもさいころの 6 つの面が出る確率は全て $\frac{1}{6}$ であり, かつ一回ごとのさいころ投げは独立と仮定する.)

これらの質問は, もっと一般に以下の様に定式化される.

問 2.1.3 独立な確率変数 X_1, X_2, X_3, \dots に対して, 新しい確率変数

$$S_N \equiv \sum_{i=1}^N X_i \quad (2.1.1)$$

を定義する (N は正の整数). S_N の分布はどうなっているか?

これらの問では「分布はどうなっているか」と, はなはだ主観的な問いかけがなされているが, これは「どのように物事を見たら分布の特徴が一番捉えられるか考える」ことまで含めて問題にしたいためである.

実は, これらの問いに対する答を一般的に与えることができる. しかも, その答は我々の直感のある程度, 支持するものであるので, この節での結果は非常に重要である. これらの結果は, 近代確率論の一つの頂点とも言える. 具体的には以下の 3 つを取り扱いたい (簡単のために, X_i の分布はみな同じで, その期待値はそれぞれ μ , 分散は σ^2 の場合を述べる).

- 大数の弱法則 (convergence in probability): かなり大ざっぱだが簡単に導出できる.

$$\text{任意の } \epsilon > 0 \text{ に対して} \quad \lim_{N \rightarrow \infty} P \left[\left| \frac{S_N}{N} - \mu \right| > \epsilon \right] = 0 \quad (2.1.2)$$

- 大数の強法則 (convergence “with probability 1”): 上よりも精密だが, 導出がちと厄介. よく考えると, 定理の意味を理解するのもちと厄介.

$$P \left[\lim_{n \rightarrow \infty} \frac{S_N}{N} = \mu \right] = 1 \quad (2.1.3)$$

- 中心極限定理 (convergence in distribution): 上の 2 つの更なる精密化.

$$\lim_{N \rightarrow \infty} P \left[a \leq \frac{S_N - N\mu}{\sigma\sqrt{N}} \leq b \right] = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad (2.1.4)$$

この段階では大体の感じを掴めば十分で, 個々の定理についてはひとつずつきっちりやる. ここでは似たような (でも異なる) 定理がいろいろある, ことをわかってもらいたくて書き連ねておいた.

2.1.1 直積空間の構成 (少し advanced)

直感的には問題設定は良いと思うが、数学的に少しだけ詰めておく。

今は確率変数 X_i ($i = 1, 2, 3, \dots$) があつたとき, $S_N \equiv \sum_{i=1}^N X_i$ と言うような和を考え, この確率変数の分布を知りたい. でもこれをやるには「 X_i が一杯あるときの X_i 達の確率分布を定義する」ことから始めないといけない. この定義は数学的には「直積測度」「直積確率空間」と言うものを使っていることになる.

定義 2.1.4 (2つの確率空間の直積) $(\Omega_j, \mathcal{F}_j, P_j)$ を確率空間とする ($j = 1, 2$). これらの直積確率空間 (Ω, \mathcal{F}, P) を以下のようにして定義する.

- まず Ω は, Ω_1 と Ω_2 の直積集合として定義する:

$$\Omega \equiv \Omega_1 \times \Omega_2 \equiv \{(\omega_1, \omega_2) \mid \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\} \quad (2.1.5)$$

- \mathcal{F} は以下のように段階的に定義する.

1. $\mathcal{C} \equiv \{A_1 \times A_2 \mid A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$

2. \mathcal{A} は \mathcal{C} の, 互いに素な元の有限和.

3. \mathcal{F} は \mathcal{A} を含む最小の σ -field, つまり $\mathcal{F} = \sigma(\mathcal{A})$

- 最後に P は,

1. $A_1 \times A_2$ の形の集合に対しては ($A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2$), $P[A_1 \times A_2] \equiv P[A_1]P[A_2]$ と定義し,

2. より一般の \mathcal{F} の元 (つまり $A_1 \times A_2$ の形のものの和集合や, \mathcal{F} が σ -field であることによる「極限」的な集合など) に対しては P に σ -加法性を課すことで拡張していつて定義する. つまり, $A \cap B = \emptyset$ の時に $P[A \cup B] = P[A] + P[B]$ などとしていく.

註 2.1.5 上の最後の部分 (P を直積の形の集合での値から一般に拡張するところ) が実際に行えるのか (拡張した測度が定義できるのか) は, 勿論, 証明を要する. しかし, この辺りは「測度の拡張定理」と呼ばれるものがいくつかあって, ちゃんと出来ることが保証されている (特に Carathéodory の拡張定理が効いてくる). この辺りをまじめにやると, それだけで測度論の講義が半学期ほど必要になるので, この講義では結果だけの引用に留める. 良くわかっている人は自分で勉強して欲しいが, 測度論が苦手な人は次善の策として, 大体このような感じ, と理解しておけばよい.

註 2.1.6 上では2つの確率空間の直積を定義したが, n 個の確率空間の直積も同様に定義する. なお, 後の方では「無限個の」確率空間の直積も必要になるが (大数の強法則に絡んで), それはその時に説明する.

2.1.2 具体的計算の奨め

以下のような問題を自分でやってみることは, この後の節の理解に役立つであろう. (レポートにしようかと思ったが, コンピューターを使わないと計算が大変なので, 各自が自主的にやるに任せる.) 問題設定は以下の通り (講義でも説明する).

- いびつなコイン (コインの表が出る確率は $\frac{2}{3}$, 裏が出る確率は $\frac{1}{3}$) を何回も投げる. 投げる毎の結果は独立であると仮定する.
- コインの表が出たら +1 点, 裏が出たら -1 点をもらえるものとし, j -回目のコイン投でもらった点を X_j と書く:

$$P[X_j = 1] = \frac{2}{3}, \quad P[X_j = -1] = \frac{1}{3}.$$

- N 回, コインを投げたときの得点の合計は勿論, $S_N \equiv \sum_{j=1}^N X_j$ であって, 以下, これに着目する.

このような設定の下で, 以下のような計算を行おう.

1. 確率変数 X_1 の期待値と分散を計算せよ. 以下, X_1 の期待値を μ , 分散の平方根を σ と書くことにする².
2. S_N の期待値と分散も計算せよ.
3. k を任意の整数として, 確率 $P[S_N = k]$ を表す式を作れ (いくつかの k に対してはこの確率はゼロかも).
4. 以下では $\frac{S_N}{N} - \mu$ がゼロに行くかどうか ($N \rightarrow \infty$ で) などを問題にする. そこで, 確率

$$P\left[\left|\frac{S_N}{N} - \mu\right| > \epsilon\right] \quad (2.1.6)$$

の具体的な値を, いくつか計算してみよう. 例えば $\epsilon = 0.2$ に対して, 上の確率を $N = 10, 20, 50$ くらいで計算したらどうなっているだろうか? (注意: 勿論, N をもっと大きくして, いろいろな ϵ を試してみるのが良いが, 大きな N の場合をまともに計算するとコンピュータでも大変な時間がかかることになるので, 注意されたし.)

来週以降, 地道ではない計算法を問う問題をレポートとして出す予定.

2.2 大数の弱法則

大数の弱法則は非常に簡単に導出できるにもかかわらず, その述べるところは強力である. まず, 定理から述べる.

定理 2.2.1 (大数の弱法則, Weak Law of Large Numbers) 質問 2.1.3 の確率変数 X_i が同じ分布に従い, 有限な分散を持っているとする. また, $\mu \equiv E[X_i]$ で X_i の期待値を表す. このとき, 任意の $\epsilon > 0$ に対して

$$\lim_{N \rightarrow \infty} P\left[\left|\frac{S_N}{N} - \mu\right| > \epsilon\right] = 0 \quad (2.2.1)$$

が成り立つ.

この定理は式 (2.2.1) の通り, 「 $\frac{S_N}{N}$ が μ からはずれる確率」が $N \rightarrow \infty$ でゼロに行く, ことを主張している. 直感的にも, これはうなずける結果であるよね.

上の定理は, 以下の評価からすぐに証明される.

²分散の平方根は「標準偏差」とよばれる

命題 2.2.2 問 2.1.3 の確率変数 X_i の平均を μ , 分散を σ^2 とし, $S_N \equiv \sum_{i=1}^N X_i$ を定義する. このとき, 任意の $\epsilon > 0$ に対して

$$P\left[\left|\frac{S_N}{N} - \mu\right| > \epsilon\right] \leq \frac{\sigma^2}{N\epsilon^2} \quad (2.2.2)$$

が成り立つ.

証明:

以下で説明するチェビシェフの不等式そのものである. □

中心極限定理への伏線として, 上の命題の主張するところを吟味しておこう. (2.2.2) の右辺は正の定数 ϵ に対しては (ϵ がどんなに小さくとも) $N \rightarrow \infty$ でゼロになってしまう. つまり, $N \gg 1$ では, $\frac{S_N}{N}$ はまあ, ほとんど μ に等しいわけだ. つまり $N \rightarrow \infty$ につれて, $\frac{S_N}{N}$ は μ の周りの非常に狭い範囲に集中して分布していくことがわかる.

2.2.1 チェビシェフの不等式とその仲間 (対数の弱法則の証明)

「分散は確率変数のばらつきの目安を与える」と以前に言い忘れたように思うが, ともかく分散はばらつきの目安である. しかし, ここではもう少し定量的な議論を行う. ここでも確率空間 (Ω, \mathcal{F}, P) 上の実数値確率変数 X を考える.

まず, $A \in \mathbb{R}$ について

$$P[X \in A] = \langle I[X \in A] \rangle \quad (2.2.3)$$

であることに注意しておこう. このところの A は正確には $A \in \mathcal{B}_1$ に対して, と言うべきであるが, いちいち断らない. この辺りがヤヤコシイ人は A として任意の開区間を考えておけばまあ十分.

命題 2.2.3 (マルコフの不等式) 正の値のみをとる確率変数 X と任意の正の数 a に対して,

$$P[X \geq a] \leq \frac{\langle X \rangle}{a} \quad (2.2.4)$$

が成立 (勿論, 右辺の期待値が存在しないときは右辺には意味がないけど.)

命題 2.2.4 (チェビシェフの不等式) 確率変数 X の期待値を μ , 分散を $\text{Var}[X]$ と書くと, 任意の正の数 a に対して,

$$P[|X - \mu| \geq a] \leq \frac{\text{Var}[X]}{a^2} \quad (2.2.5)$$

が成立 (勿論, 右辺の分散が存在しないときは右辺には意味がないけど.)

証明:

これらの不等式は (2.2.3) を用いると簡単に証明される. マルコフの不等式なら

$$\langle X \rangle \geq \langle X I[X \geq a] \rangle \geq \langle a I[X \geq a] \rangle = a \langle I[X \geq a] \rangle = a P[X \geq a]. \quad (2.2.6)$$

チェビシェフの不等式なら

$$\text{Var}[X] = \langle |X - \mu|^2 \rangle \geq \langle |X - \mu|^2, I[X \geq a] \rangle \geq \langle a^2 I[X \geq a] \rangle = a^2 \langle I[X \geq a] \rangle = a^2 P[X \geq a]. \quad (2.2.7)$$

□

調子に乗って似たような不等式を作ることもしできる．例えば，

$$P[|X - \mu| \geq a] \leq \frac{\langle |X - \mu|^n \rangle}{a^n} \quad (a > 0, n \text{ は任意の正の整数}) \quad (2.2.8)$$

同様に，任意の $a, b > 0$ に対して

$$P[|X - \mu| \geq a] \leq \frac{\langle e^{b|X - \mu|} \rangle}{e^{ab}}. \quad (2.2.9)$$

また，マルコフの不等式の仲間として (X が非負の値しかとらないとき)

$$P[X \geq a] \leq \frac{\langle e^{bX} \rangle}{e^{ab}} \quad (2.2.10)$$

など．これらの不等式は勿論，右辺の期待値が存在しなければ意味がないが，存在する場合には（特に $a \rightarrow \infty$ について）強力なものになる．これらの不等式の一つの応用は対数の弱法則の証明であるが，もちろん，それにとどまらない．

5月7日の連絡：特に連絡事項はありません．採点方針についてはまだ確定はしていませんが，一番初めの方針とは異なり，中間試験は無くすることは確実です．
 なお，先週の講義中にも訂正しましたが，定理 2.2.1（大数の弱法則）にでてくる X_i は独立かつ同分布な確率変数のつもりです（本当はもっと一般の場合でも成り立つが，ここでは簡単なものに限る．）

確率論 I, 確率論概論 I 第 2 回レポート問題

面白い問題にはならなかった（どこまで進めるかの問題もあって）．以下は単なる計算問題であるが，分量も考えて，どちらか一つだけやれば十分に合格点とする．

問 3： チェビシエフの不等式などを実地に試してみる問題．

指数分布に従う確率変数 X がある．これは非負の値をとる確率変数で， X が $[a, b]$ の値をとる確率が ($\lambda > 0$ は定数)

$$P[a \leq X \leq b] = \lambda \int_a^b e^{-\lambda x} dx \quad (0 \leq a < b) \quad (2.2.11)$$

で与えられているものである．

1. 確率変数 X の累積分布関数を求めよ．
2. $a > 0$ に対し， X が a 以上の値をとる確率 $P[X \geq a]$ を計算せよ．後のため，この結果を $f(a) \equiv P[X \geq a]$ と書く．
3. X の平均と分散を計算せよ．
4. $P[X \geq a]$ をチェビシエフの不等式を用いて評価してみよう．勿論，その結果は上で求めた $f(a)$ よりは良くないはずである．どのくらい悪いか（特に $a \rightarrow \infty$ で），実感しよう．
5. （やれる人だけで良い）チェビシエフの不等式の代わりに何か別の不等式を用いて， $P[X \geq a]$ の評価を改良することを考えてみよう．

問 4： 1 以上の値をとり，その分布が ($\lambda > 1$ は定数)

$$P[a \leq Y \leq b] \equiv (\lambda - 1) \int_a^b \frac{dx}{x^\lambda} \quad (1 \leq a < b) \quad (2.2.12)$$

で与えられている確率変数 Y がある．この Y に対して，上の問 3 の X と同じ事をくり返して行え．ただし， λ の値によっては，期待値や分散が存在しなかったり，チェビシエフの不等式が使えなかったりするかも知れないよ．不等式を用いた場合の $P[Y > a]$ の確率の評価の悪さは X の場合と比べてどう変わっただろうか？（ついでに： Y の分布がどのようなであれば，チェビシエフの不等式を用いても，そんなに損をしないだろうか？）

（実はこの後に，大数の弱法則を具体的に経験してもらおう — 命題 2.2.2 を試してみる — 問題を入れようと思ったのだが，余りにもマンマになってしまったので，やめました．各自，自分でやってみようよ．）

締め切りなど：

締め切りは 2002 年 5 月 10 日（金）の 17:00 ，

提出場所は僕の部屋（理 1-508）の前の封筒かポスト

用紙はできうる限り A4 の紙を用いる（B5 などの小さい紙は紛れてなくなるかも）

とします．

レポートのお約束:

- 友達と相談しても, 本を調べても, 何をやっても良いから, 自分で理解した範囲を書くこと. その際, 参考文献や議論した友達の名前も明記すること (友達と議論したり, 本を見たからと言って悪い点をつける, などと言うことは絶対にしない. 一番大事なのは自分でわかったところを表現することだから, それまでの過程で何をやっても問題ない.)
- なお, 問題の番外編として, 今までの講義内容・講義形態についての感想, 不満, 文句, このように改善すべしとの意見などもできるだけ書いてください. お願いします.

—以下, レジユメの続き—

2.3 確率変数の収束について

大数の法則をやっている途中だが, ここで確率変数の収束について, 4つの異なる概念を定義しておく. この内の3つは主要な極限定理に必要なものである.

定義 2.3.1 確率空間 (Ω, \mathcal{F}, P) と, その上の実確率変数 X と実確率変数の列 X_1, X_2, X_3, \dots が与えられたものとする. このとき, $X_n \rightarrow X$ のいろんなバージョンを考えたい.

- X_1, X_2, X_3, \dots が X に概収束する, または「確率1で収束する」(converges with probability 1, または converges almost surely) とは,

$$P\left[\lim_{n \rightarrow \infty} X_n \neq X\right] = 0 \quad (2.3.1)$$

なることで, $X_n \xrightarrow{\text{a.s.}} X$, または $X_n \rightarrow X$ a.s と書く.

- X_1, X_2, X_3, \dots が X に確率収束する (converges in probability) とは,

$$\forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} P\left[|X_n - X| > \epsilon\right] = 0 \quad (2.3.2)$$

なることで, $X_n \xrightarrow{P} X$, または $X_n \rightarrow X$ in pr. と書く.

- X_1, X_2, X_3, \dots が X に法則収束する (converges in distribution) とは, X の分布関数を F , X_n の分布関数を F_n と書くとき,

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad \text{が, } F \text{ の全ての連続点 } x \text{ で成り立つ} \quad (2.3.3)$$

なることで, $X_n \xrightarrow{D} X$, または $X_n \rightarrow X$ in law と書く.

- (これはおまけ) 1 以上の正の数 p を固定する. X_1, X_2, X_3, \dots が X に L^p -収束する (converges in L^p) とは,

$$\lim_{n \rightarrow \infty} E\left[|X_n - X|^p\right] = 0 \quad (2.3.4)$$

なることで, $X_n \xrightarrow{L^p} X$, または $X_n \rightarrow X$ in L^p と書く.

註 2.3.2 概収束と確率収束の定義が少しわかりにくいかも知れないので, 補足しておく. 概収束の場合, 確率空間の元 ω を一つ固定し, この固定した ω 毎に極限 $\lim_{n \rightarrow \infty} X_n(\omega)$ を考えて, こ

れが $X(\omega)$ に等しいか否かを問題にしている (等しくない確率がゼロ, つまり, 等しくないような ω が無視できるほど少ないなら良い).

一方, 確率収束の場合は, 各 n 毎に $|X_n(\omega) - X(\omega)| > \epsilon$ である確率を問題にしている. つまり, $|X_n(\omega) - X(\omega)| > \epsilon$ となるような ω は, n 毎に異なっても, とにかくその確率がゼロに行けば良い.

註 2.3.3 法則収束の定義で F の連続点だけで収束を要求しているのには, 勿論, 理由がある.

註 2.3.4 上の定義は少し解釈を拡張することで, 実確率変数とは限らない一般の確率変数 X に適用することができる (法則収束に関しては, 以下の命題 2.3.5 の定義を用いる). しかし, 簡単のため実確率変数に限定して話をしていく.

法則収束は, 以下の (2.3.5) で定義することもあるが, これは (2.3.3) と同等なので, 問題ない.

命題 2.3.5 定義 2.3.1 の法則収束の定義の条件 (2.3.3) は, 以下とも同値である: 任意の有界な²連続関数 f に対して,

$$\lim_{n \rightarrow \infty} E[f(X_n)] = E[f(X)] \quad (2.3.5)$$

が成立すること.

さて, 4 つもの収束概念が出てきて少し混乱すると思う. そこで, この講義では L^p -収束以外の 3 つ さえ理解できれば十分だとしてしよう.

2.3.1 種々の収束の相互関係

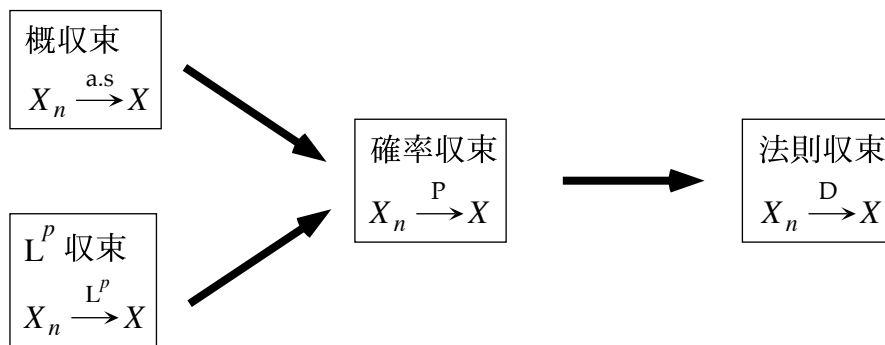
さて, このようにいろいろな定義が出てきたので, 当然, これらがどのような関係にあるのが問題にある. これについては以下の定理がなりたつ.

定理 2.3.6 (種々の収束の強弱) 定義 2.3.1 の種々の収束の間には, 以下の強弱関係がある. $\{X_n\}$ が X に

- 概収束するならば, 確率収束する.
- L^2 収束するならば, 確率収束する.
- 確率収束するならば, 法則収束する.

模式的には, 下図のようになっている. なお, 以上に掲げた強弱関係以外は, 一般には成り立たない (反例がある).

²関数 f が有界とは, 大きな定数 C によって $|f(x)| \leq C$ がなりたつ (すべての x に対して) こと



強弱関係が成り立たない例をいくつか挙げておこう.

- 独立な確率変数列 X_n を,

$$X_n \equiv \begin{cases} 1 & (\text{確率 } 1/n \text{ で}) \\ 0 & (\text{確率 } 1 - 1/n \text{ で}) \end{cases} \quad (2.3.6)$$

と定義すると, X_n は 0 に確率収束するが, 概収束しない.

- X を, 0 と 1 の値を確率 $1/2$ ずつとる確率変数と定義する. また, X_1, X_2, X_3, \dots を, X と全く同じ確率変数とする (全く同じ, の意味は, $X = 0$ なら $X_n = 0$ だし, $X = 1$ なら $X_n = 1$ だということ). また, $Y \equiv 1 - X$ も定義する. すると, X_n は Y に法則収束するが, 他のだどのような意味でも Y には収束しない.

2.3.2 概収束ならば確率収束, の証明

この辺りの関係を理解するため, 以下の命題をまず, 証明しておく.

命題 2.3.7 確率変数列 $\{X_n\}$ が 確率変数 X に概収束する (converges almost surely) 必要十分条件は, 任意の $\epsilon > 0$ に対して

$$\lim_{m \rightarrow \infty} P[m \text{ 以上の } n \text{ で } |X_n - X| > \epsilon \text{ なるものが存在}] = 0 \quad (2.3.7)$$

なることである.

註 2.3.8 上の条件 (2.3.7) は (余事象を考えると)

$$\lim_{m \rightarrow \infty} P[\text{すべての } n > m \text{ に対して } |X_n - X| \leq \epsilon] = 1 \quad (2.3.8)$$

とも同値である.

証明:

この命題を理解するため, $\epsilon > 0$ に対して

$$A_n(\epsilon) \equiv \{|X_n - X| \leq \epsilon\} \quad (2.3.9)$$

なる事象を定義し, さらに, B_m を

$$B_m(\epsilon) \equiv \bigcap_{n \geq m} A_n(\epsilon) \equiv \{ \text{すべての } n \geq m \text{ で } |X_n - X| \leq \epsilon \} \quad (2.3.10)$$

と定義すると, これは (2.3.8) に現れている事象そのものになっている.

さて, X_n が X に概収束しているとしよう. この意味は,

$$\Omega_0 \equiv \left\{ \omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \right\} \quad (2.3.11)$$

を定義すると, $P[\Omega_0] = 1$ ということである. そして, $\omega \in \Omega_0$ に対しては, 任意の $\epsilon > 0$ に対して, ϵ と ω に依存するだろう $m(\omega, \epsilon)$ があって,

$$n > m(\omega, \epsilon) \implies |X_n(\omega) - X(\omega)| < \epsilon \quad (2.3.12)$$

ということになる (ここでは単に, ω を固定した上で, 数列 $X_n(\omega)$ が $X(\omega)$ に収束する, という定義を書いただけである.)

でもこの条件と上の $B_m(\epsilon)$ の定義を比べると, このような $\omega \in \Omega_0$ は, 実は $B_m(\epsilon)$ の元になっていることがわかる ($m \geq m(\omega, \epsilon)$). つまり, $\Omega_0 \subset \bigcup_{m \geq 1} B_m(\epsilon)$ が結論される. ところが, 概収束するということは $P[\Omega_0] = 1$ ということであったし, $B_m(\epsilon)$ は (固定した ϵ に対して) m に関して単調増加であるから, 測度の単調収束定理から

$$\lim_{m \rightarrow \infty} P[B_m(\epsilon)] \geq P[\Omega_0] = 1 \quad (2.3.13)$$

が成り立つ. $P[B_m(\epsilon)]$ はいつでも 1 以下であるので,

$$\lim_{m \rightarrow \infty} P[B_m(\epsilon)] = 1 \quad (2.3.14)$$

が結論できる. つまり, 「概収束ならば (2.3.8)」が言えた.

逆に (2.3.8) が成り立っているとしよう. すると, $B(\epsilon) \equiv \bigcup_{m \geq 1} B_m(\epsilon)$ は確率 1 である. そこで

$B \equiv \bigcap_{j=1}^{\infty} B(1/j)$ を考えると, 確率の連続性から ($B(1/j)$ は j の減少列)

$$P[B] = \lim_{j \rightarrow \infty} P[B(1/j)] = \lim_{j \rightarrow \infty} 1 = 1 \quad (2.3.15)$$

が成立する (いつでも $P[B(\epsilon)] = 1$ であったことを思い出せ).

さて, $\omega \in B$ に対しては, B の定義から, (2.3.12) が $\epsilon = 1/j$ について成立する ($j = 1, 2, 3, \dots$). しかし, これは (2.3.12) が全ての $\epsilon > 0$ について成立することと同じである. つまり, 任意の $\omega \in B$ に対して $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ であることが示せた. 確率 1 である集合 B 上で X_n が X に収束するのであるから, 概収束が証明できた. \square

概収束ならば確率収束, の証明

確率収束と言うことは

$$\lim_{n \rightarrow \infty} P[|X_n - X| > \epsilon] = 0 \quad (2.3.16)$$

ということであったが, 上の条件は明らかに (why?) (2.3.7) から導かれる. すなわち, 概収束すれば確率収束する. \square

2.3.3 確率収束ならば法則収束, の証明

X_n が X に確率収束するものとし, X_n と X の分布関数を F_n と F と書く. 任意の $\epsilon > 0$ に対して

$$F_n(a) \equiv P[X_n \leq a] = P[\{X_n \leq a\} \cap \{X \leq a + \epsilon\}] + P[\{X_n \leq a\} \cap \{X > a + \epsilon\}] \quad (2.3.17)$$

と無理矢理変形することができる ($\{X \leq a + \epsilon\} \cup \{X > a + \epsilon\}$ が全事象になっていることを用いた). 前半分は

$$P[\{X_n \leq a\} \cap \{X \leq a + \epsilon\}] \leq P[X \leq a + \epsilon] = F(a + \epsilon), \quad (2.3.18)$$

後ろ半分は

$$P[\{X_n \leq a\} \cap \{X > a + \epsilon\}] \leq P[|X_n - X| > \epsilon] \quad (2.3.19)$$

と押さえられるので, 結局,

$$F_n(a) \leq F(a + \epsilon) + P[|X_n - X| > \epsilon] \quad (2.3.20)$$

となる. 同様に,

$$\begin{aligned} F(a - \epsilon) &= P[X \leq a - \epsilon] = P[\{X \leq a - \epsilon\} \cap \{X_n \leq a\}] + P[\{X \leq a - \epsilon\} \cap \{X_n > a\}] \\ &\leq F_n(a) + P[|X_n - X| > \epsilon] \end{aligned} \quad (2.3.21)$$

を得るので, 2つを併せて

$$F(a - \epsilon) - P[|X_n - X| > \epsilon] \leq F_n(a) \leq F(a + \epsilon) + P[|X_n - X| > \epsilon] \quad (2.3.22)$$

となる. ここで $n \rightarrow \infty$ を考えると (確率収束の定義から $P[|X_n - X| > \epsilon] \rightarrow 0$ なので)

$$F(a - \epsilon) \leq \liminf_{n \rightarrow \infty} F_n(a) \leq \limsup_{n \rightarrow \infty} F_n(a) \leq F(a + \epsilon) \quad (2.3.23)$$

が得られた (真ん中を単に \lim としていないのは, 極限が存在するかどうかはわかっていないから). F が a で連続ならば

$$\lim_{\epsilon \rightarrow +0} F(a - \epsilon) = \lim_{\epsilon \rightarrow +0} F(a + \epsilon) = F(a) \quad (2.3.24)$$

であるから (連続の定義), (2.3.23) で $\epsilon \rightarrow 0$ として,

$$F(a) \leq \liminf_{n \rightarrow \infty} F_n(a) \leq \limsup_{n \rightarrow \infty} F_n(a) \leq F(a) \quad (2.3.25)$$

すなわち

$$\lim_{n \rightarrow \infty} F_n(a) = F(a) \quad (2.3.26)$$

が得られ, (2.3.3) の条件を満たすことがわかった. \square

2.3.4 L^p -収束ならば確率収束, の証明 (おまけのおまけ)

だめ押しで載せておく. マルコフの不等式を $|X_n - X|^p$ という確率変数に適用して,

$$P[|X_n - X| > \epsilon] = P[|X_n - X|^p > \epsilon^p] \leq \frac{E[|X_n - X|^p]}{\epsilon^p} \quad (2.3.27)$$

が成り立つことがわかる. ϵ を固定したままで両辺で $n \rightarrow \infty$ とすると,

$$\lim_{n \rightarrow \infty} P[|X_n - X| > \epsilon] = 0, \quad (2.3.28)$$

つまり確率収束の条件が得られた. \square

5月14日の連絡: 来る6月11日は休講です.

2.4 大数の強法則

平均が μ , 分散が σ^2 の独立同分布な確率変数 X_i ($i = 1, 2, \dots$) に対して $S_N \equiv \sum_{i=1}^N X_i$ を定義する. 大数の弱法則は全ての $\epsilon > 0$ に対して

$$\lim_{n \rightarrow \infty} P \left[\left| \frac{S_N}{N} - \mu \right| > \epsilon \right] = 0 \quad (2.4.1)$$

を主張した. それに対して, 大数の強法則は

定理 2.4.1 (大数の強法則, Strong Law of Large Numbers)

平均が μ , 分散が有限の独立同分布な確率変数 X_i ($i = 1, 2, \dots$) に対して $S_N \equiv \sum_{i=1}^N X_i$ を定義する. このとき, $N \rightarrow \infty$ で

$$\frac{S_N}{N} \xrightarrow{\text{a.s.}} \mu \quad \left(\frac{S_N}{N} \text{ は } \mu \text{ に概収束する} \right) \quad (2.4.2)$$

が成り立つ. すなわち,

$$P \left[\lim_{n \rightarrow \infty} \frac{S_N}{N} = \mu \right] = 1 \quad (2.4.3)$$

が成り立つ.

と主張する. 強法則と弱法則の違いは $\lim_{n \rightarrow \infty}$ が $P[\dots]$ の外にあるか, 中にあるか (つまり, 収束が概収束か, 確率収束か) だけなのだが, これは時には大きな違いになる. 勿論, 名前の通り, 強法則の方が強い. つまり, (2.4.3) が成り立つような S_N に対しては, (2.4.1) も成り立つ.

さて, この定理を厳密に理解するには, 標本空間を無限にしないといけないので, 準備が大変である. 定理の (2.4.3) の左辺では $\lim_{n \rightarrow \infty} \frac{S_N}{N}$ が確率の中に現れている. つまり, このような「極限をとった後の事象」の確率を考えているわけで, このような確率を定義するには極限をとった後の事象が入っているような確率空間を作ってやらないといけない (弱法則の方は N が有限である場合に確率を先に計算してしまい, その確率がどんな極限に行くか, を考えているので, 確率空間は基本的に有限のものでも可能). これは下の小節で解説する.

2.4.1 無限直積空間の構成 (少し advanced)

大数の強法則における $\lim_{n \rightarrow \infty} \frac{S_N}{N}$ がちゃんと理解できるように, 舞台を設定していく. ここは少しだけ進んだ内容だから, イヤな人は大体の感じだけわかれば十分.

2.1.1 節で有限個の確率空間の直積を定義した. ここでは無限個の確率空間の直積が必要になる. つまり, 確率空間の無限列 $(\Omega_i, \mathcal{F}_i, P_i)$ ($i = 1, 2, 3, \dots$) の直積 (Ω, \mathcal{F}, P) を作りたい. しかし, $2^{\mathbb{N}}$ は非加算無限 (!) であるため, 有限個の時のようには構成できない. そこで, 以下のように可測集合を限定して作っていく.

定義 2.4.2 (無限個の確率空間の直積) 確率空間の無限列 $(\Omega_i, \mathcal{F}_i, P_i)$ ($i = 1, 2, 3, \dots$) の直積 (Ω, \mathcal{F}, P) は以下のように構成する.

- まず, $\Omega = \Omega_1 \times \Omega_2 \times \dots$ とする (これは問題なし).
- 次に \mathcal{F} だが, これは以下のように進む.
 1. 初めの n 個, つまり $(\Omega_i, \mathcal{F}_i, P_i)$ ($i = 1, 2, 3, \dots, n$) の直積確率空間を 2.1.1 節の通りに作り, これを $(\Omega^{(n)}, \mathcal{F}^{(n)}, P^{(n)})$ とする.
 2. Ω から $\Omega^{(n)}$ への射影を π_n と書く. つまり, Ω の元 $(\omega_1, \omega_2, \dots, \omega_n, \dots)$ に $\Omega^{(n)}$ の元 $(\omega_1, \omega_2, \dots, \omega_n)$ を対応させるものが π_n (要するに, $n+1$ より先を見ないことにしよう, と言うのが π_n).
 3. Ω の部分集合で, 筒集合 (cylinder set) とよばれるものに注目する. これは適当な $A^{(n)} \in \mathcal{F}^{(n)}$ を持ってきて $\pi_n^{-1}(A^{(n)})$ と書けるものの全体. 平たくいえば, Ω の要素の内, 1 番目から n 番目がみんな同じ奴らだけを集めて出来る集合のこと.
 4. \mathcal{B}_n を上の筒集合の全体とする: $\mathcal{B}_n \equiv \{\pi_n^{-1}(A^{(n)}) \mid A^{(n)} \in \mathcal{F}^{(n)}\}$.
 5. $\mathcal{A} \equiv \cup_{n=1}^{\infty} \mathcal{B}_n$ を定義すると, \mathcal{A} は Ω の加法族となる (確かめよ).
 6. \mathcal{F} を \mathcal{A} を含む最小の σ -field にとる: $\mathcal{F} = \sigma(\mathcal{A})$.
- 最後に P だが, ここまで来れば予想はつくだろう.
 1. まず $A \in \mathcal{A}$ が適当な n と $A^{(n)} \in \mathcal{F}^{(n)}$ を使って $A = \pi_n^{-1}(A^{(n)})$ と書ける場合には, $P[A] \equiv P^{(n)}[A^{(n)}]$ とする.
 2. それ以外の \mathcal{F} の元に対しては, やはり σ -加法性で拡張して定義する

このように構成した (Ω, \mathcal{F}, P) は, これが直積である事を強調するため, $(\prod_{n=1}^{\infty} \Omega_n, \prod_{n=1}^{\infty} \mathcal{F}_n, \prod_{n=1}^{\infty} P_n)$ とも書く.

ようやく, 大数の強法則の正確な意味を理解できるようになった. つまり, 大数の強法則では,

$$\lim_{n \rightarrow \infty} \frac{S_N}{N} = \mu \quad (2.4.4)$$

である事象を, 上のようにして構成した確率空間で考える. その結果, この確率は 1 である (つまり, S_N/N が μ に概収束する), と主張している.

2.4.2 大数の強法則の証明 I

大数の強法則の証明はなかなか厄介であるので, まず余分な仮定をおくことで, 簡単な証明を紹介する.

警告: この節では大数の強法則の定理 2.4.1 に余分の仮定

各 X_i は有限な 4 次のモーメントを持つ, つまり, $\langle X_i^4 \rangle < \infty$ である

を付け加えた証明を紹介する.

定理 2.4.3 (Cantelli による大数の法則) X_1, X_2, \dots を独立な確率変数とする (同分布でなくてもよい). 更に, ある定数 C が存在して,

$$\left\langle \left| X_n - \langle X_n \rangle \right|^4 \right\rangle \leq C \quad (n \geq 1) \quad (2.4.5)$$

であるとせよ. このとき, $S_N \equiv \sum_{i=1}^N X_i$ に対して,

$$\frac{S_N - \langle S_N \rangle}{N} \xrightarrow{\text{a.s.}} 0 \quad (2.4.6)$$

が成り立つ. つまり, $\frac{S_N - \langle S_N \rangle}{N}$ は 0 に概収束する.

補題 2.4.4 確率変数 X_n と X が $X_n \xrightarrow{\text{a.s.}} X$ を満たすための十分条件は,

$$\sum_{n=1}^{\infty} P[|X_n - X| \geq \epsilon] < \infty \quad (2.4.7)$$

である.

証明:

$\sum_{n=1}^{\infty} P[|X_n - X| \geq \epsilon]$ が有限であるということは, この和の tail がゼロに行くということである:

$$\lim_{N \rightarrow \infty} \sum_{n=N}^{\infty} P[|X_n - X| \geq \epsilon] = 0 \quad (2.4.8)$$

しかし, これは

$$\sum_{k=N}^{\infty} P[|X_k - X| \geq \epsilon] \geq P\left[\bigcup_{k \geq N} \{|X_k - X| \geq \epsilon\}\right] \geq P\left[\sup_{k \geq N} |X_k - X| \geq \epsilon\right] \quad (2.4.9)$$

に注目すると,

$$\lim_{N \rightarrow \infty} P\left[\sup_{k \geq N} |X_k - X| \geq \epsilon\right] = 0 \quad (2.4.10)$$

を意味する. でもこれは, 正に命題 2.3.7 での概収束の必要十分条件に他ならない. 従って, X_n は X に概収束する. \square

定理 2.4.3 の証明

まず, 注意. $\langle X_n \rangle \neq 0$ の場合は, 新しく $Y_n \equiv X_n - \langle X_n \rangle$ を考えることにすると, この Y_n の期待値はゼロである. だから, 定理の X_n の期待値は全てゼロだと思っても一般性を失わない. さて, 上の補題から,

$$\forall \epsilon > 0, \quad \sum_{N=1}^{\infty} P\left[\left|\frac{S_N}{N}\right| > \epsilon\right] < \infty \quad (2.4.11)$$

が言えれば, 定理は証明されたことになるので, 以下では (2.4.11) を証明していく.

さて, チェビシェフの不等式から,

$$P\left[\left|\frac{S_N}{N}\right| > \epsilon\right] = P\left[\left|\frac{S_N}{N}\right|^2 > \epsilon^2\right] \leq \frac{\left\langle \left|\frac{S_N}{N}\right|^4 \right\rangle}{\epsilon^4} \quad (2.4.12)$$

であるから, (2.4.11) のためには

$$\sum_N \left\langle \left|\frac{S_N}{N}\right|^4 \right\rangle < \infty \quad (2.4.13)$$

が言えれば十分.

そこで, 以下では (2.4.13) を証明する. まず, X_i の独立性を使いやすいように, S_N^4 を分解する:

$$\begin{aligned} S_N^4 &= \left(\sum_{j=1}^N X_j\right)^4 = \sum_{j=1}^N X_j^4 + \frac{4!}{2!2!} \sum_{i<j} X_i^2 X_j^2 \\ &\quad + \frac{4!}{2!1!1!} \sum_{\substack{i \neq j \\ i \neq k \\ j < k}} X_i^2 X_j X_k + 4! \sum_{i \neq j} X_i^3 X_j + \frac{4!}{3!1!} \sum_{i<j<k<l} X_i X_j X_k X_l \end{aligned} \quad (2.4.14)$$

ここで S_N^4 の期待値をとるのであるが, 独立性と $\langle X_j \rangle \equiv 0$ であることを用いると,

$$\langle S_N^4 \rangle = \sum_{j=1}^N \langle X_j^4 \rangle + 6 \sum_{i<j} \langle X_i^2 \rangle \langle X_j^2 \rangle \quad (2.4.15)$$

が得られる. 右辺の第一項は NC で押さえられる. また第 2 項は, Schwartz の不等式から,

$$\langle X^2 \rangle^2 \leq \langle X^4 \rangle \quad \implies \quad \langle X^2 \rangle \leq \sqrt{\langle X^4 \rangle} \quad (2.4.16)$$

が成り立つので, 結局

$$\langle S_N^4 \rangle \leq NC + 3N(N-1)C < 3N^2C \quad (2.4.17)$$

がわかった. これから

$$\sum_N \left\langle \left|\frac{S_N}{N}\right|^4 \right\rangle < \sum_N \frac{3N^2C}{N^4} = 3C \sum_N \frac{1}{N^2} < \infty \quad (2.4.18)$$

となり, (2.4.13) が証明できた. □

2.4.3 大数の強法則の証明 II (おまけ: かなり趣味が入ってます)

余分な条件無しで大数の強法則を証明するのはなかなか厄介ではある. ここでは普通の確率論の本にはあまり載っていない(?), ひねくれた方法 (method of subsequence) を紹介しよう. 定理 2.4.3 の証明と同じく, X_n の期待値はゼロだとしても一般性を失わない.

Step 1. さて, 上の定理 2.4.3 の証明と同じく, 補題 2.4.4 を利用したいのだが, 今度は 4 次のモーメントの有限性を仮定していないので, うまくは行かない. そこで, 列 S_N/N を考えることは諦めて, $Z_N \equiv \frac{S_{N^2}}{N^2}$ で与えられる Z_N をまず, 考える. これについては, 任意の $\epsilon > 0$ に対して, チェビシェフの不等式から

$$P[Z_N > \epsilon] \leq \frac{\text{Var}[Z_N]}{\epsilon^2} = \frac{\text{Var}[S_{N^2}]}{(N^2)^2 \epsilon^2} = \frac{N^2 \text{Var}[X_1]}{N^4 \epsilon^2} = \frac{\text{Var}[X_1]}{N^2 \epsilon^2} \quad (2.4.19)$$

が成り立ち,

$$\sum_{N=1}^{\infty} P[Z_N > \epsilon] \leq \sum_{N=1}^{\infty} \frac{\text{Var}[X_1]}{N^2 \epsilon^2} < \infty \quad (2.4.20)$$

となる. これは Z_N が補題 2.4.4 の条件を満たすことを意味するので, Z_N が 0 に概収束することが言えた. つまり,

$$\frac{S_{n^2}}{N^2} \xrightarrow{\text{a.s.}} 0 \quad (2.4.21)$$

である³.

Step 2. 勿論, 我々は $\frac{S_n}{n}$ が 0 に概収束することを言いたい. そのためには, n^2 から $(n+1)^2$ までの間の $\frac{S_m}{m}$ が $\frac{S_{n^2}}{n^2} = Z_n$ などとほとんど変わらないことを言えばよい. 以下, これを目指す. まず, この差を測るために,

$$D_n \equiv \max_{n^2 < m < (n+1)^2} |S_m - S_{n^2}| = \max_{n^2 < m < (n+1)^2} \left| \sum_{k=n^2+1}^m X_k \right| \quad (2.4.22)$$

を導入する. この D_n については以下の不等式

$$E[D_n^2] \leq 3n^2 E[X_1^2] \quad (2.4.23)$$

が成り立つ (証明は Step 4 で). これを認めると, チェビシェフの不等式から

$$P[D_n > n^2 \epsilon] \leq \frac{3n^2 E[X_1^2]}{(n^2 \epsilon)^2} = \frac{3E[X_1^2]}{n^2 \epsilon^2} \quad (2.4.24)$$

が得られるが, これは (2.4.19) と同じ形であるから, S_{n^2} と同じ議論で,

$$\frac{D_n}{N^2} \xrightarrow{\text{a.s.}} 0 \quad (2.4.25)$$

が結論できる.

Step 3. 以上の 2 つを組み合わせよう. まず, $n^2 + 1 \leq m < (n+1)^2$ に対して

$$\frac{|S_m|}{m} \leq \frac{|S_{n^2}| + D_n}{m} \leq \frac{|S_{n^2}| + D_n}{n^2} = \frac{|S_{n^2}|}{n^2} + \frac{D_n}{n^2} \quad (2.4.26)$$

であるので, ここで $m \rightarrow \infty$ とすることを考える. このとき $n \rightarrow \infty$ となるが, 右辺の 2 つの量はともに 0 に概収束することが (2.4.21) と (2.4.25) で保証されている. 従って, 左辺にでてくる S_m/m も 0 に概収束する.

Step 4. 最後に (2.4.23) を証明しよう. まず, 各 ω に対して, D_n の定義中の \max_m が実現されるような m は $n^2 + 1$ 以上 $(n+1)^2 - 1$ 以下ではあるから, 最大値を与える m を m_0 と書いて

$$(D_n)^2 = |S_{m_0} - S_{n^2}|^2 \leq \sum_{m=n^2+1}^{(n+1)^2-1} |S_m - S_{n^2}|^2 \quad (2.4.27)$$

³これは次の一般的な定理の特殊例だとも考えられる. X_n が X に確率収束するとき, その適当な部分列をとって, その部分列は X に概収束するようになれる.

はいつでも成り立つ (最後の所は単に, どんな m_0 でも右辺の和に含まれている, ことを表現しただけ). この両辺の期待値をとると (期待値をとるときに m_0 はいろいろと変わるが, 上の式の右辺は m_0 にかかわらずいつでも成り立つからこいつの期待値をとればよい),

$$E[D_n^2] \leq \sum_{m=n^2+1}^{(n+1)^2-1} E[|S_m - S_{n^2}|^2] = \sum_{m=n^2+1}^{(n+1)^2-1} E\left[\left(\sum_{k=n^2+1}^m X_k\right)^2\right] \quad (2.4.28)$$

ここで $\left(\sum_{k=n^2+1}^m X_k\right)^2$ を展開してから X_i 達が互いに独立であることを用いると

$$E\left[\left(\sum_{k=n^2+1}^m X_k\right)^2\right] = \sum_{j=n^2+1}^m \sum_{k=n^2+1}^m E[X_j X_k] = \sum_{j=n^2+1}^m E[X_j^2] = (m - n^2)E[X_1^2] \quad (2.4.29)$$

であるから, 結局

$$E[D_n^2] \leq \sum_{m=n^2+1}^{(n+1)^2-1} (m - n^2) E[X_1^2] = n(2n + 1) E[X_1^2] \quad (2.4.30)$$

を得る. □

自主的に計算してみる問題 (以下は講義中に黒板で説明したものである.)

来週からの中心極限定理に備えて, 以下の計算問題をやってみておいて欲しい. 講義時間の関係もあり, 以下のような泥臭い計算を講義中に紹介することはできないだろう. しかし綺麗な定理のみならず, 以下のような地道な計算も, 理解を深めるためには重要である. これはレポート問題ではないが, 来週からの講義の予習だと思って出来る範囲で取り組んで欲しい.

独立・同分布な確率変数 X_1, X_2, \dots を以下のように定義する.

$$P[X_i = -2] = \frac{1}{3}, \quad P[X_i = 1] = \frac{2}{3}$$

これについていつものように $S_N \equiv \sum_{i=1}^N X_i$ を定義し, 確率 $P\left[a \leq \frac{S_N}{\sqrt{N}} \leq b\right]$ を計算したい. 目標はこの確率が, $N \rightarrow \infty$ で, 中心極限定理で与えられるものに行くことを納得する事である. ただし, これを近似無しで計算することは不可能に近いので, 以下のような方針で行えばよい.

- S_N がいろいろな値をとる確率は二項分布になるから, その確率を書き下すことは簡単である.
- そのままでは 2 項係数が出てきて始末に負えないはずだから, Stirling の公式を使うことにする.

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \quad (n \rightarrow \infty)$$

と言うのがそれだが, これを小さい n についても等式として使って良いから ($n = 0$ はちょっと別扱いが必要), S_N がいろいろな値をとる確率を少しでも簡単にせよ.

- 目標は

$$\lim_{N \rightarrow \infty} P\left[a \leq \frac{S_N}{\sqrt{N}} \leq b\right] = \int_a^b f(x) dx$$

と書けるか? 書けるとすれば $f(x)$ は何か? を見つけることである. 上で簡単にしたこととうまく用いて, このような $f(x)$ があるのか, あればそれは何なのか, 考えよ.

5月21日の連絡: 6月11日は休講にします. また, 期末テストを講義の最終日(7月23日)に行います.

なお, 成績を計算する式は, 大体のところ, 以下のようにする予定(細部を変更する可能性が有ります):

- 期末試験の点を $G_{\text{試験}}$, レポートの点を $G_{\text{レポ}}$ とする(いずれも 100 点満点で).
- レポート点 $G_{\text{レポ}}$ は学期中に出题したレポートの合計点であるが, 簡単な問題を解いた部分を 60% 以上にする(つまり, ある程度解く努力をしていれば 60% はもらえる ということ).
- 最終成績 $G_{\text{最終}}$ は以下で算出する(細かい比率は少し変えるかも):

$$G_a = (0.5) \times G_{\text{試験}} + (0.5) \times G_{\text{レポ}}, \quad G_b = (0.8) \times G_{\text{試験}} + (0.2) \times G_{\text{レポ}},$$

$$G_{\text{最終}} = \max \{ G_a, G_b \}.$$

確率論 I, 確率論概論 I 第 3 回レポート問題

余りに簡単な問題だと退屈だろうから, 簡単なのと少し難しいのを混ぜてみた. 以下の 5-a, 5-b はみんなができるはずだからやって下さい(これだけであれば合格点はある). 5-c, 5-d も, そんなには難しくないけど, ちと面倒かな.

問 5: 対数の弱法則などを実地に証明してみる問題. 講義では独立・同分布の X_i に対して大数の弱法則などを証明したが, 「同分布」の条件は少しくらい緩めても大数の弱法則は成り立つ. それを実感しよう.

確率変数 X_j ($j = 1, 2, 3, \dots$) に対して, 今までのように $S_N \equiv \sum_{j=1}^N X_j$ と定義する. 以下のそ

れぞれの場合に, $\frac{S_N}{N}$ に対して大数の弱法則が成り立つこと, すなわち, 適当な定数 c が存在して

$$\frac{S_N}{N} \xrightarrow{P} c \quad (N \rightarrow \infty)$$

であることを証明せよ.

- a. 確率変数 X_j は同分布で ($0 < p < 1$)

$$P[X_1 = 1] = p, \quad P[X_1 = -1] = 1 - p$$

を満たす(これは講義でやった大数の弱法則でカバーされているが, 足慣らしの意味でやりなおすことを奨める).

- b. 確率変数 X_j は独立だが同分布ではなく,

$$P[X_j = 0] = 1 - \frac{1}{j^2}, \quad P[X_j = j] = P[X_j = -j] = \frac{1}{2j^2}$$

を満たす.

- c. 確率変数 X_j は独立だが同分布ではなく, ある $\epsilon > 0$ に対して

$$P[X_j = 0] = 1 - \frac{1}{j^{1+\epsilon}}, \quad P[X_j = j] = P[X_j = -j] = \frac{1}{2j^{1+\epsilon}}$$

を満たす.

d. (ちょっとムズイかも) 確率変数 X_j は独立だが同分布ではなく,

$$P[X_j = 0] = 1 - \frac{1}{j \log(j+1)}, \quad P[X_j = j] = P[X_j = -j] = \frac{1}{2j \log(j+1)}$$

を満たす.

問 6: (完全なボーナス問題: 根性のある人は考えてみよう) 上のそれぞれの場合, S_N/N は概収束もしているだろうか?

締め切りなど:

締め切りは 2002 年 5 月 27 日 (月) の 14:00 (曜日と時刻の変更に注意!),

提出場所は僕の部屋 (理 1-508) の前の封筒かポスト

用紙はできうる限り A4 の紙を用いる (B5 などの小さい紙は紛れてなくなるかも)

とします.

レポートのお約束:

- 友達と相談しても, 本を調べても, 何をやっても良いから, 自分で理解した範囲を書くこと. その際, 参考文献や議論した友達の名前も明記すること (友達と議論したり, 本を見たからと言って悪い点をつける, などと言うことは絶対にしない. 一番大事なのは自分でわかったところを表現することだから, それまでの過程で何をやっても問題ない.)
- なお, 問題の番外編として, 今までの講義内容・講義形態についての感想, 不満, 文句, このように改善すべしとの意見などもできるだけ書いてください. お願いします.

—————以下, レジユメの続き—————

2.5 おまけ (ちょっと一休み): Weierstrass の多項式近似定理

この節の内容は本当のおまけであるが, 大数の法則などの面白い応用例になるので, ここに載せておく. レジユメが多くなってイヤだ, という人は, この小節は無視しても一向に構わない. 講義でもほとんど触れることはできないだろう.

微分積分学で習ったかも知れないが, Weierstrass の近似定理と言うのがある. これは

定理 2.5.1 閉区間 $[0, 1]$ で定義された連続関数 $f(x)$ がある. これに対して, 適当な多項式をとることにより, $f(x)$ をいくらでも良い精度で一様に近似できる. 具体的には, 任意の $\epsilon > 0$ に対し, 適当な多項式 $p(x)$ をとって,

$$|f(x) - p(x)| < \epsilon \quad (\forall x \in [0, 1]) \quad (2.5.1)$$

を成立させることができる.

と言う定理である.

この定理はもちろん「普通の」解析的手段で証明できるが、大数の法則などにでてくるアイディアによる簡単な別証明がある。以下に紹介しよう。

証明：

まず、 f に対してベルンシュタインの多項式を

$$B_n(x) \equiv \sum_{k=0}^n \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k} \quad (2.5.2)$$

と定義する。この $B_n(x)$ が $f(x)$ の近似多項式の候補である。実際にどのくらい近似できているかを調べるため、以下のように強引な確率論的解釈を行う。

n 個の確率変数 X_i ($i = 1, 2, \dots, n$) が ($0 \leq x \leq 1$)

$$X_i = \begin{cases} 0 & (\text{確率 } 1-x \text{ で}) \\ 1 & (\text{確率 } x \text{ で}) \end{cases} \quad (2.5.3)$$

となっている場合を考え、 $S_n \equiv \sum_{i=0}^n X_i$ と定義しよう。 $S_n = k$ となる確率は

$$\binom{n}{k} x^k (1-x)^{n-k} \quad (2.5.4)$$

であるので、ベルンシュタインの多項式の定義を思い出すと、

$$B_n(x) = \left\langle f\left(\frac{S_n}{n}\right) \right\rangle \quad (2.5.5)$$

と書けることがわかる(ここで $\langle \cdot \rangle$ は確率変数 X_1, X_2, \dots, X_n の分布に関する期待値を表す)。

以上をもとにして、証明を進める。キーになるのは、 n を大きくすると、大数の法則によって

$$\frac{S_n}{n} \approx \langle X_1 \rangle = x \quad \text{なので,} \quad B_n(x) \approx f(x) \quad (?) \quad (2.5.6)$$

となるだろう、とのアイディアである — もしこうなら、 $f(x)$ を多項式 $B_n(x)$ で近似できたことになる。大筋はこうなのだが、数学的に厳密にするために以下のように工夫する。

まず、 $f(x)$ 自身は確率変数とは何の関係もないから $\langle \cdot \rangle$ の中へ入れることができる：

$$f(x) - B_n(x) = \langle f(x) \rangle - \left\langle f\left(\frac{S_n}{n}\right) \right\rangle = \left\langle f(x) - f\left(\frac{S_n}{n}\right) \right\rangle \quad (2.5.7)$$

両辺の絶対値をとると、

$$\left| f(x) - B_n(x) \right| = \left| \left\langle f(x) - f\left(\frac{S_n}{n}\right) \right\rangle \right| \leq E \left\langle \left| f(x) - f\left(\frac{S_n}{n}\right) \right| \right\rangle \quad (2.5.8)$$

となる(最後の不等式は $\int |f(x)dx| \leq \int |f(x)|dx$ と同じ)。この最後の項を更に二つに分ける(チェビシェフの不等式の証明を逆にたどっている感じ): $\delta > 0$ を小さい数として

$$\left\langle \left| f(x) - f\left(\frac{S_n}{n}\right) \right| \right\rangle = \left\langle \left| f(x) - f\left(\frac{S_n}{n}\right) \right| I\left[\left|\frac{S_n}{n} - x\right| \geq \delta\right] \right\rangle + \left\langle \left| f(x) - f\left(\frac{S_n}{n}\right) \right| I\left[\left|\frac{S_n}{n} - x\right| < \delta\right] \right\rangle. \quad (2.5.9)$$

さて、第一項については、

$$\left| f(x) - f\left(\frac{S_n}{n}\right) \right| \leq 2 f_{\max} \quad (2.5.10)$$

を用いて (f_{\max} は $[0, 1]$ 上での f の最大値),

$$\left\langle \left| f(x) - f\left(\frac{S_n}{n}\right) \right| I\left[\left|\frac{S_n}{n} - x\right| \geq \delta\right] \right\rangle \leq 2 f_{\max} \left\langle I\left[\left|\frac{S_n}{n} - x\right| \geq \delta\right] \right\rangle = 2 f_{\max} P\left[\left|\frac{S_n}{n} - x\right| \geq \delta\right] \quad (2.5.11)$$

とできるし, 第 2 項は

$$\left| f(x) - f\left(\frac{S_n}{n}\right) \right| I\left[\left|\frac{S_n}{n} - x\right| < \delta\right] \leq \max_{\substack{|s-t| < \delta \\ s, t \in [0, 1]}} |f(s) - f(t)| \quad (2.5.12)$$

を用いて

$$\left\langle \left| f(x) - f\left(\frac{S_n}{n}\right) \right| I\left[\left|\frac{S_n}{n} - x\right| < \delta\right] \right\rangle \leq \max_{\substack{|s-t| < \delta \\ s, t \in [0, 1]}} |f(s) - f(t)| \quad (2.5.13)$$

と押さえることができる. そこで (2.5.8), (2.5.9) から

$$\left| f(x) - B_n(x) \right| \leq 2 f_{\max} P\left[\left|\frac{S_n}{n} - x\right| \geq \delta\right] + \max_{\substack{|s-t| < \delta \\ s, t \in [0, 1]}} |f(s) - f(t)| \quad (2.5.14)$$

が得られた. 更に, この第一項はチェビシエフの不等式から ($\text{Var}[X_1] = x(1-x)$ に注意)

$$P\left[\left|\frac{S_n}{n} - x\right| \geq \delta\right] \leq \frac{x(1-x)}{n\delta^2} \leq \frac{1}{4n\delta^2} \quad (2.5.15)$$

と押さえられているので, 結局

$$\left| f(x) - B_n(x) \right| \leq \frac{f_{\max}}{2n\delta^2} + \max_{\substack{|s-t| < \delta \\ s, t \in [0, 1]}} |f(s) - f(t)| \quad (2.5.16)$$

が得られた.

さてここまで $\delta > 0$ は任意だったから, 右辺第二項が $\epsilon/2$ より小さくなるように δ を選ぼう (これは f が閉区間で連続な関数だからできる — ここは本当は「一様連続性」という概念を使っている). その上で (こうやって選んだ δ に対して) n を十分大きくとって第一項も $\epsilon/2$ より小さくすることができる. つまり, このような δ, n の取り方によって,

$$\left| f(x) - B_n(x) \right| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \quad (2.5.17)$$

が成り立つことがわかった. □

2.6 中心極限定理

本題に戻る. 今までの「大数の法則」では, S_N がその平均値の周りに集中していくことを見た. そこで, 集中していくとしたら, その集中の幅はどのくらいか, またその行った先はどうなっているのか, に答えるのが中心極限定理である.

定理 2.6.1 (中心極限定理, Central Limit Theorem, CLT) X_i ($i = 1, 2, \dots$) を独立, かつ同分布な確率変数とし, その平均と, 分散の平方根をそれぞれ

$$\mu \equiv E[X_i], \quad \sigma \equiv \sqrt{\text{Var}[X_i]} \quad (2.6.1)$$

とする．このとき，

$$S_N \equiv \sum_{i=1}^N X_i, \quad Z_N \equiv \frac{1}{\sigma \sqrt{N}} \sum_{i=1}^N (X_i - \mu) = \frac{S_N - \langle S_N \rangle}{\sigma \sqrt{N}} \quad (2.6.2)$$

を定義すると，任意の $a < b$ に対して

$$\lim_{N \rightarrow \infty} P[a \leq Z_N \leq b] = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \quad (2.6.3)$$

が成り立つ．

右辺に出てきた分布を「正規分布」(normal distribution) と言う．今までの言葉では，上の定理は Z_N は正規分布に法則収束する とまとめられる．通常，正規分布の累積分布関数を

$$\Phi(x) \equiv \int_{-\infty}^x \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy \quad (2.6.4)$$

と書く．以下に $1 - \Phi(x) = \int_x^{\infty} \frac{e^{-y^2/2}}{\sqrt{2\pi}} dy$ のいくつかの値を載せておく⁴：

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}

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，と言うのが一つの主張．更に， $S_N - N\mu$ を \sqrt{N} で割ることによって Z_N を定義することで， $N \rightarrow \infty$ でも（大抵は）有限にとどまるような量を定義できる，と言うのがもう一つの主張．更に詳しく，定理は，この Z_N が「正規分布」に近づいていくことを主張している．

2.6.1 中心極限定理の（特殊な場合についての）証明

中心極限定理そのものは上に書いたとおりに非常に一般に成り立つ．しかし，その証明にはいくつかの道具が必要である．そこで，まず，非常に特殊な場合に具体的な計算をする事で，この特殊な場合には実際に定理が成立していることを確かめることにも意味がある（それでもかなり大変なので，細かいところの厳密性ははしょってある）．その後で一般的な場合の証明を行うことにする．なお，この証明自身にはそんなに意味はないが，先週の「自主的に行う問題」の解答を兼ねてこのレジュメに載せることにした．

具体的には X_i が ($0 < p < 1$)

$$X_i = \begin{cases} 0 & (\text{確率 } 1-p \text{ で}) \\ 1 & (\text{確率 } p \text{ で}) \end{cases} \quad (2.6.5)$$

⁴この積分は特殊な x の値に対して以外は，陽には計算できない．以下の表は数値計算によるものである

となっている場合のみを考える(「自主的に考える問題」の状況にあわせるには $p = \frac{1}{3}$ とした上で, $Y_i = -2 + 3X_i$ なる Y_i を考えればよい). この場合,

$$\mu = p, \quad \sigma^2 = \text{Var}[X_1] = p(1-p) \quad (2.6.6)$$

となっている.

Step 1. さて, 問題になっている確率を書き直すところから始めよう. 今, 確率変数の数 N を固定し, N 個の X_i のうちで m 個の値が 1 である確率を考えると, これは

$$p_{N,m} \equiv \binom{N}{m} p^m (1-p)^{N-m} \quad (2.6.7)$$

である. またこのとき, Z_N と m の間には

$$Z_N \equiv \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^N (X_i - \mu) = \frac{m - Np}{\sigma\sqrt{N}} \quad (2.6.8)$$

の関係がある. よって, 問題になる確率を m に対する条件を用いて計算すると

$$P[a \leq Z_N \leq b] = P[Np + a\sigma\sqrt{N} \leq m \leq Np + b\sigma\sqrt{N}] = \sum_{Np + a\sigma\sqrt{N} \leq m \leq Np + b\sigma\sqrt{N}} \binom{N}{m} p^m (1-p)^{N-m} \quad (2.6.9)$$

ということになる. この右辺に対して $N \rightarrow \infty$ の極限が計算できればよい. そこで, 右辺に出てきている確率を計算することにする.

Step 2. そのためにはまず, 二項係数を計算する必要があるが, それには Stirling の公式

$$n! \sim \sqrt{2\pi} n^{n+1/2} e^{-n} \quad (n \rightarrow \infty \text{ で}) \quad (2.6.10)$$

を用いる. $n \rightarrow \infty$ で成り立つ式を有限の n で使うには誤差の評価が必要である. しかし, この公式は小さな n でも異常に正確 ($n=2$ での相対誤差は 4.05%, $n=3$ なら 2.73%) なので, あたかも等式であるかのように進めて行く. さて, $N \rightarrow \infty$ では (2.6.9) の和に出てくる m については m も, $N-m$ も, 共に無限大になる. そこで Stirling の公式が使えて,

$$\binom{N}{m} = \frac{N!}{m!(N-m)!} \approx \frac{1}{\sqrt{2\pi}} \left(\frac{N}{m(N-m)} \right)^{1/2} \frac{N^N}{m^m (N-m)^{N-m}} \quad (2.6.11)$$

これを更に変形しないといけないが, $N \rightarrow \infty$ で何が起こるかを見やすくするために, 今興味のある m を $m = Np + \ell\sqrt{N}$ と書くことにする ((2.6.9) の和に出てくる m に対しては, $a\sigma \leq \ell \leq b\sigma$ である). すると ($q = 1-p$ と書く),

$$\binom{N}{m} \approx \frac{1}{\sqrt{2\pi N}} \left(p + \frac{\ell}{\sqrt{N}} \right)^{-1/2 - Np - \sqrt{N}\ell} \left(q - \frac{\ell}{\sqrt{N}} \right)^{-1/2 - Nq + \sqrt{N}\ell} \quad (2.6.12)$$

となるので, (2.6.11) は

$$\begin{aligned}
 \binom{N}{m} p^m q^{N-m} &\approx \frac{1}{\sqrt{2\pi N}} \left(p + \frac{\ell}{\sqrt{N}}\right)^{-1/2-Np-\sqrt{N}\ell} \left(q - \frac{\ell}{\sqrt{N}}\right)^{-1/2-Nq+\sqrt{N}\ell} \times p^{Np+\sqrt{N}\ell} q^{Nq-\sqrt{N}\ell} \\
 &= \frac{1}{\sqrt{2\pi Npq}} \left(1 + \frac{\ell}{\sqrt{N}p}\right)^{-1/2-Np-\sqrt{N}\ell} \left(1 - \frac{\ell}{\sqrt{N}q}\right)^{-1/2-Nq+\sqrt{N}\ell} \\
 &= \frac{1}{\sqrt{2\pi Npq}} \left(1 + \frac{\ell}{\sqrt{N}p}\right)^{-1/2} \left(1 - \frac{\ell}{\sqrt{N}q}\right)^{-1/2} \\
 &\quad \times \left(1 + \frac{\ell}{\sqrt{N}p}\right)^{-Np} \left(1 - \frac{\ell}{\sqrt{N}q}\right)^{-Nq} \times \left(1 + \frac{\ell}{\sqrt{N}p}\right)^{-\sqrt{N}\ell} \left(1 - \frac{\ell}{\sqrt{N}q}\right)^{+\sqrt{N}\ell}
 \end{aligned} \tag{2.6.13}$$

となる.

Step 3. さて, 上の第一行は $N \rightarrow \infty$ で $1/\sqrt{2\pi Npq}$ に行く. また, 最後の2つは

$$\lim_{N \rightarrow \infty} \left(1 + \frac{x}{N}\right)^N = e^x \tag{2.6.14}$$

から,

$$\left(1 + \frac{\ell}{\sqrt{N}p}\right)^{-\sqrt{N}\ell} \left(1 - \frac{\ell}{\sqrt{N}q}\right)^{+\sqrt{N}\ell} \xrightarrow{N \rightarrow \infty} e^{-\ell^2/p} \times e^{-\ell^2/q} = \exp\left(-\frac{\ell^2}{p} - \frac{\ell^2}{q}\right) \tag{2.6.15}$$

となる. 真ん中の二つはちょっと厄介なので, \log をとってみると,

$$\begin{aligned}
 \log \left[\left(1 + \frac{\ell}{\sqrt{N}p}\right)^{-Np} \left(1 - \frac{\ell}{\sqrt{N}q}\right)^{-Nq} \right] &= -Np \log \left(1 + \frac{\ell}{\sqrt{N}p}\right) - Nq \log \left(1 - \frac{\ell}{\sqrt{N}q}\right) \\
 &\approx -Np \left[\frac{\ell}{\sqrt{N}p} - \frac{1}{2} \left(\frac{\ell}{\sqrt{N}p}\right)^2 \right] - Nq \left[-\frac{\ell}{\sqrt{N}q} - \frac{1}{2} \left(\frac{\ell}{\sqrt{N}q}\right)^2 \right] = \frac{\ell^2}{2} \left(\frac{1}{p} + \frac{1}{q}\right)
 \end{aligned} \tag{2.6.16}$$

となるから,

$$\left(1 + \frac{\ell}{\sqrt{N}p}\right)^{-Np} \left(1 - \frac{\ell}{\sqrt{N}q}\right)^{-Nq} \approx \exp\left(\frac{\ell^2}{2p} + \frac{\ell^2}{2q}\right) \tag{2.6.17}$$

が結論できる.

Step 4. 以上から $m = Np + \ell\sqrt{N}$ に対して,

$$\begin{aligned}
 P[Z_N = \ell/\sigma] &= \binom{N}{m} p^m q^{N-m} \approx \frac{1}{\sqrt{2\pi Npq}} \times \exp\left(\frac{\ell^2}{2p} + \frac{\ell^2}{2q}\right) \times \exp\left(-\frac{\ell^2}{p} - \frac{\ell^2}{q}\right) \\
 &= \frac{1}{\sqrt{2\pi N}\sigma} \exp\left[-\frac{\ell^2}{2} \left(\frac{1}{p} + \frac{1}{q}\right)\right]
 \end{aligned} \tag{2.6.18}$$

となることがわかった (最後のところでは $\sigma = \sqrt{pq}$ であることを用いた). $\ell/\sigma \equiv z$ と書くと,

$$P[Z_N = z] \approx \frac{1}{\sqrt{2\pi N}\sigma} \exp\left[-\frac{\sigma^2 z^2}{2} \left(\frac{1}{p} + \frac{1}{q}\right)\right] = \frac{1}{\sqrt{2\pi N}\sigma} \exp\left(-\frac{z^2}{2}\right) \tag{2.6.19}$$

となる. ただし, 上の式が意味を持つには, z が許される値 (つまり, $S_N = Np + \sigma\sqrt{N}z$ が整数になるように) をとることが必要である.

Step 5. これで漸く, (2.6.9) に戻ることができる. 問題の確率は

$$P[a \leq Z_N \leq b] = P[a \leq z \leq b] \approx \sum_z \frac{1}{\sqrt{2\pi N} \sigma} \exp\left(-\frac{z^2}{2}\right) \quad (2.6.20)$$

なのだ. ただし, z の和は, $S_N = Np + \sigma \sqrt{N}z$ が整数になるような z のみをとる. さて, そのような z は, $\frac{1}{\sigma\sqrt{N}}$ 毎に分布している. そこで上のを以下のように書いてみると, 結果は積分のリーマン和の形になる:

$$P[a \leq Z_N \leq b] \approx \frac{1}{\sqrt{2\pi}} \sum_z \frac{1}{\sigma \sqrt{N}} \exp\left(-\frac{z^2}{2}\right) \approx \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{z^2}{2}\right) dz. \quad (2.6.21)$$

と言うわけで定理が「証明」された. □

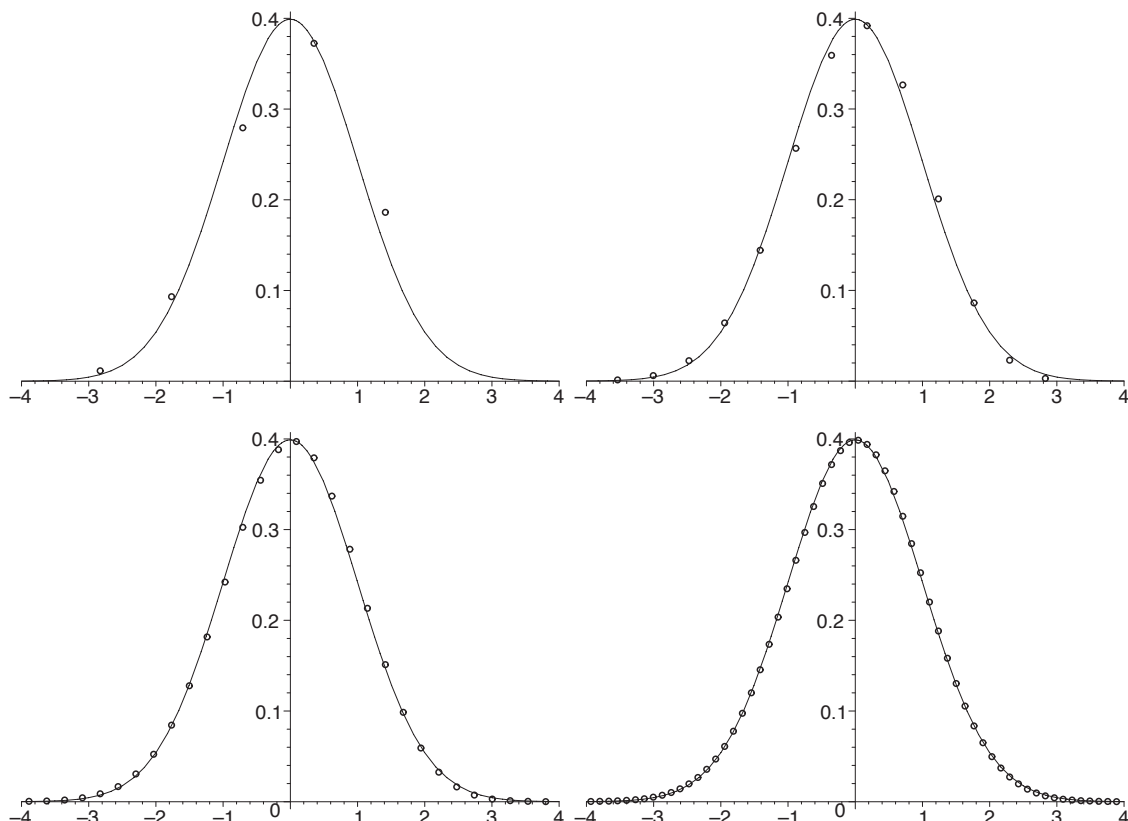
少しおまけ: 上の計算をグラフにしてみたのが以下の図である. 図では正規分布の分布密度関数(実線)と, $p = 1/3$ に相当する二項分布を規格化したもの(いくつかの点)を $N = 4$ (左上), $N = 16$ (右上), $N = 64$ (左下), $N = 256$ (右下) の場合についてそれぞれ描いた. より正確には, 「自主的に計算してみる問題」の場合を考え,

$$S_N \equiv \sum_{j=1}^N X_j, \quad Z_N \equiv \frac{S_N}{\sqrt{2N}} \quad (2.6.22)$$

として定義した Z_N を横軸に, また,

$$P[Z_N = z] \times \frac{\sqrt{2N}}{3} \quad (2.6.23)$$

なる量を縦軸にとったものを描いている. N が増えて行くにつれ, 正規分布に近づいている様子がわかる.



2.7 特性関数と中心極限定理

さて, 一般の場合に中心極限定理を証明するのに上のようなことをやるわけにはいかない. 上では特殊な場合に限定したにもかかわらず, それなりに大変だった. X_i の分布が 2 項分布でなく多項分布だったら, それよりも連続的に分布している指数分布だったら, と考えると, 同様の計算をくり返す気はしないだろう. 幸いなことに「特性関数」というものを使うと, 中心極限定理を非常に一般的に証明できる.

そこで以下では特性関数を導入し, その基本的な性質を述べると共に, 中心極限定理の証明を与えることにしよう.

2.7.1 特性関数

定義 2.7.1 実確率変数 X の特性関数 $\phi(t)$ とは,

$$\phi(t) \equiv E[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} \mu(dx) \quad (t \in \mathbb{R}) \quad (2.7.1)$$

で定義される t の関数のことである. 上の右辺では, X の分布を表す測度を μ として, 期待値をこの測度に関する積分の形で書いた.

以下, 特性関数に関する基本的な性質を列挙する.

命題 2.7.2 X を実確率変数, $\phi(t)$ をその特性関数とすると,

- 全ての $t \in \mathbb{R}$ に対して, $|\phi(t)| \leq 1 = \phi(0)$. また, $\phi(-t) = \bar{\phi}(t)$. ここで $\bar{\phi}$ は ϕ の複素共役.
- ϕ は t について一様連続

証明:

簡単なので略. □

註 2.7.3 ランダウの O と o についての注意. x の関数 $f(x), g(x)$ があって, $x \rightarrow \infty$ の極限に興味があるとき,

- $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0$ の時, $f(x) = o(g(x))$ と書く (小文字の o).
- $x \rightarrow \infty$ において $\frac{f(x)}{g(x)}$ が有界のとき, $f(x) = O(g(x))$ と書く (大文字の O).

同様の定義は, $x \rightarrow 0$ を考えている場合にも用いる.

命題 2.7.4 X を実確率変数, $\phi(t)$ をその特性関数とする. また, k を 1 以上の整数とする.

(1) もし $\phi(t)$ の原点での $2k$ -階微分 $\phi^{(k)}(0)$ が存在するならば, $E[|X^{2k}|]$ は有限である.

同様に, $\phi^{(2k-1)}(0)$ が存在するならば, $E[|X^{2k-2}|]$ は有限である.

(2) 逆に $E[|X^k|]$ が有限であるならば,

$$\phi(t) = \sum_{j=0}^k \frac{E[X^j]}{j!} (it)^j + o(t^k) \quad (2.7.2)$$

と展開できる. 特に, $\phi^{(k)}(0) = i^k E[X^k]$ である.

証明:

本質的にはテイラー展開をしたものをどう積分するか, と言った話であり, 解析の標準的な話なので, 略 (ただし, これができなければこの講義についていけない話ではないので, わからない人も気にしないように.) \square

2.7.2 一意性と反転公式

特性関数が有効なのは, この節で述べる以下の性質のためである. なお, これらの定理の証明はそれなりに大変で, 時間の関係で割愛せざるを得ない. 標準的な確率論の本には絶対載っているものであるから, 興味のある人は各自調べることをお奨めする.

定理 2.7.5 (特性関数と分布の一対一対応) 実確率変数 X と Y の特性関数が一致するならば, X と Y の分布も一致する. つまり, 特性関数と分布の間には一対一対応がある.

まず, この定理は, 特性関数が分布を一意的に決める (逆は定義から明らか) ことを主張しており, 分布を決めるために特性関数の決定から入る動機付けになる. 次の定理はこれをもっと押し進めたものである.

定理 2.7.6 (Lévy の反転公式) 実確率変数 X の累積分布関数を $F(a)$, 特性関数を $\phi(t)$ とする. このとき:

(1) 一般に $a < b$ の時,

$$\frac{1}{2} (P[X = a] + P[X = b]) + P[a < X < b] = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt \quad (2.7.3)$$

がなりたつ. なお, 2点 a, b ($a < b$) において F が連続ならば, 上の左辺は $F(b) - F(a) = P[a < X \leq b]$ に等しい.

(2) もしも $\int_{-\infty}^{\infty} |\phi(t)| dt < \infty$ ならば, $F(a)$ は連続である. 更に $F(a)$ の密度関数 $\rho(a)$ が存在して,

$$F(a) = \int_{-\infty}^a \rho(x) dx \quad (2.7.4)$$

および

$$\rho(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi(t) dt \quad (2.7.5)$$

が成り立つ.

見ての通りではあるが、上の反転公式は特性関数から分布関数を再構成する具体的方法を教えてくれている。このおかげで、「先に特性関数を計算してから分布関数を再現する」ような芸当も可能になるわけだ。実際の応用場面では、分布関数を直接計算するよりも特性関数の方が計算しやすい局面は良くある（中心極限定理の証明もその一つで、これから見ていく）。

（少しおまけ）反転公式には $P[X = a]$ のような、ある一点の値をとる確率が現れていた。このように一点に集中した確率を与えるのが次の公式である（用語の註： $P[X = a] > 0$ の場合、 a は X の分布の atom であると言う）。

定理 2.7.7 (おまけ：“atom” の確率) 実確率変数 X の特性関数を $\phi(t)$ とすると、
 (1) 任意の $a \in \mathbb{R}$ に対して

$$P[X = a] = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T e^{-ita} \phi(t) dt \quad (2.7.6)$$

が成り立つ。

(2) さらに、

$$\sum_{a \in \mathbb{R}} \left(P[X = a] \right)^2 = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |\phi(t)|^2 dt \quad (2.7.7)$$

も成り立つ。ここで左辺の a についての和で寄与するのは atomic な所だけで、これは高々、加算個しかない。

最後に、特性関数と法則収束には以下で見るような、密接な関係がある：

定理 2.7.8 (特性関数と法則収束) 特性関数の収束と確率変数の法則収束はほとんど同値である。より正確には：

(1) 実確率変数 X と実確率変数列 X_n ($n = 1, 2, 3, \dots$) があり、 X の特性関数を $\phi(t)$ 、 X_n の特性関数を $\phi_n(t)$ と書く。もし $X_n \xrightarrow{D} X$ (X_n が X に法則収束) するならば、全ての $t \in \mathbb{R}$ において $\phi_n(t) \rightarrow \phi(t)$ が成立する。

(2) 逆に、 $\phi(t) \equiv \lim_{n \rightarrow \infty} \phi_n(t)$ が全ての $t \in \mathbb{R}$ で存在し、かつ $\phi(t)$ が $t = 0$ で連続だとしよう。このとき、 $\phi(t)$ を特性関数に持つような実確率変数 X が存在し、 $X_n \xrightarrow{D} X$ (X_n が X に法則収束する) が成立する。

2.7.3 中心極限定理の証明

では、特性関数を用いて、中心極限定理を証明しよう。証明の方針は定理 2.7.8(2) による。

Step 1. 記号を簡単にして見通しを良くすることから始める。まず、

$$Y_n \equiv \frac{X_n - \mu}{\sigma} \quad (2.7.8)$$

を定義すると、 Y_1, Y_2, \dots は互いに独立かつ同分布で、

$$\langle Y_n \rangle = 0, \quad \text{Var}[Y_n] = 1 \quad (2.7.9)$$

がなりたつ．また，この Y_n を用いると Z_n は $Z_n \equiv \frac{1}{\sqrt{n}} \sum_{j=1}^n Y_j$ と書けることになる．以下では Y_1 の特性関数を $\phi_Y(t)$ ， Z_n の特性関数を ϕ_n と書くことにする．

Step 2. 計算を始める．

$$\phi_n(t) \equiv E[e^{itZ_n}] = E\left[\exp\left\{\frac{it}{\sqrt{n}} \sum_{j=1}^n Y_j\right\}\right] = E\left[\prod_{j=1}^n \exp\left\{\frac{itY_j}{\sqrt{n}}\right\}\right] \quad (2.7.10)$$

であるが，積の中身は互いに独立な確率変数であるから，積の期待値は期待値の積にわかれ，更に Y_i が同分布だから，結局

$$\phi_n(t) = \prod_{j=1}^n E\left[\exp\left\{\frac{itY_j}{\sqrt{n}}\right\}\right] = \left(E\left[\exp\left\{\frac{it}{\sqrt{n}}Y_1\right\}\right]\right)^n = \left\{\phi_Y\left(\frac{t}{\sqrt{n}}\right)\right\}^n \quad (2.7.11)$$

という形になる（一番最後の等式は ϕ_Y の定義をよく睨むと良い）．こいつの $n \rightarrow \infty$ での極限が知りたいのであるから， n が入っている 2 カ所に注意して進む．

Step 3. さて，ともかく，右辺の $\phi_Y\left(\frac{t}{\sqrt{n}}\right)$ を何とかしないことには始まらない．これについては $n \rightarrow \infty$ に従って引数がゼロに近づくことに注意して，テイラー展開を行う．つまり，定理の仮定（ Y_1 の分散が有限）の下では，

$$\phi_Y(t) = 1 - \frac{t^2}{2} + o(t^2) \quad (2.7.12)$$

が成り立つことを用いる（ここで $o(t^2)$ とは， $t \rightarrow 0$ の時に， t^2 より速くゼロに行く項を表す）．この証明は後回し（後の命題 2.7.4）にすると，これから直ちに，

$$\phi_Y\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o(t^2/n) \quad (2.7.13)$$

が得られる．従って，(2.7.11) から

$$\lim_{n \rightarrow \infty} \phi_n(t) = \lim_{n \rightarrow \infty} \left\{1 - \frac{t^2}{2n} + o(t^2/n)\right\}^n = e^{-t^2/2} \quad (t \in \mathbb{R}) \quad (2.7.14)$$

が得られた（最後の極限の計算は，学部一年生の解析だから良いでしょう）．

Step 4. 以上の結果を定理 2.7.8(2) によって解釈する．(2.7.14) で得られた極限を $\phi(t)$ と書くと，定理 2.7.8(2) の条件は全て満たされた．従って定理 2.7.8(2) から， $\phi(t) \equiv e^{-t^2/2}$ を特性関数に持つ確率変数を Z として， Z_n が Z に法則収束することが言える．

Step 5. ところで， $\phi(t) \equiv e^{-t^2/2}$ を特性関数に持つ確率変数は何かと言うと，正規分布である．実際，定理 2.7.6(2) が使える状況なので， Z の分布密度関数が

$$\rho(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} e^{-t^2/2} dt = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad (2.7.15)$$

と計算できる．これは正に，正規分布の分布密度関数そのものである！

□

6月4日の連絡：6月11日は休講にします。

確率論 I, 確率論概論 I 第4回レポート問題

来週を休講にする関係上、レポート問題は多めに出しました(半分くらいは骨のある問題?)。以下の7, 8, 9はみんなができるはずだからやって下さい(これだけであれば合格点はある)。問10も出来ると思うんだけど... 問11と12は、講義では触れにくかった話題などを補うために、興味とやる気のある人向けに出しました。これらの問題は講義だけでは解けないかも知れないので「ボーナス問題」です。

(大嘘)以下の問題のいくつかでは、以下のようなウソを認めて解答してください。中心極限定理は

$$\lim_{n \rightarrow \infty} P[a \leq Z_n \leq b] = \int_a^b \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \quad (\text{正しい})$$

で、あくまで左辺の確率の $n \rightarrow \infty$ の極限が右辺だと主張している(ここまでは正しい)。しかし、これを極限をとる前でも(大嘘!)左辺の確率が右辺に等しい、つまり

$$P[a \leq Z_n \leq b] = \int_a^b \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \quad (\text{大嘘!})$$

だと思って、解答して欲しい、のである(余談:正しい評価は定理2.7.9にある。)

問7: (中心極限定理を安易に使う問題)今、中心極限定理が現れる典型例として、測定値の誤差を考えてみる。 j 回目の測定での結果(測定値)を X_j ($j = 1, 2, 3, \dots$) とすると、自然な仮定としては

- X_j は互いに独立, 同分布で
- X_j の期待値は μ , 分散は σ^2

となるであろう。 μ が真の値なのだが、いろいろな要因で測定値がばらついて、確率変数 X_j のように見えるわけ。大数の法則により、 $\frac{1}{N} \sum_{j=1}^N X_j$ は真の値 μ に収束するはずだが、実際問題としては N 回測定したときの平均値 $\frac{1}{N} \sum_{j=1}^N X_j$ が真の値からどれくらいずれているのかが大事である。この問に答えるため、以下のように考えよう。

1. 問題にしたいのは、 $\frac{1}{N} \sum_{j=1}^N X_j$ と μ の差(誤差)が a より大きい確率(a は後から決める)、つまり $P\left[\left|\frac{S_N}{N} - \mu\right| > a\right]$ である。この確率が小さければ、誤差は a より小さいと言えよう。上の枠で困った「大嘘」によると、この確率はどのような積分で表されるか?
2. 次に、上で求めた、「 $\frac{1}{N} \sum_{j=1}^N X_j$ と μ の差(誤差)が a より大きい確率」が 5% 以下になるように、 N を選びたい。このような N を σ と a で表せ。ここも、上の「大嘘」を前提にして答えて良い(ヒント: 講義でも注意したが、積分を正確に遂行することはできないから、正規分布の累積分布関数 $\Phi(x)$ の表を使うべし。)

3. 具体的な数値を入れてみよう．長さの測定の場合を考え， $\mu = 10mm$ ， $\sigma = 1mm$ とする．このとき，誤差（上の a ）を $0.1mm$ 以下にするには，大体，何回くらい測定すればよいか？同様に，誤差を $0.01mm$ 以下にするには，大体，何回くらい測定すればよいか？
4. (根性のある人だけでよい)「大嘘」を使わず，定理 2.7.9 を用いたらどうなるだろうか？

問 8： 上の問 7 では「大嘘」を使って中心極限定理を近似的に適用して解答してもらった．これだけでは哀しいので，中心極限定理などは忘れて，いつでも成り立つ正しい評価をしてみよう．つまり，チェビシエフの不等式を使って上の問 7 をやり直してみよう．問 7 のようにズルイことをした場合に比べて，最終的な答え（どのくらいの回数，測定すべきか）はどのように変わってくるだろうか？

問 9： (もう一つ，中心極限定理を安易に使う問題)サイコロを 100 回投げること考え， j 回目に出た目を X_j と書く．このとき， $1 \leq a \leq 6$ について，確率

$$P\left[\prod_{j=1}^{100} X_j < a^{100}\right]$$

の表式を，上の「大嘘」を用いて求めよ．勿論， $X_1, X_2, X_3, \dots, X_{100}$ は互いに独立だと仮定して良い（ヒント：積が出ていてギョツとするかも知れないが，積の \log をとれば和に直すことができるよ．なお「表式」を求めると言っても，適当な積分などで表せば十分．）

問 10： 以下を証明せよ（定義通りやれば出来る）:

同じ確率空間で定義されている確率変数 X_n, X, Y_n, Y について $X_n \xrightarrow{P} X$ かつ $Y_n \xrightarrow{P} Y$ であるとする（勿論， $n \rightarrow \infty$ に関する極限）．このとき，

- $X_n \pm Y_n \xrightarrow{P} X \pm Y$ が成り立つ（複合同順）．
- $X_n Y_n \xrightarrow{P} XY$ も成り立つ．

問 11： (ボーナス問題，その 1) X_1, X_2, X_3, \dots を独立・同分布な確率変数とし，その期待値は 0，分散は $\sigma^2 > 0$ だとする．中心極限定理によれば， $Z_n \equiv \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n X_j$ は正規分布に法則収束するのであるが，確率収束はしない．後半部分（確率収束はしないこと）を証明せよ．

ヒント 1： Z_n と Z_{2n} を考えてみよ．

ヒント 2： 上の問 10．

問 12： (ボーナス問題，その 2) 同じ確率空間で定義された確率変数 X, Y があり， X と Y は同分布でその期待値は 0，分散は 1 とする．更に，確率変数 $\frac{X+Y}{\sqrt{2}}$ の分布も， X や Y の分布と同じであると言う．このとき， $X, Y, \frac{X+Y}{\sqrt{2}}$ は全て正規分布に従うことを証明せよ．

ヒント：これらの確率変数の特性関数を考えよ．特に， $\frac{X+Y}{\sqrt{2}}$ の特性関数は， X, Y の特性関数を用いて，どのように書けるか？

締め切りなど：

締め切りは 2002 年 6 月 18 日（火）の 12:55（曜日と時刻の変更に注意！），

提出場所は僕の部屋 (理 1-508) の前の封筒かポスト

用紙はできうる限り A4 の紙を用いる (B5 などの小さい紙は紛れてなくなるかも)

とします.

(いつも通りの) レポートのお約束:

- 友達と相談しても, 本を調べても, 何をやっても良いから, 自分で理解した範囲を書くこと. その際, 参考文献や議論した友達の名前も明記すること (友達と議論したり, 本を見たからと言って悪い点をつける, などと言うことは絶対にしない. 一番大事なのは自分でわかったところを表現することだから, それまでの過程で何をやっても問題ない.)
- なお, 問題の番外編として, 今までの講義内容・講義形態についての感想, 不満, 文句, このように改善すべしとの意見などもできるだけ書いてください. お願いします.

—————以下, レジュメの続き—————

2.7.4 中心極限定理の収束の速さ

上のレポート問題7では「大嘘」をやったけども, 真面目にこのような誤差の問題に取り組むためには「大嘘」でない理論が必要である. 特に, 有限の n について, $P[a \leq Z_n \leq b]$ が正規分布で与えられる積分とどのくらい異なるのか, がわからないと話にならない. これについて代表的な結果として, 以下が挙げられる.

定理 2.7.9 X_1, X_2, X_3, \dots を独立・同分布な確率変数とし, X_j の期待値を μ , 分散を σ^2 とする. 更に, X_j の3次のモーメントを

$$\beta_3 \equiv E[|X_1 - \mu|^3] \quad (2.7.16)$$

と定義しておく. いつものように, $Z_n \equiv \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n (X_j - \mu)$ を定義し, Z_n の累積分布関数を $F_n(x)$, 正規分布の累積分布関数を $\Phi(x)$ と書くと,

$$\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{C\beta_3}{\sigma^3\sqrt{n}} \quad (2.7.17)$$

が成り立つ. ここで C は定数で, $\frac{1}{\sqrt{2\pi}}$ と 0.8 の間にあることがわかっている.

レポートの7番をちゃんと解くには, 上の定理のようなものを援用する必要が, 本当はあるわけだけども, 簡単のために「大嘘」をやったわけ.

2.8 Large Deviations (少しおまけ)

いままで, 大数の法則と中心極限定理を見てきた. 大数の法則が最も粗く, (独立同分布の X_1, X_2, \dots の場合, $S_N \equiv \sum_{j=1}^N X_j$, $\mu \equiv \langle X_1 \rangle$ とおいて)

$$\frac{S_N - N\mu}{N} \xrightarrow{\text{a.s.}} 0 \quad (2.8.1)$$

を主張し, 次に中心極限定理がこの近づき方を議論した. 特に $S_N - N\mu$ は \sqrt{N} くらいのオーダーである, ことが主張の一つであった.

この節では $S_N - N\mu$ が \sqrt{N} よりも大きなオーダーの場合, 特に $S_N - N\mu$ が N くらいのオーダーの確率を議論する. 中心極限定理から, この確率は N と共にゼロになるのは確かであるが, どのくらいの速さでゼロになるかを問題にしたいのである. これに答えるため, いくつかの量を定義する.

まず, X_1 の キュムラント母関数を

$$\Lambda(t) \equiv \log \langle e^{tX_1} \rangle \quad (2.8.2)$$

として定義する. 次に, Λ の ルジャンドル変換を

$$\Lambda^*(a) \equiv \sup_{t \in \mathbb{R}} [at - \Lambda(t)] \quad (a \in \mathbb{R}) \quad (2.8.3)$$

として定義する. すると,

定理 2.8.1 (Large Deviation) X_1, X_2, \dots を独立・同分布な確率変数とし, それらの期待値を $\mu \equiv \langle X_1 \rangle$ と書く. X_1 のモーメント母関数 $\langle e^{tX_1} \rangle$ が $t = 0$ の近傍で有限であるとしよう. 上のように $\Lambda(t)$ と $\Lambda^*(a)$ を定義し, $a > \mu$ なる a を一つ固定する. すると, $P[X_1 > a] > 0$ である限り,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P[S_N > Na] = -\Lambda^*(a) \quad (2.8.4)$$

が成り立つ. 更にこのとき, $\Lambda^*(a) > 0$ である.

この定理によれば,

$$P[S_N > Na] \approx e^{-N\Lambda^*(a)} \quad (N \rightarrow \infty)$$

と言う感じで, この確率は $\Lambda^*(a) > 0$ の場合, N と共に指数関数的に小さくなっていくことがわかる. なお, $a < \mu$ については $-X_j$ を上の X_j と思って定理を使えばよい.

例: 典型的な例で, X_j が ± 1 の値を確率 $\frac{1}{2}$ ずつでとる場合を考える. このとき, $\mu = 0$ であり,

$$\Lambda(t) = \log \langle e^{tX_1} \rangle = \log(\cosh t), \quad \Lambda^*(a) = \frac{1}{2} \left[(1+a) \log(1+a) + (1-a) \log(1-a) \right] \quad (2.8.5)$$

となる.

これで, 大数の法則から延々と続いてきた話が一段落した.

3 ランダムウォーク

この節では物理などでも重要な、ランダムウォークについて考えていく。

3.1 背景説明

ランダムウォークは色々なところに顔を出す。

- 「酔っぱらいのおっさん」の問題。酔っぱらってフラフラ歩いていたら、そのうちに自分の家にたどり着けるか？
- ブラウン運動（例えば）タバコの煙が空気の分子にぶつかられてフラフラ動く運動。
- 気体中の拡散の問題。部屋の隅に置いた香水の香りが部屋の中程まで伝わるのにはどのくらいの時間がかかる？
- 株価の動き。

これらの運動に共通するのは、粒子（や人や株の値）が、周りから色々な力を受け、あっちこっちへ移動することである。その結果として、これらの運動には共通の性質がみられるが、これはまた、空間の性質に敏感に依存する。

この節ではこのような性質を調べることを目的とする。

3.2 1次元のランダムウォーク

3.2.1 定義

まず、一次元で粒子が動く場合を考えよう。株価の変動などが例になる。

粒子の動く場所は一次元の x -軸の上で、粒子は座標が整数の点のみを移動するものとする。粒子は原点 ($x = 0$) から出発し、確率的に左右に移動していく。その移動のルールは、今までの履歴には全く関係なく¹ ($0 \leq p \leq 1$)

確率 p で左へ一歩、確率 $(1 - p)$ で右へ一歩

動くものとする。問題は n 歩経ったときにどこにいるか、とか、 n 歩までにどのような点を通ってきているか、などである。以下、 $q \equiv 1 - p$ と書く。

3.2.2 大体の振るまい

今までの問題と関係づけるために、以下のように定式化しよう。まず、確率変数 X_i ($i = 1, 2, \dots$) を

$$X_i = \begin{cases} -1 & (\text{確率 } p \text{ で}) \\ 1 & (\text{確率 } 1 - p \text{ で}) \end{cases} \quad (3.2.1)$$

となるように定義し、 $\{X_i\}$ は互いに独立であるとする。そして

$$S_n = \sum_{i=1}^n X_i \quad (3.2.2)$$

¹余談であるが、「今までの履歴には全く関係なく」動く性質をマルコフ性と言う

と定義すると, この S_n が n 歩たったときにどこにいるかを表す数になる. あれ, この S_n って大数の法則なんかででてきた奴じゃん. と言うわけで, これまでやって来た極限定理と関係してくるわけ (ただし, それよりももっと詳しい情報を見ることになる — 以下を参照). その前にすぐに計算できることとして:

$$\langle S_n \rangle = (q - p)n, \quad \text{Var}[S_n] = 4pqn \quad (3.2.3)$$

に注意しておこう. 2 番目の式 (特に $p = q = \frac{1}{2}$ の時) は Einstein's relation と呼ばれている.

さて, n 歩経ったときに x にいる確率を $P_n(x)$ と書いて, この $P_n(x)$ を求めることにしよう. これは簡単に計算できる (今まで何回かやった 2 項分布). 実際, n 歩で x にいると言うことは, これまでに

$$\text{左へ } (n - x)/2 \text{ 歩, 右へ } (n + x)/2 \text{ 歩}$$

動いていると言うことだ. 右へ動く確率が $(1 - p)$, 左へ動く確率が p だから, このように動く確率は

$$P_n(x) = \binom{n}{\frac{n+x}{2}} p^{(n-x)/2} (1-p)^{(n+x)/2} \quad (3.2.4)$$

と計算できる. 上で $P_n(x)$ がゼロにならない x の値は (1) $x - n$ が偶数で, かつ (2) $|x| \leq n$ であることは容易にわかる.

この $P_n(x)$ は n が大きくなると正規分布に近づく (中心極限定理). 実は今考えている S_n は中心極限定理の時に出てきたものと同じだから, それを思い出すと, 確率変数

$$Z_n \equiv \frac{1}{\sqrt{4pqn}} [S_n - n(q - p)] \quad (3.2.5)$$

が

$$\lim_{n \rightarrow \infty} P[a \leq Z_n \leq b] = \int_a^b \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \quad (3.2.6)$$

をみtasることになる. ここで $q \equiv 1 - p$ であり, また $(q - p)$ と $4pq$ はそれぞれ X_i の平均と分散である.

これからすぐにわかること:

- n 歩後の位置は中心が $n(q - p)$, 拡がりが大體 $\sqrt{4pqn}$ になっている.
- $p \neq \frac{1}{2}$ では, 中心は右か左にどんどん移動していき, その位置は n に比例している. つまり, 中心の移動速度は一定 ($q - p$) である.
- 一方, 位置の拡がりは大體 \sqrt{n} のオーダーである. だから, $p \neq \frac{1}{2}$ ではこの拡がり位置が移動した距離 (オーダー n) に比べて非常に小さい. つまり, $n \gg 1$ では粒子は $n(q - p)$ の周りに集中しているように見える.
- $p = \frac{1}{2}$ では話は別で, このときだけ, 中心が動かない. 位置の拡がり \sqrt{n} で, 今までと本質的な差はないが, 何分, 中心が動かないので, この「位置の拡がり」が主役を演じる.

用語: $p \neq \frac{1}{2}$ の場合は左か右に n に比例してそれていくので, この振る舞いを ballistic な振るまい, と呼ぶ. 一方, $p = \frac{1}{2}$ の振る舞いを diffusive (拡散的) な振るまい, と呼ぶ.

6月18日の連絡: 6月4日出題のレポートは来週, 返却の予定です. また, 来週には, 新しいレポート問題を出す予定です.

3.2.3 再帰性

こんな訳で, $p = \frac{1}{2}$ が一番面白いから, 以下, $p = \frac{1}{2}$ に話を限る

さて, ランダムウォークの「再帰性」について考えよう. ランダムウォークはフラフラ動いている訳なので, 色々なところへ行く. あるサンプルをとれば出発点に何回も戻ってくるだろうし, 別のサンプルでは戻ってこないだろう. それで

無限の(十分大きい)時間待ってやったら, どんなサンプルでも出発点に戻ってくるか. もっと正確には「出発点に戻ってくる確率が1かどうか」

と言う問題を考えたい. 上の答が YES, つまり無限時間待ったら原点に確率1で戻ってくる時, ランダムウォークは「再帰的」(recurrent)と言う. 一方, 無限時間待っても戻ってこない確率がゼロでない時, ランダムウォークは「推移的」(transient)と言う.

(注意) 上で $P_n(x)$ を求めたから, $P_n(0)$ もわかっている. けども, この $P_n(0)$ だけでは上の問には答えられない. と言うのは, $P_n(0)$ は「 n 歩後にたまたま原点に戻っている確率」であって, これまでに何回でも原点に戻っているものも数えてしまっているから.

$P_n(x)$ とその仲間を以下のように定義する:

- $P_n(x)$ とは「原点から出発したランダムウォークが時刻 n に x にいる確率」.
- $F_n(x)$ とは「原点から出発したランダムウォークが時刻 n で初めて x に到達する確率」. ただし, $x = 0$ の場合は $F_n(0)$ を「ランダムウォークが時刻 n で初めて原点に戻ってくる確率」とする ($n \geq 2$). また, $F_0(0) = 0$ と決めておく.
- $G(x) \equiv \sum_{n=0}^{\infty} P_n(x)$
- $r(x) \equiv \sum_{n=0}^{\infty} F_n(x)$. 言葉では「ランダムウォークがいつかは x に到達する確率」.

我々の考えたい確率は $r(x)$ であり, これは $F_n(x)$ の和で書けている. そこで $F_n(x)$ を求めたいのだが, これは良くわからない. 一方, $P_n(x)$ については既に求めてある — (3.2.4) 参照. そこで F_n と P_n の関係をつけよう.

このためには以下のように考える. $P_n(x)$ にはいろんなランダムウォークが寄与している. あるものは n 歩目で初めて x に来た. あるものは x に来るのは n 歩目で2回目だ. あるものは3回目だ... 式では

$$P_n(x) = \sum_{\ell=1}^{\infty} P[\text{時刻 } n \text{ には } x \text{ にいるが, これは } x \text{ へは } \ell \text{ 回目の訪問である}] \quad (3.2.7)$$

と書いて良からう.

この内, $\ell = 1$ のは「 x に来るのはこの n 歩目が初めてだ」と言うことだから, これは定義から $F_n(x)$ そのものである. 次に, x に 2 回以上来るもの ($\ell \geq 2$) をまとめて考える. この場合, n 歩目の前に少なくとも一回は x に来ている訳なので, そのときの時刻を k と書くと,

$$\begin{aligned} & \sum_{\ell=2}^{\infty} P[\text{時刻 } n \text{ には } x \text{ にいるが, これは } x \text{ へは } \ell \text{ 回目の訪問である}] \\ &= P[\text{時刻 } n \text{ には } x \text{ にいるが, 以前にも } x \text{ に来たことがある}] \\ &= \sum_{0 < k < n} P[\text{時刻 } n \text{ には } x \text{ にいるが, 時刻 } k \text{ に初めて } x \text{ に来た}] \end{aligned} \quad (3.2.8)$$

と書けている. ところが, 「時刻 n には x にいるが, 時刻 k に初めて x に来た」と言うランダムウォークを時刻 k で 2 つに分けて書くと, 時刻 k までの部分は $F_k(x)$ への寄与であるし, k 以降の部分は $P_{n-k}(0)$ への寄与と考えられる (ここで k 以降の部分は x から出発して $n-k$ 歩で x へ戻る確率なので, 一見 $P_{n-k}(0)$ とは異なる — $P_{n-k}(0)$ は 0 から出発して 0 へ戻る. しかし, 今のランダムウォークは平行移動不変性があるって, 出発点と終点が同じならどこから出発しても同じだから問題の確率は $P_{n-k}(0)$ に等しい.) 以上から上の確率は

$$= \sum_{0 < k < n} F_k(x) P_{n-k}(0) \quad (3.2.9)$$

となる.

以上から (3.2.7) は

$$P_n(x) = F_n(x) + \sum_{k:0 < k < n} F_k(x) P_{n-k}(0) = \sum_{k:0 < k \leq n} F_k(x) P_{n-k}(0) \quad (3.2.10)$$

と書くことができる (第 2 の等号は, $P_0(0) = 1$ である事から出る). この両辺を $n \geq 1$ について和をとると

$$\sum_{n=1}^{\infty} P_n(x) = \sum_{n=1}^{\infty} \sum_{k:0 < k \leq n} F_k(x) P_{n-k}(0) = \sum_{k:k>0} \sum_{n=k}^{\infty} F_k(x) P_{n-k}(0) = \left(\sum_{k:k>0} F_k(x) \right) \left(\sum_{\ell=0}^{\infty} P_{\ell}(0) \right) \quad (3.2.11)$$

従って

$$r(x) = \sum_{k:k>0} F_k(x) = \frac{\sum_{n=1}^{\infty} P_n(x)}{\sum_{\ell=0}^{\infty} P_{\ell}(0)} = \frac{\sum_{n=0}^{\infty} P_n(x) - \delta_{0,x}}{\sum_{\ell=0}^{\infty} P_{\ell}(0)} \quad (3.2.12)$$

と書ける. 最後のところは分子の和を $n = 0$ からにする代わりに $n = 0$ の寄与 1 をひいておいた (ここで $\delta_{0,x}$ とは, $x = 0$ の時のみ 1, 他は 0 の値をとるもの.)

以上で長い準備が終わった. $r(x) = 1$ か否かを見るには, (3.2.12) の分子を計算すればよい. まず $P_n(0)$ は n が偶数の時のみゼロではなくて, Stirling の公式を用いると,

$$P_{2m}(0) = \binom{2m}{m} \left(\frac{1}{2} \right)^{2m} = \frac{(2m)!}{m! m!} \frac{1}{4^m} \approx \frac{1}{\sqrt{\pi m}} \quad (3.2.13)$$

である事がわかり,

$$\sum_{n=0}^{2N} P_n(0) \approx 1 + \sum_{m=1}^N \frac{1}{\sqrt{\pi m}} \approx 1 + \frac{2}{\sqrt{\pi}} \sqrt{N} \quad (3.2.14)$$

となって, $N \rightarrow \infty$ では発散することがわかる. これと (3.2.12) から直ちに,

$$r(0) = 1 - \frac{1}{\sum_{\ell=0}^{\infty} P_{\ell}(0)} = 1 - \frac{1}{\infty} = 1 \quad (3.2.15)$$

が得られる. すなわち, 1次元ランダムウォークは再帰的である.

3.2.4 特別付録: $x \neq 0$ についての $r(x)$

以下の議論は, 相当に細かいし, 少し技術的なので, あくまで「おまけ」である. 興味のある人は読んでください (この節の内容は講義では触れない予定).

上では $r(0)$ (いつかは原点に戻ってくる確率) を計算したが, $r(x)$ (ゼロでない点 x に戻ってくる確率) はまだだった. そこで, 講義では触れなかったが, $x \neq 0$ に対して $r(x)$ も求めておこう.

(3.2.12) を用いたいだが, 分母子共に発散するので, 注意が必要である. そこで (3.2.10) から (3.2.11) 辺りまで戻る. N を大きな数として, $1 \leq n \leq N$ なる n について (3.2.10) の両辺の和をとってみると,

$$\sum_{n=1}^N P_n(x) = \sum_{k=1}^N F_k(x) \sum_{\ell=0}^{N-k} P_{\ell}(0) \quad (3.2.16)$$

となる. このままでは右辺の和が別れないので, 強引に以下のように変形する. まず, 右辺の k の和を $1 \leq k \leq N/2$ に制限すると和の項が少なくなるので右辺の値は下がる:

$$\sum_{n=1}^N P_n(x) \geq \sum_{k=1}^{N/2} F_k(x) \sum_{\ell=0}^{N-k} P_{\ell}(0) \quad (3.2.17)$$

更に, この範囲の k に対しては $N - k \geq N/2$ なので, ℓ の和も $1\ell \leq N/2$ に制限すると, 更に値は下がる:

$$\geq \sum_{k=1}^{N/2} F_k(x) \sum_{\ell=0}^{N/2} P_{\ell}(0) = \left(\sum_{k=1}^{N/2} F_k(x) \right) \left(\sum_{\ell=0}^{N/2} P_{\ell}(0) \right) \quad (3.2.18)$$

従って,

$$\sum_{k=1}^{N/2} F_k(x) \leq \frac{\sum_{n=1}^N P_n(x)}{\sum_{\ell=0}^{N/2} P_{\ell}(0)} \quad (3.2.19)$$

が得られた.

逆側の不等式を作るには, まず, (3.2.16) の ℓ の和を $0 \leq \ell \leq N$ まで広げる. すると, 和の数が多くなるから, 右辺の方が大きくなる:

$$\sum_{n=1}^N P_n(x) \leq \sum_{k=1}^N F_k(x) \sum_{\ell=0}^N P_{\ell}(0) = \left(\sum_{k=1}^N F_k(x) \right) \left(\sum_{\ell=0}^N P_{\ell}(0) \right) \quad (3.2.20)$$

すなわち,

$$\sum_{k=1}^N F_k(x) \geq \frac{\sum_{n=1}^N P_n(x)}{\sum_{\ell=0}^N P_\ell(0)} \quad (3.2.21)$$

も, 得られた.

この節の中間報告: $r(x)$ については

$$\frac{\sum_{n=1}^N P_n(x)}{\sum_{\ell=0}^N P_\ell(0)} \leq \sum_{k=1}^N F_k(x) \leq \frac{\sum_{n=1}^{2N} P_n(x)}{\sum_{\ell=0}^N P_\ell(0)} \quad (3.2.22)$$

が成り立つ.

後は (3.2.22) のそれぞれの分母子を計算すればよい. $\sum_{\ell=0}^N P_\ell(0)$ については (3.2.14) で計算してある. $P_n(x)$ については, まず $x = 2y$ が偶数の時から考える (ついでに $x > 0$ としておこう). この場合, n も偶数でないと $P_n(x) = 0$ になってしまうので,

$$\begin{aligned} \sum_{n=1}^{2N} P_n(2y) &= \sum_{m=y}^N P_{2m}(2y) = \sum_{m=y}^N \frac{1}{4^m} \binom{2m}{m-y} = \sum_{m=y}^N \frac{1}{4^m} \frac{(2m)!}{(m-y)!(m+y)!} \\ &= \sum_{m \approx y}^N \frac{1}{\sqrt{\pi m}} \left(1 - \frac{y}{m}\right)^{-(m-y+1/2)} \left(1 + \frac{y}{m}\right)^{-(m+y+1/2)} \\ &= \sum_{m \approx y}^N \frac{1}{\sqrt{\pi m}} \left(1 - \frac{y^2}{m^2}\right)^{-(m+1/2)} \left(1 - \frac{y}{m}\right)^y \left(1 + \frac{y}{m}\right)^{-y} \end{aligned} \quad (3.2.23)$$

和は $m = y$ からとるべきだが, 実際に (3.2.22) で重要になるのはこの和の $N \rightarrow \infty$ での発散の仕方である. これは m が大きいところだけで決まるから, 上で y/m が小さいものとして展開すると,

$$\approx \sum_{m \approx y}^N \frac{1}{\sqrt{\pi m}} \left(1 - \frac{y^2}{m} + \dots\right) \approx \frac{2}{\sqrt{\pi}} \sqrt{N} - (\text{定数}) \quad (3.2.24)$$

となる (最後のは $N \rightarrow \infty$ で正しい). これを (3.2.22) に入れると, どちらの分母子も全く同じ速さ $\frac{2}{\sqrt{\pi}} \sqrt{N}$ で発散することがわかる. すなわち,

$$r(x) = \sum_{k>0} F_k(x) = 1 \quad (3.2.25)$$

となり, この場合も (長い間待っておれば) そのうちに x を訪問することがわかる.

x が奇数の時も本質的に同じ計算なので省略する. 結果は同じで (3.2.25) が成り立ち, やはりいずれは x を訪問することがわかる.

$x = 0$ の時も本当は上のような議論をして, $N \rightarrow \infty$ の極限を考えるべきであるが, 話をややこしくしないために講義では触れなかった.

以上の結論：一次元ランダムウォークは十分に長い時間経てば必ず（確率 1 で），出発点に戻ってくる．また，十分に長い時間経てば必ず（確率 1 で），どの点（ $x \in \mathbb{Z}$ ）にでも到達する．

3.3 2次元のランダムウォーク

3.3.1 定義

では，2次元のランダムウォークを考えよう．今回は2次元平面で，座標の値が整数値である点を考える．これを \mathbb{Z}^2 と書くことにする：

$$\mathbb{Z}^2 = \{(x, y) \mid x, y \in \mathbb{Z}\} \quad (3.3.1)$$

$+x$ 方向を東， $-x$ 方向を西， $+y$ 方向を北， $-y$ 方向を南と呼ぶ．

さて，この \mathbb{Z}^2 上で粒子が動くことを考える．その規則は

- 粒子はやはり原点 $(0, 0)$ から出発する．
- 粒子は整数時刻毎に今いるところから隣の点へうごく．
- 隣の跳び方は，東西南北のいずれにでも，それぞれ確率 $\frac{1}{4}$ で跳ぶ．
- 跳び方は粒子の過去の履歴には全くよらない（今までいたところを避けようとしたり，逆に今までいたところに移行したり，はない）．

と言うものである．1次元ランダムウォークの $p = 1/2$ の場合には左右へ同確率で跳んだが，この2次元では，東西南北へ同確率（ $1/4$ ）で跳ぶのである．この意味で1次元ランダムウォークの自然な拡張になっている．

1次元の時と同じく，少し抽象的な定式化をしておこう．2次元ベクトルである確率変数（確率ベクトル） \vec{X}_i で以下のように値をとるものを導入する．

$$\vec{X}_i = \begin{cases} (1, 0) & (\text{確率 } 1/4 \text{ で}) \\ (-1, 0) & (\text{確率 } 1/4 \text{ で}) \\ (0, 1) & (\text{確率 } 1/4 \text{ で}) \\ (0, -1) & (\text{確率 } 1/4 \text{ で}) \end{cases} \quad (3.3.2)$$

すると，今度は

$$\vec{S}_n \equiv \sum_{i=1}^n \vec{X}_i \quad (3.3.3)$$

が時刻 n での粒子の位置を表す．

3.3.2 大体の振る舞い

少し今までやった「中心極限定理」の範囲を超えるが，ともかく以下のように考えてみる．時刻 n でどのくらい原点から離れているか，見てみたい．そのためにまず， \vec{S}_n の期待値と「分散」を考える．期待値は

$$\langle \vec{S}_n \rangle = \left\langle \sum_{i=1}^n \vec{X}_i \right\rangle = \sum_{i=1}^n \langle \vec{X}_i \rangle = \sum_{i=1}^n \vec{0} = \vec{0} \quad (3.3.4)$$

上のはベクトルの各成分毎に期待値をとっていることに注意 (講義で説明) .

「分散」に相当するものとしては, 原点からの距離の二乗 $|\vec{S}_n|^2$ の期待値を考えるのが自然である. これは期待値の線形性から

$$\langle |\vec{S}_n|^2 \rangle = \langle \vec{S}_n \cdot \vec{S}_n \rangle = \left\langle \left(\sum_{i=1}^n \vec{X}_i \right) \cdot \left(\sum_{j=1}^n \vec{X}_j \right) \right\rangle = \left\langle \sum_{i=1}^n \sum_{j=1}^n \vec{X}_i \cdot \vec{X}_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n \langle \vec{X}_i \cdot \vec{X}_j \rangle \quad (3.3.5)$$

となる. ところが,

$$\langle \vec{X}_i \cdot \vec{X}_j \rangle = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \quad (3.3.6)$$

となることがわかる (講義で説明). 従って,

$$\langle |\vec{S}_n|^2 \rangle = n \quad (3.3.7)$$

となる.

以上の解釈: 粒子の分布の中心は原点 $(0, 0)$ である. その周りの粒子の拡がりは大体 $\sqrt{\langle |\vec{S}_n|^2 \rangle} = \sqrt{n}$ ぐらいである.

以上を見る限り, 1次元のランダムウォーク (の $p = 1/2$ の場合) と定性的に同じ振る舞いをしているように見える.

3.3.3 $P_n(x, y)$ の計算

更に, 時刻 n で (x, y) にいる確率を $P_n(x, y)$ と書いて, これを求めよう. 今度は1次元に比べてちと厄介である. 上下左右に n 歩動いた結果として (x, y) にいると言うことは, それぞれの方向にどれくらい動いたことになるだろうか?

東に n_e 歩, 西に n_w 歩, 北に n_n 歩, 南に n_s 歩

だけ動いているとすると (e, w, n, s はそれぞれ, east, west, north, south の略), n 歩目での位置は,

$$x = n_e - n_w, \quad y = n_n - n_s \quad (3.3.8)$$

となっているはずである. また, 全体で n 歩なのだから,

$$n_e + n_w + n_n + n_s = n \quad (3.3.9)$$

の関係がある. この3条件をみたく n_e, n_w, n_n, n_s なら何でも良い.

さて, このような n_e, n_w, n_n, n_s 歩で n 歩を行くやり方は何通りあるかと言うと, 全体で n の時刻 (ステップ) の中から n_e 個の「東へ行くステップ」, n_w 個の「西へ行くステップ」, 等々をとるやり方だけある. これは

$$\binom{n}{n_e} \times \binom{n - n_e}{n_w} \times \binom{n - n_e - n_w}{n_n} \times \binom{n - n_e - n_w - n_n}{n_s} = \binom{n}{n_e \ n_w \ n_n \ n_s} \quad (3.3.10)$$

通りある (最後の量は「多項係数」). それぞれの行き方の確率はいつも $(\frac{1}{4})^n$ であるから, 最終的に,

$$P_n(x, y) = \sum'_{n_e, n_w, n_n, n_s} \frac{1}{4^n} \binom{n}{n_e \ n_w \ n_n \ n_s} \quad (3.3.11)$$

となる。ただし、上の和は (3.3.8) と (3.3.9) の条件を満たすような n_e, n_w, n_n, n_s についてとる。(このように条件が付いていることを明示するため、和の記号にプライムを付けた。)

一つ、具体例をやってみよう。時刻 n で原点にいる確率 $P_n(0, 0)$ を考える。 $P_n(0, 0)$ がゼロにならないためには n は偶数である必要がある。更に、(3.3.8), (3.3.9) の条件は ($n = 2m$ と書く)

$$n_e = n_w, \quad n_n = n_s, \quad n_e + n_n = \frac{n}{2} = m \quad (3.3.12)$$

ということになる。よって、(3.3.11) より

$$P_{2m}(0, 0) = \sum_{n_e+n_n=m} \frac{1}{4^{2m}} \binom{n}{n_e \ n_e \ n_n \ n_n} = \frac{1}{4^{2m}} \sum_{\ell=0}^m \frac{(2m)!}{(\ell!)^2 ((m-\ell)!)^2} \quad (3.3.13)$$

となっている(最後のところでは n_e を ℓ と書き、多項係数を具体的に書き下した)。

3.3.4 再帰性と推移性

2次元のランダムウォークが再帰的かどうか、1次元の時と同様に判定できる。1次元の時に倣って、

- $P_n(x, y)$ とは「原点から出発したランダムウォークが時刻 n に (x, y) にいる確率」。
- $F_n(x, y)$ とは「原点から出発したランダムウォークが時刻 n で初めて (x, y) に到達する確率」。

ただし、 $x = y = 0$ の場合は $F_n(0, 0)$ を「ランダムウォークが時刻 n で初めて原点に戻ってくる確率」とする ($n \geq 2$)。また、 $F_0(0, 0) = 0$ と決めておく。

- $G(x, y) \equiv \sum_{n=0}^{\infty} P_n(x, y)$
- $r(x, y) \equiv \sum_{n=0}^{\infty} F_n(x, y)$ 。言葉では「ランダムウォークがいつかは (x, y) に到達する確率」。

を導入すると、(3.2.22) がそのまま成り立つ：

$$\frac{\sum_{n=1}^N P_n(\vec{x})}{\sum_{\ell=0}^{\infty} P_{\ell}(\vec{0})} \leq \sum_{k=1}^N F_k(\vec{x}) \leq \frac{\sum_{n=1}^{2N} P_n(\vec{x})}{\sum_{\ell=0}^{\infty} P_{\ell}(\vec{0})} \quad (3.3.14)$$

従って、ここに出てくる分母を計算すればよい。一般の \vec{x} は大変なので、 $\vec{x} = \vec{0}$ (原点に戻ってくるかどうか) を考えよう。

ちょっとずるいけども、以下の公式：

$$\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n} \quad (3.3.15)$$

を使うのが一番簡単である。この公式自身は、

$$(1+t)^n \times (1+t)^n = (1+t)^{2n} \quad (3.3.16)$$

の両辺を展開して t^n の係数を比較すると証明できる．ともかく (3.3.13) に (3.3.15) を用いると，

$$P_{2m}(0,0) = \frac{1}{4^{2m}} \sum_{\ell=0}^m \frac{(2m)!}{(m!)^2} \times \frac{(m!)^2}{(\ell!)^2 ((m-\ell)!)^2} = \frac{1}{4^{2m}} \binom{2m}{m} \sum_{\ell=0}^m \binom{m}{\ell}^2 = \frac{1}{4^{2m}} \binom{2m}{m}^2 = \left[\frac{1}{4^m} \binom{2m}{m} \right]^2 \quad (3.3.17)$$

となるので，更に Stirling の公式を用いて — 右辺に出ている量は (3.2.13) の 2 乗であることに注意

$$\approx \frac{1}{\pi m} \quad (3.3.18)$$

となる．そこで

$$\sum_{m=1}^N P_{2m}(0,0) \approx \frac{1}{\pi} \sum_{m=1}^N \frac{1}{m} \approx \frac{1}{\pi} \log N \quad (3.3.19)$$

が得られるので， $N \gg 1$ では (3.3.14) から

$$\sum_{k=1}^N F_{2k}(0,0) \approx 1 - \frac{1}{\sum_{m=1}^N P_{2m}(0,0)} \approx 1 - \frac{\pi}{\log N} \quad (3.3.20)$$

ぐらいであることがわかる．(3.3.14) には N だけでなく $2N$ も入っているが， $\log(2N) = \log N + \log 2$ であることなどを考えると，大まかな振る舞いは (3.3.20) で良い．

と言うわけで，2次元でも

$$r(0,0) = \sum_{k=1}^{\infty} F_{2k}(0,0) = 1 \quad (3.3.21)$$

であり，ランダムウォークは再帰的であることがわかった．

(次回の予告，のつもり) まあ，このように計算すれば良いのだが，これでは次元が増えていったら大変だと言うことはすぐにわかるだろう．また，隣の点に跳ぶだけでなく，斜めの点や隣の隣の点へも跳ぶようなランダムウォークを考えて行くと，この節のような計算をくり返して行くと死んでしまうこともわかるだろう．

ところが，再帰性については，このように一般化したランダムウォークについても，非常に一般的な性質があることがわかっている．と言うわけで，来週は，この一般的な性質を導くためにはどうすればよいか，を考えて行く．結論から言えば，フーリエ変換を用いるのである．

6月25日の連絡：特になし．宣言通り，期末テストは7月23日だよ．

確率論 I, 確率論概論 I 第 5 回レポート問題

今回は厳密にやるにはちと大変かもしれないので，ボーナス問題気味．

問 13：（1次元や2次元で，再帰的でないランダムウォークを作る問題）1次元や2次元のランダムウォークは，かなり緩い条件の下で再帰的であることは，今日，講義する予定である．そこで，1次元で再帰的にならないようなランダムウォークの例を作ってみよう．（1次元より2次元の方がやりやすい側面もある．2次元が良い人は2次元をやっても良い．）

1. 具体的には，ランダムウォークの遷移確率を $p_{x,y} = f(y-x)$ と書いたとき， f を適当に与えて，再帰的でないものを作るとよい．例えば， $f(x) = C|x|^{-\alpha}$ （ C は規格化定数， α はこれから選ぶ定数）の形で f を探してみるとどうなるか — どのように α をとるべきかなどと考えてみると良い．ただし，ランダムウォークが対称になるように， $f(x) = f(-x)$ なる f の範囲で探すこと（一般に対称でなければ再帰的でないので，つまらない．）
2. 上で選んだ $f(x)$ が，再帰的でないランダムウォークを与えることを証明せよ（厳密にやるのは少し難しいかも）．

締め切りなど：

締め切りは 2002 年 7 月 1 日（月）の 15:00（曜日と時刻の変更に注意！），

提出場所は僕の部屋（理 1-508）の前の封筒かポスト

用紙はできうる限り A4 の紙を用いる（B5 などの小さい紙は紛れてなくなるかも）

とします．

（いつも通りの）レポートのお約束：

- 友達と相談しても，本を調べても，何をやっても良いから，自分で理解した範囲を書くこと．その際，参考文献や議論した友達の名前も明記すること（友達と議論したり，本を見たからと言って悪い点をつける，などと言うことは絶対にしない．一番大事なのは自分でわかったところを表現することだから，それまでの過程で何をやっても問題ない．）
- なお，問題の番外編として，今までの講義内容・講義形態についての感想，不満，文句，このように改善すべしとの意見などもできるだけ書いてください．お願いします．

3.4 6月4日出題のレポート略解

（まずは先週が切だったレポートの解答（略解）から．ここで新しい節というのも変だけど， \LaTeX の都合上，こうやらせてもらいまふ．）

問 7： 中心極限定理の簡単な応用問題（ただし「大嘘」つき）である．

1. 定義に従って，中心極限定理の使いそうな形に持っていく．

$$\begin{aligned}
 P\left[\left|\frac{S_N}{N} - \mu\right| > a\right] &= P\left[\left|\frac{S_N - N\mu}{\sigma\sqrt{N}}\right| > \frac{a\sqrt{N}}{\sigma}\right] = P\left[|Z_N| > \frac{a\sqrt{N}}{\sigma}\right] \\
 &= P\left[Z_N > \frac{a\sqrt{N}}{\sigma}\right] + P\left[Z_N < -\frac{a\sqrt{N}}{\sigma}\right] \stackrel{\text{大嘘}}{=} \int_{a\sqrt{N}/\sigma}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz + \int_{-\infty}^{-a\sqrt{N}/\sigma} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \\
 &= 2 \int_{a\sqrt{N}/\sigma}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = 2\left[1 - \Phi\left(\frac{a\sqrt{N}}{\sigma}\right)\right] \tag{3.4.1}
 \end{aligned}$$

である.

2. 上の積分が $0.05 = 1/20$ より小さければ良い. つまり,

$$1 - \Phi\left(\frac{a\sqrt{N}}{\sigma}\right) < \frac{1}{40} \quad (3.4.2)$$

であれば良いわけだ. このためには表から,

$$\frac{a\sqrt{N}}{\sigma} > 1.960, \quad \text{つまり } N > \left(\frac{1.960 \times \sigma}{a}\right)^2 \quad (3.4.3)$$

であればよいことがわかる.

3. 具体的な数値を入れると, $a = 0.1$ の場合には, $N > (1.960)^2 \approx 385 \approx 400$ 回, となる. $a = 0.01$ なら $N > (196)^2 \approx 38416 \approx 4 \times 10^4$ 回, である.

4. まあ, 定理の通り, 誤差を評価します. 具体的な数値を入れることはしないけども, 定理による誤差評価が中心極限定理に出てくる誤差と同じオーダーである事(従って, 「大嘘」もまんざらウソではないかも知れない — せいぜい, 誤差を実際の半分くらいに見積もる, 位のウソ — こと)は知っていて損はないと思います.

問 8: チェビシェフの不等式を用いると,

$$P\left[P\left[\left|\frac{S_N}{N} - \mu\right| > a\right]\right] = P\left[|S_N - N\mu| > Na\right] \leq \frac{\text{Var}[S_N]}{N^2 a^2} = \frac{\sigma^2}{N a^2} \quad (3.4.4)$$

となるので, これがやはり $1/20$ 以下であるためには,

$$\frac{\sigma^2}{N a^2} < \frac{1}{20} \quad \implies \quad N > 20 \left(\frac{\sigma}{a}\right)^2 \quad (3.4.5)$$

であれば良い. これを中心極限定理の場合と比べると, 必要な N の回数が

$$\frac{20}{(1.96)^2} \approx 5.206 \quad (3.4.6)$$

倍になっている. チェビシェフの不等式を用いるこの方法は「大嘘」もなく, 正しいものであるから, 必要な N がある程度大きくなるのは仕方ない.

具体的な数値としては, $a = 0.1$ で $N > 2000 = 2 \times 10^3$, $a = 0.01$ で $N > 2 \times 10^5$ となる.

問 9: 一般の場合を考えるため, 100 を N と書く. 考えたい事象は $\prod_{j=1}^N X_j < a^N$ というもの

であるが, 両辺の対数をとると, これは $\sum_{j=1}^N \log X_j < N \log a$ となる. 従って, $Y_j \equiv X_j$ なる確率変数を導入して, 確率

$$P\left[\sum_{j=1}^N Y_j < N \log a\right] \quad (3.4.7)$$

を計算すればよい. さて,

$$\mu \equiv E[Y_1] = \frac{1}{6} \sum_{k=1}^6 \log k = \frac{1}{6} \log 6! \approx 1.0965, \quad (3.4.8)$$

かつ

$$\sigma^2 \equiv \text{Var}[Y_1] = \frac{1}{6} \sum_{k=1}^6 (\log k)^2 - \mu^2 \approx 0.3659, \quad \implies \quad \sigma \approx 0.6049 \quad (3.4.9)$$

である．従って, $S_N = \sum_{j=1}^N Y_j$ とおくと,

$$\begin{aligned} P\left[\sum_{j=1}^N Y_j < N \log a\right] &= P[S_N < N \log a] = P\left[\frac{S_N - N\mu}{\sigma\sqrt{N}} < \sqrt{N}(\log a - \mu)/\sigma\right] \\ &\approx \int_{-\infty}^A \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz, \quad A \equiv \sqrt{N}(\log a - \mu)/\sigma \end{aligned} \quad (3.4.10)$$

となる．これが一応の答え．

しかし, よく考えると, 上では非常に損をしていたことに気づく． $X_j \geq 1$ なので $Y_j \geq 0$ であり, $S_N \geq 0$ でもある．だから, 本当は

$$\begin{aligned} P\left[\sum_{j=1}^N Y_j < N \log a\right] &= P\left[-10\mu/\sigma < \frac{S_N - N\mu}{\sigma\sqrt{N}} < \sqrt{N}(\log a - \mu)/\sigma\right] \\ &\approx \int_{-B}^A \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz, \quad A \equiv \frac{\sqrt{N}(\log a - \mu)}{\sigma}, \quad B \equiv \frac{\sqrt{N}\mu}{\sigma} \end{aligned} \quad (3.4.11)$$

とやる方が良い．

問 10: この問題は本来, X_n, Y_n などが有界ではない場合も込みでやってもらいたかったのだが, かなりの方が有界性を仮定してしまっていた．

ともかく, 和の方は易しい．仮定から, 任意の $\epsilon > 0$ に対して,

$$\lim_{n \rightarrow \infty} P[|X_n - X| > \epsilon] = 0, \quad \lim_{n \rightarrow \infty} P[|Y_n - Y| > \epsilon] = 0 \quad (3.4.12)$$

が成立している．我々が言いたいのは, 任意の $\epsilon > 0$ に対して,

$$\lim_{n \rightarrow \infty} P[|(X_n + Y_n) - (X + Y)| > \epsilon] = 0 \quad (3.4.13)$$

となる事である．さて, 三角不等式から

$$|X_n - X| \leq \frac{\epsilon}{2} \quad \text{かつ} \quad |Y_n - Y| \leq \frac{\epsilon}{2} \quad \implies \quad |(X_n - X) + (Y_n - Y)| \leq \epsilon \quad (3.4.14)$$

であるので, この対偶をとると,

$$\begin{aligned} P[|(X_n + Y_n) - (X + Y)| > \epsilon] &\leq P\left[|X_n - X| > \frac{\epsilon}{2} \text{ または } |Y_n - Y| > \frac{\epsilon}{2}\right] \\ &\leq P\left[|X_n - X| > \frac{\epsilon}{2}\right] + P\left[|Y_n - Y| > \frac{\epsilon}{2}\right] \end{aligned} \quad (3.4.15)$$

が得られる．この両辺で $n \rightarrow \infty$ とすると, 右辺は仮定からゼロに行くので, 左辺もゼロに行く．つまり, 目標が証明された．

さて, 問題は積の方である．積の方は X, Y が有界であると仮定すれば易しいので, この場合をまずは解説する．それには,

$$X_n Y_n - XY = (X_n - X)(Y_n - Y) + (X_n - X)Y + X(Y_n - Y) \quad (3.4.16)$$

に注意すると良い．すると，

$$\begin{aligned}
 P[|X_n Y_n - XY| > \epsilon] & \\
 & \leq P\left[|(X_n - X)(Y_n - Y)| > \frac{\epsilon}{3} \text{ または } |(X_n - X)Y| > \frac{\epsilon}{3} \text{ または } |X(Y_n - Y)| > \frac{\epsilon}{3}\right] \\
 & \leq P\left[|(X_n - X)(Y_n - Y)| > \frac{\epsilon}{3}\right] + P\left[|(X_n - X)Y| > \frac{\epsilon}{3}\right] + P\left[|X(Y_n - Y)| > \frac{\epsilon}{3}\right]
 \end{aligned} \tag{3.4.17}$$

が成り立つ．ところで， $|X|, |Y|$ の上限を M とすると，

$$\begin{aligned}
 P\left[|(X_n - X)Y| > \frac{\epsilon}{3}\right] & \leq P\left[|X_n - X| > \frac{\epsilon}{3M}\right], \\
 P\left[|Y(Y_n - Y)| > \frac{\epsilon}{3}\right] & \leq P\left[|Y_n - Y| > \frac{\epsilon}{3M}\right],
 \end{aligned} \tag{3.4.18}$$

であるし，

$$P\left[|(X_n - X)(Y_n - Y)| > \frac{\epsilon}{3}\right] \leq P\left[|X_n - X| > \sqrt{\frac{\epsilon}{3}}\right] + P\left[|Y_n - Y| > \sqrt{\frac{\epsilon}{3}}\right] \tag{3.4.19}$$

である．これらは全て， $n \rightarrow \infty$ でゼロに行くので，

$$\lim_{n \rightarrow \infty} P[|X_n Y_n - XY| > \epsilon] = 0 \tag{3.4.20}$$

が結論できるわけ．

しかし，有界でない場合も扱って欲しい，と言うのが本音であった． X, Y が有界でない場合，上のやり方では少し困る．実際， Y が非常に大きい（例： $|Y| \gg 1/\epsilon$ ）場合は， $|(X_n - X)Y| > \frac{\epsilon}{3}$ は $X_n - X$ が $O(1)$ でも成り立つので，この事象の確率も小さくないことが起こりうる．

この問題を解決するには，大きな数 $A > 0$ をまずは固定し， $|X|, |Y|$ が A を超える場合と超えない場合に分類して考えると良い（この A 自身，最終段階でゼロへ持っていく）．詳しく言うと，以下ようになる．まず，考えている事象 $|X_n Y_n - XY| > \epsilon$ を3つの事象に分ける：

$$\begin{aligned}
 \{|X_n Y_n - XY| > \epsilon\} & \subset \left[\{|X_n Y_n - XY| > \epsilon\} \cap \{|X| > A\} \right] \cup \left[\{|X_n Y_n - XY| > \epsilon\} \cap \{|Y| > A\} \right] \\
 & \quad \cup \left[\{|X_n Y_n - XY| > \epsilon\} \cap \{|X| \leq A\} \cap \{|Y| \leq A\} \right]
 \end{aligned} \tag{3.4.21}$$

最初の2つの事象の確率は，単純に $|X| > A$ などの確率で押さえてしまう：

$$P\left[|X_n Y_n - XY| > \epsilon \text{ かつ } |X| > A\right] \leq P[|X| > A] \tag{3.4.22}$$

$$P\left[|X_n Y_n - XY| > \epsilon \text{ かつ } |Y| > A\right] \leq P[|Y| > A] \tag{3.4.23}$$

一方，最後の確率は X, Y が有界な時と全く同じで，

$$\begin{aligned}
 & P\left[\{|X_n Y_n - XY| > \epsilon\} \cap \{|X| \leq A\} \cap \{|Y| \leq A\}\right] \\
 & \leq P\left[|X_n - X| > \frac{\epsilon}{3A}\right] + P\left[|Y_n - Y| > \frac{\epsilon}{3A}\right] + P\left[|Y_n - Y| > \sqrt{\frac{\epsilon}{3}}\right]
 \end{aligned} \tag{3.4.24}$$

が成立する．結局，

$$\begin{aligned}
 P[|X_n Y_n - XY| > \epsilon] & \leq P[|X| > A] + P[|Y| > A] \\
 & \quad + P\left[|X_n - X| > \frac{\epsilon}{3A}\right] + P\left[|Y_n - Y| > \frac{\epsilon}{3A}\right] + P\left[|Y_n - Y| > \sqrt{\frac{\epsilon}{3}}\right]
 \end{aligned} \tag{3.4.25}$$

が任意の $A > 0, \epsilon > 0$ について成り立つことがわかった。さて、ここで A を固定したまま n を無限大に送ると、第 2 行の寄与が全部ゼロに収束して、

$$\lim_{n \rightarrow \infty} P[|X_n Y_n - XY| > \epsilon] \leq P[|X| > A] + P[|Y| > A] \quad (3.4.26)$$

が得られる。さて、ここで A は任意だったので、今度は $A \rightarrow \infty$ の極限を考えると、 $P[|X| > A]$, $P[|Y| > A]$ はゼロに行く（このところ、極限をとる順序に注意！）つまり、任意の $\epsilon > 0$ に対して

$$\lim_{n \rightarrow \infty} P[|X_n Y_n - XY| > \epsilon] = 0 \quad (3.4.27)$$

が証明されたので、 $X_n Y_n$ は XY に確率収束する。□

問 11: ヒントでは Z_n と Z_{2n} を比べよ、としたが、 Z_{4n} の方が（平方根が出ないだけ）楽だったな。まあ、とにかく証明の方針は背理法による。

正規分布に従う確率変数を Z と書いて、 Z_n が Z に確率収束したと仮定しよう。当然、 Z_{2n} も (Z_n の部分列ゆえ) Z に確率収束するはずである。しかし、これに問 10 を用いると、 $Z_{2n} - Z_n$ が $Z - Z \equiv 0$ に確率収束する、ことが結論されてしまう。そこで、以下では「 $Z_{2n} - Z_n$ が 0 に確率収束する」のはあり得ない、ことを示して、矛盾を導こう。

そのためには、具体的に計算してみる。先を見越して、

$$A_n \equiv \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n X_j, \quad B_n \equiv \frac{1}{\sigma\sqrt{n}} \sum_{j=n+1}^{2n} X_j \quad (3.4.28)$$

を定義しておく、

$$Z_{2n} - Z_n = \frac{1}{\sigma\sqrt{2n}} \sum_{j=1}^{2n} X_j - \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^n X_j = \frac{A_n + B_n}{\sqrt{2}} - A_n = \frac{-\sqrt{2} + 1}{\sqrt{2}} A_n + \frac{1}{\sqrt{2}} B_n \quad (3.4.29)$$

であり、かつ（非常に重要） A_n と B_n は独立である。従って、任意の $\epsilon > 0$ に対して、

$$\begin{aligned} P[|Z_{2n} - Z_n| > \epsilon] &= P\left[\left|\frac{-\sqrt{2} + 1}{\sqrt{2}} A_n + \frac{1}{\sqrt{2}} B_n\right| > \epsilon\right] = P\left[|-(\sqrt{2} - 1)A_n + B_n| > \sqrt{2}\epsilon\right] \\ &\geq P[A_n \leq 0 \text{ かつ } B_n > \sqrt{2}\epsilon] \stackrel{\text{独立}}{=} P[A_n \leq 0] \times P[B_n > \sqrt{2}\epsilon] \end{aligned} \quad (3.4.30)$$

が成り立つ。ところが、 A_n, B_n は共に Z に法則収束するので、 $n \rightarrow \infty$ では上の右辺は

$$\int_{-\infty}^0 \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \times \int_{\epsilon}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = \frac{1}{2} \int_{\sqrt{2}\epsilon}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \quad (3.4.31)$$

に収束してしまう。つまり

$$\lim_{n \rightarrow \infty} P[|Z_{2n} - Z_n| > \epsilon] \geq \frac{1}{2} \int_{\sqrt{2}\epsilon}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \quad (3.4.32)$$

が証明されたわけだが、 ϵ が有限の限り、上の積分の値は正なので、これは $Z_{2n} - Z_n$ がゼロに確率収束することはあり得ないことを示す。□

問 12 :

大変に申し訳ない。この問題では X と Y が独立であるという条件が不可欠であるが、それをなぜか問題文に書き忘れてしまった。何人かの人は多いに悩んだようで、申し訳ない。

独立性を仮定する解答例は以下の通り：

(方針) X の特性関数を $\phi(t)$ と書く。 Y も同分布だから、 Y の特性関数も $\phi(t)$ だし、 $\frac{X+Y}{\sqrt{2}}$ の特性関数も $\phi(t)$ である。 $\frac{X+Y}{\sqrt{2}}$ の特性関数を X, Y の特性関数で表すことにより、 ϕ の満たすべき方程式を作って、それを解く。

(以下、方針の実行) さて、 X, Y が独立の場合、 $\frac{X+Y}{\sqrt{2}}$ の特性関数 $\psi(t)$ は、

$$\begin{aligned}\psi(t) &= E\left[\exp\left(it\frac{X+Y}{\sqrt{2}}\right)\right] = E\left[\exp\left(it\frac{X}{\sqrt{2}}\right)\exp\left(it\frac{Y}{\sqrt{2}}\right)\right] \stackrel{\text{独立}}{=} E\left[\exp\left(it\frac{X}{\sqrt{2}}\right)\right]E\left[\exp\left(it\frac{Y}{\sqrt{2}}\right)\right] \\ &= \phi\left(\frac{t}{\sqrt{2}}\right) \times \phi\left(\frac{t}{\sqrt{2}}\right)\end{aligned}\quad (3.4.33)$$

を満たすが、同分布なので、この ψ 自身が ϕ に等しい。つまり、 ϕ は

$$\phi(t) = \left\{\phi\left(\frac{t}{\sqrt{2}}\right)\right\}^2 \quad (3.4.34)$$

を全ての実数値 t に対して満たすのである。後はこれを解けばよい。

どのようにしても解けるが、一つの手は上の関係式を何回も使って

$$\phi(t) = \left\{\phi\left(\frac{t}{\sqrt{2}}\right)\right\}^2 = \left\{\phi\left(\frac{t}{2}\right)\right\}^4 = \left\{\phi\left(\frac{t}{2\sqrt{2}}\right)\right\}^8 = \dots = \left\{\phi\left(\frac{t}{2^n}\right)\right\}^{4^n} \quad (3.4.35)$$

が任意の $n \geq 1$ で成り立つことに注目することであろう。命題 2.7.4 から

$$\phi\left(\frac{t}{2^n}\right) = 1 - \frac{1}{2}\left(\frac{t}{2^n}\right)^2 + o\left(\left(\frac{t}{2^n}\right)^2\right) \quad (3.4.36)$$

が成り立っているので、これを代入して $n \rightarrow \infty$ を考えれば (中心極限定理の証明と同じく)

$$\phi(t) = \left\{\phi\left(\frac{t}{2^n}\right)\right\}^{4^n} = \left\{1 - \frac{1}{2}\frac{t^2}{4^n} + o\left(\left(\frac{t}{2^n}\right)^2\right)\right\}^{4^n} \xrightarrow{n \rightarrow \infty} e^{-t^2/2} \quad (3.4.37)$$

を得る。特性関数が正規分布の特性関数 $e^{-t^2/2}$ に一致したので、 $X, Y, \frac{X+Y}{\sqrt{2}}$ の分布は全て正規分布である。 □

—以下、レジュメの続き—

3.5 再帰性について：一般的な方法

さて、本題にもどる。ランダムウォークの再帰性について考えていたのだった。そこでの一般的結論 (3.2.12) から従う) は、

$$\text{ランダムウォークが再帰的} \iff \sum_{\ell=0}^{\infty} P_{\ell}(0) = \infty$$

であったので, $\sum_{\ell=0}^{\infty} P_{\ell}(0)$ が無限大かどうかを調べればよい. それで実際に 1 次元と 2 次元の場合についての泥臭い計算をやってみたのだが, これは大変だった.

この小節では, もっと格好の良いやり方を紹介する. この方法によって, もっと一般のランダムウォークの再帰性を判定することも可能になる.

3.5.1 フーリエ変換へ持ち込むための準備

少し一般的に考えたいが, 考えやすいように, まずは 1 次元で考える. 今までのランダムウォークを少しだけ一般化して, 点 x にいる粒子が次に y に跳ぶ確率を $p_{x,y}$ と書くことにする. ただし, 平行移動不変性 ($p_{x,y} = p_{0,y-x}$) を仮定しておく. 例えば 3.2 節で考えた, 隣の点へ跳ぶものなら,

$$p_{xy} = \begin{cases} p & (y = x - 1, \text{つまり左へ跳ぶ}) \\ 1 - p & (y = x + 1, \text{つまり右へ跳ぶ}) \\ 0 & (\text{それ以外}) \end{cases} \quad (3.5.1)$$

に相当する.

$p_{x,y}$ を使って $P_{\ell}(z)$ を表すことが出来る (z は任意の点). 実際,

$$P_1(z) = p_{0,z}, \quad P_2(z) = \sum_x p_{0,x} p_{x,z} = (p^2)_{0,z}, \quad P_3(z) = \sum_{x,y} p_{0,x} p_{x,y} p_{y,z} = (p^3)_{0,z} \quad (3.5.2)$$

などとなり (p^2, p^3 などは行列としての p の積を表す), 一般に

$$P_{\ell}(z) = \sum_{x_1, x_2, x_3, \dots, x_{\ell-1}} p_{0, x_1} p_{x_1, x_2} \cdots p_{x_{\ell-1}, z} = (p^{\ell})_{0, z} \quad (3.5.3)$$

が成り立つ. つまり, P_{ℓ} は行列 p の ℓ 個の積でかけることがわかる.

と言うわけで, 問題は, 行列 p の ℓ 個の積を求め, 更に ℓ について和をとること, となった. これは普通は大変なのであるが, 今考えているように p が並進不変性を持っている場合には, 「フーリエ変換」を用いて簡単に解ける.

3.5.2 フーリエ変換の復習

フーリエ変換について完全に解説することは出来ないので, 要点だけをまとめておく.

整数点 $x \in \mathbb{Z}$ 上で定義された関数 $f(x)$ と $k \in [-\pi, \pi]$ に対して

$$\hat{f}(k) = \sum_{x \in \mathbb{Z}} e^{-ikx} f(x) \quad (3.5.4)$$

を, f のフーリエ変換と言う (勿論, この和が収束するときのみ定義). このとき, f または \hat{f} が適当な条件を満たすならば,

$$f(x) = \int_{-\pi}^{\pi} \hat{f}(k) e^{ikx} \frac{dk}{2\pi} \quad (3.5.5)$$

も成り立つ. つまり, f を \hat{f} によって表すことができる.

フーリエ変換の著しい性質は、「 f の方の畳み込み (convolution) が \hat{f} の方では積になる」ことにある。具体的には、 f, g のフーリエ変換をそれぞれ \hat{f}, \hat{g} とし、また、 f, g の畳み込み $f * g$ を

$$(f * g)(x) \equiv \sum_{y \in \mathbb{Z}} f(x - y)g(y) \quad (3.5.6)$$

で定義すると、 $f * g$ のフーリエ変換 $\widehat{f * g}$ は

$$\widehat{f * g}(k) = \hat{f}(k)\hat{g}(k) \quad (3.5.7)$$

を満たす。

3.5.3 $P_\ell(z)$ の計算

以上を基にして、 $P_\ell(z)$ を計算してみよう。対応関係を明確にするために $f(x) = p_{0,x}$ と書くことにすると、(3.5.3) は

$$P_\ell(z) = \underbrace{(f * f * f * \dots * f)}_{\ell \text{個}}(z) \quad (3.5.8)$$

と、 ℓ 個の f の畳み込みで書けるから、上に述べたことから

$$P_\ell(z) = \int_{-\pi}^{\pi} (\hat{f}(k))^\ell e^{ikz} \frac{dk}{2\pi} \quad (3.5.9)$$

が成り立つ。特に $z = 0$ を見たいのであるが、これは

$$P_\ell(0) = \int_{-\pi}^{\pi} (\hat{f}(k))^\ell \frac{dk}{2\pi} \quad (3.5.10)$$

と書け、 ℓ についての和をとってしまうと (少々厳密性を無視して和と積分を入れ替える)²

$$\sum_{\ell=0}^{\infty} P_\ell(0) = \sum_{\ell=0}^{\infty} \int_{-\pi}^{\pi} (\hat{f}(k))^\ell \frac{dk}{2\pi} = \int_{-\pi}^{\pi} \sum_{\ell=0}^{\infty} (\hat{f}(k))^\ell \frac{dk}{2\pi} = \int_{-\pi}^{\pi} \frac{1}{1 - \hat{f}(k)} \frac{dk}{2\pi} \quad (3.5.11)$$

が得られる。つまり、 $\sum_{\ell=0}^{\infty} P_\ell(0)$ が有限かどうかは、上の k -積分が収束するかどうかにかかっている — 分母の $1 - \hat{f}(k)$ はゼロになるのか? となるとすればどのくらいの速さで?

3.5.4 1次元の計算

例として1次元の場合をやり直してみよう。 $p = q = 1/2$ の場合を考えていたので、

$$f(x) = \begin{cases} 1/2 & (x = \pm 1) \\ 0 & (\text{その他}) \end{cases} \quad \hat{f}(k) = \cos(k) \quad (3.5.12)$$

である。従って、問題の (3.5.11) の積分は

$$\int_{-\pi}^{\pi} \frac{1}{1 - \cos(k)} \frac{dk}{2\pi} \quad (3.5.13)$$

²このところは勿論、厳密にできる。 ℓ についての和を有限に限って和と積分を入れ替え、後で ℓ の範囲を無限大にすればよい。一般にはこの手順そのものが疑問なのだが、今は和の各項が正またはゼロだから許される

となる．考えている積分範囲ではこの分母は $k = 0$ でのみゼロになり，かつ $k \approx 0$ では非積分関数は $1/k^2$ のように振る舞う．これは $k = 0$ の近傍で可積分ではないので，この積分は発散する．つまり，1次元のランダムウォークは再帰的である（二項係数から Stirling の公式を用いたことに比べれば，何と簡単に出了事よ！）．

この方法なら，もっと一般の1次元ランダムウォークに拡張することもできる．例えば L を大きな数として，

$$p_{0,x} = a_x \quad (|x| < L); \quad \text{ただし } a_x = a_{-x} \quad (3.5.14)$$

となっているような（隣以外への点へも跳べる）ランダムウォークを考えると，

$$\hat{f}(k) = \sum_x a_x e^{-ikx} = a_0 + 2 \sum_{x>0} a_x \cos kx \quad (3.5.15)$$

である．全確率が1であることから出る

$$\sum_x a_x = 1 \quad (3.5.16)$$

をも考慮に入れると，

$$\begin{aligned} 1 - \hat{f}(k) &= \sum_x a_x - \left(a_0 + 2 \sum_{x>0} a_x \cos kx \right) = 2 \sum_{x>0} a_x (1 - \cos kx) \\ &\leq 2 \sum_{x>0} a_x \frac{(kx)^2}{2} = k^2 \sum_{x>0} x^2 a_x \end{aligned} \quad (3.5.17)$$

が得られる．よって，

$$\sum_{\ell=0}^{\infty} P_{\ell}(0) = \int_{-\pi}^{\pi} \frac{1}{1 - \hat{f}(k)} \frac{dk}{2\pi} \geq \left(\sum_{x>0} x^2 a_x \right)^{-1} \int_{-\pi}^{\pi} \frac{1}{k^2} \frac{dk}{2\pi} \quad (3.5.18)$$

となつて後ろの積分は発散するから， $\sum_{x>0} x^2 a_x < \infty$ である限りは $\sum_{\ell=0}^{\infty} P_{\ell}(0) = \infty$ で，これらのランダムウォークは全て再帰的だと結論できる．

3.5.5 2次元の計算

2次元でも基本的な違いはない．ただし，今まで x, k と書いてきたものがそれぞれ2次元のベクトル $\vec{x} = (x_1, x_2), \vec{k} = (k_1, k_2)$ になるところだけが異なる．フーリエ変換と逆変換の式は

$$\hat{f}(k_1, k_2) = \sum_{x_1, x_2} \exp\{-i(x_1 k_1 + x_2 k_2)\} f(x_1, x_2), \quad (3.5.19)$$

$$f(x_1, x_2) = \int_{-\pi}^{\pi} \frac{dk_1}{2\pi} \int_{-\pi}^{\pi} \frac{dk_2}{2\pi} \exp\{i(x_1 k_1 + x_2 k_2)\} \hat{f}(k_1, k_2) \quad (3.5.20)$$

となる．3.3.1節のモデルに対しては，

$$f(x_1, x_2) = \frac{1}{4} \quad (x_1^2 + x_2^2 = 1) \quad (3.5.21)$$

（それ以外はゼロ）であるので，

$$\hat{f}(k_1, k_2) = \frac{\cos k_1 + \cos k_2}{2} \quad (3.5.22)$$

である．従って，

$$\sum_{\ell=0}^{\infty} P_{\ell}(0) = \int_{-\pi}^{\pi} \frac{dk_1}{2\pi} \int_{-\pi}^{\pi} \frac{dk_2}{2\pi} \frac{2}{2 - (\cos k_1 + \cos k_2)} \quad (3.5.23)$$

の収束，発散が問題になる．この積分の収束，発散は $k_1 \approx k_2 \approx 0$ での非積分関数の振る舞いで決まるが，

$$0 \leq 2 - (\cos k_1 + \cos k_2) \leq \frac{k_1^2 + k_2^2}{2} \implies \frac{2}{2 - (\cos k_1 + \cos k_2)} \geq \frac{4}{k_1^2 + k_2^2} \quad (3.5.24)$$

である事に注意すると，この積分は発散することがわかる．すなわち，この2次元の場合もランダムウォークは再帰的なのである．

1次元の時と同じように，もっと一般的なランダムウォークにこの結果を一般化する事も勿論可能である．要するに (k_1, k_2 の小さいところで)

$$1 - \hat{f}(k_1, k_2) \approx k_1^2 + k_2^2 \quad (3.5.25)$$

となるとヤバイ (積分が発散する) わけであるが，1次元の時と同じように考えると，

$$p_{\vec{0}, \vec{x}} = a_{x_1, x_2}, \quad \sum_{x_1, x_2} (x_1^2 + x_2^2) a_{x_1, x_2} < \infty \quad (3.5.26)$$

であるようなランダムウォークは全て再帰的，とわかる (ただし，係数 a_{x_1, x_2} については対称性 $a_{x_1, x_2} = a_{\pm x_1, \pm x_2}$ を仮定して) ．

3.6 3次元のランダムウォーク

今までの拡張として3次元のランダムウォークを考える．今度は確率 $\frac{1}{6}$ で「東西南北上下」の6方向にランダムに動く．これは煙の粒子のブラウン運動などのモデル化である．

大まかな振る舞いは1次元，2次元と同じである．時刻 n での位置 \vec{S}_n の期待値はゼロ，また原点からの距離は大体 \sqrt{n} ，つまり (3.3.4)，(3.3.7) が成り立つ．この観点からは次元による差異は認められない．

大きな差が出るのは再帰性に関してである．2次元でもなかなか帰って来なかったんだから3次元ならもっと帰ってこないことは予想できる (1次元では2方向ある内の一つを選べば帰れるが，2次元では4方向の内1つを選ばないといけない．3次元では6方向中の1方向で，いよいよ大変) ．

この問題はやはり，(3.2.22) や (3.3.14) に相当する式を計算することによって解決される．特に問題になるのは今までと同じく， $\sum_{n=0}^{\infty} P_n(\vec{0})$ が有限か否か，である．この計算を多項係数を使ってやるのは非常に大変であるが，不可能ではない (少なくとも3次元では) ．しかし，ここでは折角学んだフーリエ変換を用いてやってみよう．3次元では $\vec{x} = (x_1, x_2, x_3)$ ， $\vec{k} = (k_1, k_2, k_3)$ と3成分になるので，問題の和は

$$\sum_{n=0}^{\infty} P_n(\vec{0}) = \int_{-\pi}^{\pi} \frac{dk_1}{2\pi} \int_{-\pi}^{\pi} \frac{dk_2}{2\pi} \int_{-\pi}^{\pi} \frac{dk_3}{2\pi} \frac{1}{1 - \hat{f}(k_1, k_2, k_3)} \quad (3.6.1)$$

となる．単純ランダムウォークの場合は

$$\hat{f}(k_1, k_2, k_3) = \frac{\cos k_1 + \cos k_2 + \cos k_3}{3} \quad (3.6.2)$$

であるから

$$\sum_{n=0}^{\infty} P_n(\vec{0}) = \int_{-\pi}^{\pi} \frac{dk_1}{2\pi} \int_{-\pi}^{\pi} \frac{dk_2}{2\pi} \int_{-\pi}^{\pi} \frac{dk_3}{2\pi} \frac{3}{3 - (\cos k_1 + \cos k_2 + \cos k_3)} \quad (3.6.3)$$

となる.

さて, 今回は上の積分は有限である (why?). 従って, 3次元の単純ランダムウォークは推移的であることがわかる. 4次元以上でも同様の計算を行うと, 以下が結論できる:

3次元では(更に4次元以上でも), $\sum_n P_n(\vec{0})$ は有限である. 従って, ランダムウォークは推移的で, いくら待っても原点に戻ってこないものがたくさん存在する.

いろんな次元における $r(\vec{0})$ (いつかは原点に戻ってくる確率) の値を表にしておこう.

d	3	4	5	6	7	8	9	10
r	0.3405373	0.1932017	0.1351786	0.1047155	0.0858449	0.0729126	0.0634477	0.0561975

次元が大きくなるほど帰って来にくい. 例えば, 仮想的に10次元(!)の空間を考えると, いくら待っても5.6%しか帰ってこない事がわかる.

3.7 戻ってくるための歩数 (おまけ)

最後に(確実に)原点に戻ってくるためにはどのくらいの歩数が必要か, 考えておこう. ランダムウォークには色々なものがあるから, あるものは短い歩数で戻ってくるし, あるものはなかなか戻ってこない. 「無限時間まで待てば確実に戻ってくる」と言うのが上の結論だが, 無限時間も待たない場合, 何パーセントくらいのランダムウォークが戻ってきていないだろうか, と言うことを考えたい.

この問に対する答は1次元と2次元でかなり異なる. 1次元の場合は(3.3.20)に相当する計算を行うと,

$$\sum_{k=1}^N F_{2k}(0) \approx 1 - \frac{1}{\sum_{m=1}^N P_{2m}(0)} \approx 1 - \frac{\sqrt{\pi}}{2\sqrt{N}} \quad (3.7.1)$$

となっている(ちょっと N と $2N$ をごまかしたが...). これから時刻 N では全体の $\frac{1}{\sqrt{N}}$ くらいの割合のランダムウォークが原点に戻って来ていない, と結論できる. 2次元では(3.3.20)であるから, 時刻 N で帰ってきていないのは $\frac{\pi}{\log N}$ である.

勿論, 1次元でも2次元でも, 帰ってきていないウォークの割合は, $N \rightarrow \infty$ ではゼロになる. ただ, ここで見たように, ゼロになるなり方は2次元では非常に遅い. $N = 10^8$ の時, 1次元ではたったの0.01%だけが帰ってきていないのに対し, 2次元では17%くらいも帰ってきていないのだ. $N = 10^{32}$ まで待っても, 2次元では4%も帰っていない.

上では原点に戻ってくることを考えたが, 一般の点 (x, y) にたどり着く確率も同様の振る舞いをする. つまり, 無限時間待てば, 任意の点 (x, y) に必ずたどり着く. ただし, たどり着くまでの時間は2次元では非常にかかる.

(と言うことは酔っぱらいのおっさんは家にたどり着く前に餓死している公算が高い...)

最後に, 2次元のランダムウォークの例を挙げておこう.

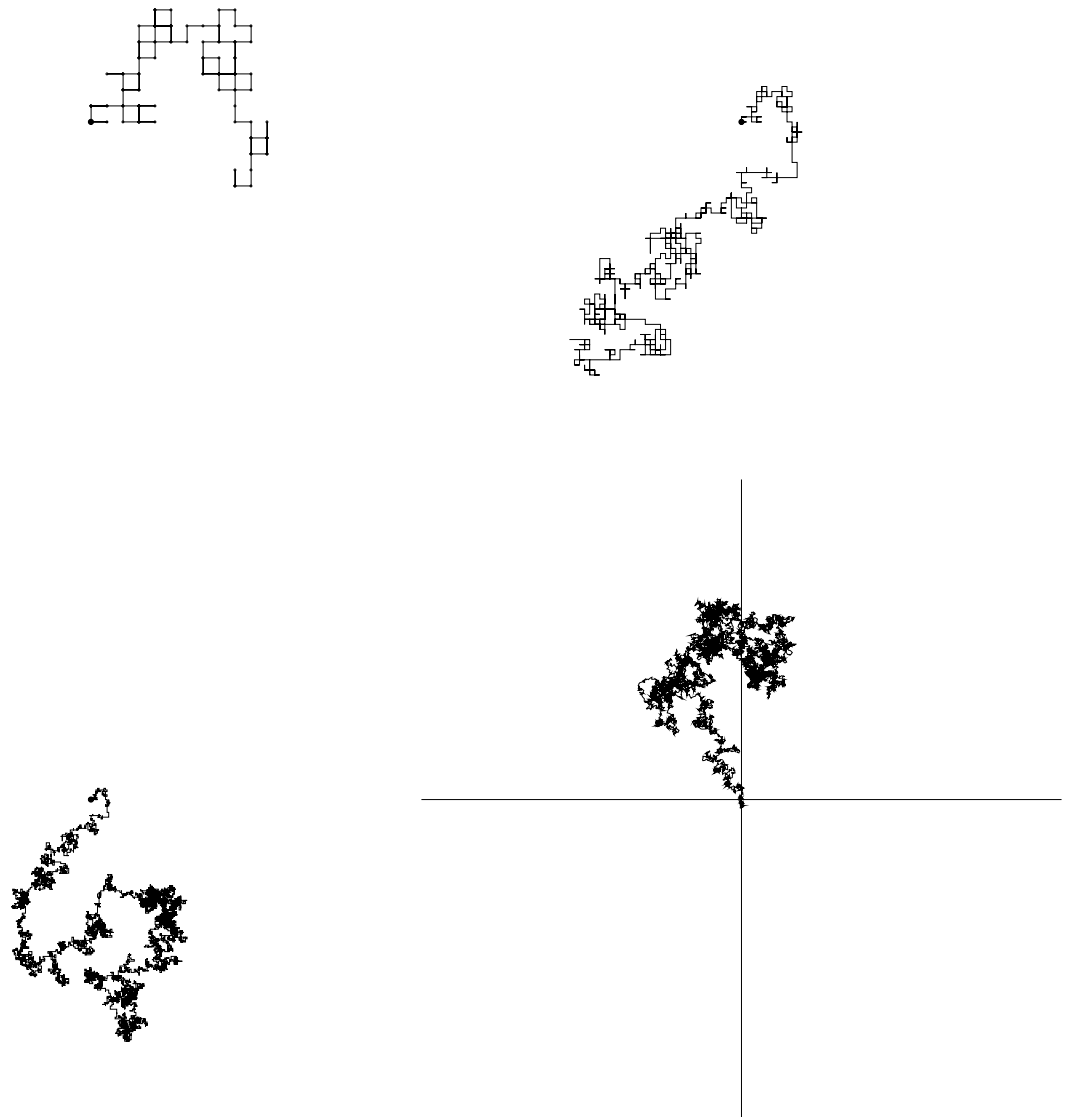


図 1: 2次元ランダムウォークの例 : 左上から $10^2, 10^3, 10^4, 10^7$ steps. 横軸は x , 縦軸は y で, 図示している範囲は $2\sqrt{n}$.

4 ブラウン運動

この講義も終盤にさしかかった。最後の章では「ブラウン運動」について考える。実のところ、これまでにややコシイ「高度な」理論を避けてきたことのツケが回ってきて、ブラウン運動を厳密に定義する（と言うか、定義したものが実際に存在することを示す）ことはなかなか難しい状況である。ある程度、高度な理論を仮定しての話になることをお断りしておく。

記号について：特に断らない限り、この節では n, m と書けばどちらも整数値（特に非負の整数値）をとるものとする。 s, t の方は実数値（時には非負の実数値）をとるものとする。

4.1 確率過程としてのランダムウォークとその paths

まず、ブラウン運動への布石として、ランダムウォークを復習し、これを例にとって「確率過程」の概念を導入する。話を簡単にするために1次元に限るが、多次元の場合もほとんど同じように議論できる。

4.1.1 ランダムウォークの復習

左右に等確率で動く1次元ランダムウォークを思いだそう。これは

- 粒子は原点 ($x = 0$) から出発し、確率的に左右に移動していく。
- その移動のルールは、今までの履歴には全く関係なく確率 $\frac{1}{2}$ で左へ一歩、確率 $\frac{1}{2}$ で右へ一歩である

と言うものであった。確率 $\frac{1}{2}$ で ± 1 の値をとる、独立・同分布の確率変数 X_1, X_2, \dots を用いると、 n 歩めでの粒子の位置 S_n を

$$S_n = X_1 + X_2 + \dots + X_n \quad (4.1.1)$$

表すこともできた。

4.1.2 確率過程とその paths

今まではランダムウォークを粒子の運動としてとらえてきたが、これからはこの粒子の軌跡にも注目したい。

まず、「確率過程」の概念を導入しよう。

定義 4.1.1 実数のパラメータ t で番号づけられた確率変数の集まり $\{W_t\}_{t \in \mathbb{R}}$ を確率過程と言う。実数のパラメータ t が整数値（やその一部分）のみをとる場合も確率過程と言う。確率変数自身は実数値をとる場合を考えると多いが、もっと一般の空間の値をとっても良い。

(例) ランダムウォークの場合： n 歩目の位置を S_n と書くと、 n が実数のパラメータ、 S_n が確率変数として、上の確率過程の定義にかなっている。この場合、1次元ランダムウォークな

らば S_n は実数値確率変数, n 次元ランダムウォークならば S_n は \mathbb{R}^n に値をとる確率変数, になる.

確率過程とはこのように、「時刻」 t のそれぞれの値が確率的に決まるようなもの(過程)のことである. ただし, わざわざ確率過程と言うときは, それぞれの時刻での W_t をバラバラに考えるのではなく, 色々な時刻 t での W_t の値を全体として見る視点を意識している.

それぞれの時刻 t 毎に W_t が確率変数になっているのであるから, W_t をまとめて扱うには, 背後に(かなり大きな)各 t に共通の確率空間 (Ω, \mathcal{F}, P) があって, その上でそれぞれの W_t が定義されている必要がある.

なお, しつこいけども, W_t 自身は確率変数であるから, Ω の元 ω 毎に値が変わる. この意味で, W_t のランダム性 (ω) への依存度を強調するときには $W_t(\omega)$ と書くこともある.

さて, $\omega \in \Omega$ を一つ固定する毎に, 各時刻 t での W_t の値 $W_t(\omega)$ が定まる. 各時刻 t での $W_t(\omega)$ の値をつなぐと, 各 ω 毎に t の関数 $W_t(\omega)$ が決まる. これは W_t が値をとる空間の中の道のように見えるので (t が動いた時の粒子の軌跡のイメージ), 確率過程 $\{W_t\}$ の path と言う¹.

4.1.3 ランダムウォークの paths

一般論はこのくらいにして, ランダムウォークに戻ろう. 後でやるようにブラウン運動は連続時間 t で番号づけられた確率変数の集まりで, その path も連続時間 t の関数になっている. 一方, 今まで考えてきたランダムウォークは t が非負の整数の所でしか定義されていないので, その path も整数時間 n の関数になっている.

これでは対応関係がよく見えないので, 後への準備として, ランダムウォークの path を連続にする操作を行っておこう. ランダムウォーク $\{S_n\}_{n \geq 0}$ が与えられたとき, 隣り合った整数時刻 n での粒子の位置 S_n を折れ線で結ぶと $s \geq 0$ で定義された連続な関数(折れ線) \tilde{S}_s ができる². 念のために数式で書いておくと, $n \leq s \leq t+1$ の時,

$$\tilde{S}_s \equiv S_n + (s - n) \times (S_{n+1} - S_n) \quad (4.1.2)$$

と定義するわけである. ちょっと用語をごまかして, 以下ではこの $\{\tilde{S}_s\}_{s \geq 0}$ を, ランダムウォーク S_n の path と呼ぶことにする.

4.1.4 ランダムウォークの paths 上の測度

最後に, このようにして定義したランダムウォークの path を用いて, ランダムウォークを少し別の見方で見ておこう.

今まで, 確率過程としてのランダムウォーク $\{S_n\}_{n \geq 0}$, 更に $\omega \in \Omega$ を決める毎に sample path の点をつないでできる折れ線 $\{\tilde{S}_s\}_{s \geq 0}$ を定義してきた. しつこいけども, これらは元々の確率空間の一点 $\omega \in \Omega$ を決める毎に決まるランダムなものなので, $\{\tilde{S}_s(\omega)\}_{s \geq 0}$ とでも書くべきものである. つまり, ランダムウォークの一本一本はその path と 1 対 1 の対応がついている.

¹確率論では path を「道」と言うが, どうも抵抗がある!「道」よりは「グラフ」だと思っんですがね. まあ, この意味では path と言ったって同じなんだけど...

²すぐ後で「時刻」 t が別にてでくるので, ここでは s で時刻を表している

ところで, ランダムウォークの一本一本の実現確率は決めてある — 長さが n のランダムウォークなら, その実現確率は 2^{-n} . ランダムウォークとその path が 1 対 1 対応しているから, これから, ランダムウォークの 可能な path $\{\tilde{S}_s(\omega)\}_{s \geq 0}$ の空間に, それぞれの path の実現確率が定義されることになる.

つまり, ランダムウォークを考えると言うことは, その実現可能な path の空間上での測度 (確率) を考えることと同等なのである. 以下ではこの「path の空間上での測度」の立場から, ブラウン運動へ入っていく.

4.2 ランダムウォークからブラウン運動へ I

以上の準備の下に, ブラウン運動に取りかかろう. 教科書などではブラウン運動を天下一に定義することが多いが, この講義では敢えてそれを避け, 段階的に進むことにする.

ここで考える基本的な問は以下の通り:

ランダムウォークの path の「連続極限」は何か? つまり, ランダムウォークが $1/N$ 秒ごとに 1 ステップ進むことにした場合, $N \rightarrow \infty$ での極限で, path の全体はどんな分布になるだろうか?

(註) 上ではわざと曖昧な訊き方をしている. どのように「連続極限」をとったらよいか, と言うのも質問のうち.

このような質問は単に数学的な興味だけのものではない. 例えばブラウン粒子の運動を考えるつもりなら, 空気の分子がガンガンとぶつかる時間間隔は非常に短いはずである. この場合, $1/N$ 秒ごとにぶつかる, と考えると N が非常に大きい (ほとんど無限大) と思うべし. つまり, $N \rightarrow \infty$ の極限から物事を見る方がよく見える可能性が高いわけだ.

簡単のため, 長さが N のランダムウォークの全体を考える. それぞれのランダムウォークを (t, x) 平面の折れ線と対応づけることは既に行った. ここで問題にしているのは, この折れ線の横軸 (t -軸) を $1/N$ に縮めた場合に³何が起こるか, である. 数式で言えば, 今までの s と

$$t = \frac{s}{N} \quad (4.2.1)$$

で関係づけられた t を導入し, ランダムウォークの折れ線をこの t の関数として見よう, と言うことである.

さて, これでは良くないことはすぐにわかる. ランダムウォークの n 歩目の位置 S_n が

$$E[S_n] = 0, \quad \text{Var}[S_n] = n \quad (4.2.2)$$

を満たすことは何回も見た. ゼロでない t に対応する元々の歩数は Nt くらいだが, 上によると⁴

$$\text{Var}[S_{Nt}] = Nt \quad (\rightarrow \infty \text{ as } N \rightarrow \infty) \quad (4.2.3)$$

となってしまうぞ. と言うことは単純に $N \rightarrow \infty$ とやっては縦軸方向に無限大に振動する折れ線ができるだけで, ヤバイ.

³時刻を縮めるだけでは足りないことはすぐにわかる

⁴正確には以下で, S_{Nt} の代わりに $S_{[Nt]}$ などと書くべきである — $[Nt]$ は Nt の整数部分. でも, こんな所を気にしても記号が複雑になるだけなので, 敢えて無視することにした

この問題を回避するには、縦方向にもある程度縮めてやる必要がある。どのくらい縮めるべきかは (4.2.3) から読みとれる。すなわち、 $|S_{Nt}| = O(\sqrt{N})$ なのだから、縦方向には $1/\sqrt{N}$ に縮めればよいはずだ。

と言うわけで、ランダムウォークの時に作った \tilde{S}_s から

$$W_t^{(N)} \equiv \frac{1}{\sqrt{N}} \tilde{S}_{Nt} \quad (4.2.4)$$

となるような対応関係を付ければ良いように思える。キーは時刻の方は N で割っているのに、空間の方は \sqrt{N} で割っていることである⁵。実際にこれでうまく行きそうなことはこれから見ていく。

以上をまとめておこう (これまでは長さが N のランダムウォークに限っていたが、もっと長いウォークを考えても同じことなので、そうしておく。)

ランダムウォークの path \tilde{S}_s が一本与えられると、その時間方向を $1/N$ 、空間方向を $1/\sqrt{N}$ して (N は大きな整数)、

$$W_t^{(N)} \equiv \frac{1}{\sqrt{N}} \tilde{S}_{Nt} \quad (4.2.5)$$

なる、 (t, x) 平面での曲線を作る。 $N \rightarrow \infty$ で、この $W_t^{(N)}$ の集まり (元々のランダムウォークをいろいろと変えたときの $W_t^{(N)}$ の全体) がどのような分布になるかを知りたい。

本来ならば、これから実際に $W_t^{(N)}$ の集まりがどのような分布に従っているか、その分布の $N \rightarrow \infty$ の極限は何か、と考えていくべきなのだが、これは (1) かなり大変で長い道のりになる (2) 数学的にもかなり高度なことをやらないといけないので難しい。ので、天下りになるが $N \rightarrow \infty$ での極限の測度 (これがブラウン運動を与える) をまず書き下し、後でランダムウォークとの対応に戻ってくることにする。

4.3 ブラウン運動とは？

(以上の前置きの下に、ブラウン運動を定義しよう。ランダムウォークの極限としてのイメージから、ブラウン運動の paths はかなりギザギザな連続曲線になりそうな気はすると思う。)

さて、ブラウン運動とは、非負の実数時刻 t ($0 \leq t$) でパラメータづけられた (連続無限個の) 確率変数の集まり $\{B_t\}$ で⁶、以下を満たすものの事を言う。

定義 4.3.1 ある確率空間 (Ω, \mathcal{F}, P) 上で定義された実数値確率過程 $\{B_t\}_{t \geq 0}$ が一次元ブラウン運動である、とは、 B_t が以下を満たすことである：

⁵このところがわかりにくい人は、(4.2.4) を

$$W_t^{(N)} \approx \frac{1}{\sqrt{N}} S_n, \quad t \approx \frac{n}{N}$$

と書いてみるとよい

⁶添え字で書くと縮小したら小さすぎて見えにくいことは承知しているが、 $B_t(\omega)$ と書きたくなることを見越して、敢えて添え字を使う

1. (出発点と連続性) 確率 1 で, $B_0 = 0$ であり (原点から出発), B_t は t の関数として連続 (つまり, B_t の path は連続)
2. (各時刻での分布) X_t の分布は平均値 0, 分散が t の正規分布である.
3. (独立性) 任意の $n \geq 1$ と任意の時刻の列 $0 = t_0 < t_1 < t_2 \dots < t_n$ をとったとき, $B_{t_1} - B_0, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$ は互いに独立である.
4. (定常性) 任意の $0 \leq s < t$ に対して $B_t - B_s$ の分布は $t - s$ だけで決まる.

しつこいかも知れないが, 確率過程 $\{B_t\}_{t \geq 0}$ の全体が定義されているような共通の確率空間 (Ω, \mathcal{F}, P) を背後に考えており, 「確率 1 で」と言うのもこの共通の確率空間での確率を言っている.

註 4.3.2 実は 4 番目の定常性は上の 3 つから導かれるのだが (深い理由はないけど) 付けておいた.

註 4.3.3 上のようにブラウン運動を定義すると, これから自然にブラウン運動の paths の分布が定義される (ランダムウォークのときの同様の事情を思い出せ). ブラウン運動の paths の分布を表す測度を Wiener measure と言う. このようにして作った Wiener measure は本質的に一通りであることがわかっている.

註 4.3.4 上で「定義」したような確率過程が存在するのか, (Ω, \mathcal{F}, P) がとれるのか, また, Wiener measure は存在するのか, は全く自明なことではない. 自分に都合の良い条件をどんどん付けていくと, その条件全てを満たすものは存在しなくなってしまう可能性も十分にある! ブラウン運動と Wiener measure の存在 (構成) はかなり高度な問題で, これについては後の方で少しだけ触れる予定. ここではともかく, このような測度が存在すると仮定して話を進めていく.

註 4.3.5 (少し先走り) 「ランダムウォークの連続極限としてのブラウン運動」と言うのは, 数学的に言うと,

今まで作ってきた $\{W_t^{(N)}\}$ の分布が, B_t の分布に, 法則収束する

と要約できる. これについては後の小節である程度詳しく説明する. その過程で Wiener measure そのものをも構成して, 「定義 4.3.1 を満たすものが実際に存在する」ことを言う. なお, ランダムウォークにもいろいろなものがある (例えば, 隣の隣へ跳ぶようなもの) が, ある程度の条件を満たしていれば「ランダムウォークの連続極限としてのブラウン運動」は (その分散の大きさだけ調節すれば) 本質的に一意に定まることもわかっている.

4.3.1 ランダムウォークの連続極限としての, 定義 4.3.1 の理解

ブラウン運動の性質を調べていく前に, 定義 4.3.1 の諸性質が, 元になっているランダムウォークの性質から自然にでてくることを見ておこう.

1. ランダムウォークの出発点はいつも原点だから, B_t の出発点も原点なのはまあ, 自然である. 連続性の方は一見, 自然に見えるが, 「連続関数の極限が不連続」な例はいくらでもあるか

ら, 自然と言ってしまうのは問題だ (実際, Wiener measure の存在証明の一つの山場はこの連続性を示すことである.) 別の言い方をすると, このような連続極限をとっても path の連続性が保たれているのが驚異, とも言える.

2. 元のランダムウォークにもどると, イメージとしては

$$B_t \approx W_t^{(N)} \approx \frac{1}{\sqrt{N}} S_{Nt} \quad (4.3.1)$$

であったが, この右辺は正に「中心極限定理」を使ってくれ, と言わんばかりの形である. $E[S_{Nt}] = 0$ および $\text{Var}[S_{Nt}] = Nt$ を考えに入れると, (4.3.1) が分散 t の正規分布になるのは非常に自然.

3. 元々のランダムウォークに戻すと,

$$B_{t_i} \approx W_{t_i}^{(N)} \approx \frac{1}{\sqrt{N}} S_{Nt_i} \quad (4.3.2)$$

と対応しているから, 結局

$$B_{t_{i+1}} - B_{t_i} \approx \frac{1}{\sqrt{N}} (S_{Nt_{i+1}} - S_{Nt_i}) = \frac{1}{\sqrt{N}} \sum_{j=Nt_i+1}^{Nt_{i+1}} X_j \quad (4.3.3)$$

となっている. つまり, $B_{t_{i+1}} - B_{t_i}$ と $B_{t_{k+1}} - B_{t_k}$ は, $k \neq i$ なら元々のランダムウォークでは互いに独立な X_j 達に相当している. だから当然, 独立であろう.

4. 上と同じように考えると,

$$B_t - B_s \approx \frac{1}{\sqrt{N}} \sum_{j=Ns+1}^{Nt} X_j \quad (4.3.4)$$

となるが, X_j が独立・同分布だったから, 「右辺は X_j が何個あるか」つまり $Nt - Ns$ のみで決まる.

と言うわけで, 1 の連続性以外は, 定義 4.3.1 の諸性質は非常に自然 (ランダムウォークの極限として) だと言えることがわかった.

ただし, 自然にでてくるからと言って, そのような性質を満たすものが存在するかどうかはわからない. また, 定義 4.3.1 に書いた性質だけで定義したいものが一意に決まるかどうかもわからない. 逆に言うと, 定義 4.3.1 に書いた性質を満たすものが存在し, かつ本質的に一意に定まる, のはやはり驚異である.

確率論 I, 確率論概論 I 第 6 回レポート問題

今回の問題は、「ブラウン運動」と言う名前に負けずに落ち着いて考えれば、決して難しくない。ブラウン運動の定義にはいろいろな流儀がある。そのような定義の同等性について考えてみよう。

あ、ヤバイ！今ミスプリに気づいた。この前の定義 4.3.1 の条件の 2, 「 X_t の分布は平均値 0 ...」となっているのは、「 B_t の分布は平均値 0 ...」の間違いです。ヤヤコシイので、もう一回載せておく。

定義 4.3.1. [再録] ある確率空間 (Ω, \mathcal{F}, P) 上で定義された実数値確率過程 $\{B_t\}_{t \geq 0}$ が一次元ブラウン運動である、とは、 B_t が以下を満たすことである：

- (i) (出発点と連続性) 確率 1 で、 $B_0 = 0$ であり (原点から出発)、 B_t は t の関数として連続 (つまり、 B_t の path は連続)
- (ii) (各時刻での分布) B_t の分布は平均値 0, 分散が t の正規分布である。
- (iii) (独立増分性) 任意の $n \geq 1$ と任意の時刻の列 $0 = t_0 < t_1 < t_2 \dots < t_n$ をとったとき、 $B_{t_1} - B_0, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$ は互いに独立である。
- (iv) (定常性) 任意の $0 \leq s < t$ に対して $B_t - B_s$ の分布は $t - s$ だけで決まる。

(しつこいけども、上の (iv) は他の 3 つからでる。)

更に、問 15 のために、別の定義も与えておく。

定義 4.3.6 ある確率空間 (Ω, \mathcal{F}, P) 上で定義された実数値確率過程 $\{B_t\}_{t \geq 0}$ が一次元ブラウン運動である、とは、 B_t が以下を満たすことである：

- (i) (出発点と連続性) 確率 1 で、 $B_0 = 0$ であり (原点から出発)、 B_t は t の関数として連続 (つまり、 B_t の path は連続)
- (iii) (独立増分性) 任意の $n \geq 1$ と任意の時刻の列 $0 = t_0 < t_1 < t_2 \dots < t_n$ をとったとき、 $B_{t_1} - B_0, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$ は互いに独立である。
- (v) (定常性 + α) 任意の $0 \leq s, t$ に対して $B_{s+t} - B_s$ の分布は平均 0, 分散 t の正規分布に従う。

問 14： 定義 4.3.1 では (i) から (iv) の性質を仮定しているが、この内の (iv) (定常性) は他から導かれる、と既に言っている。これを確かめよう。

1. 定義 4.3.1 の (ii) と (iii) を仮定する。このとき、 $s, t > 0$ に対して確率変数 $D_t = B_{s+t} - B_s$ を考えると、 D_t 自身は平均 0, 分散が t の正規分布に従うことを示せ (ヒント: B_t や D_t などの特性関数を考えよ。)
2. このことから、定義 4.3.1 の (iv) の性質が、それ以外から導かれることを確かめよ。

問 15： 問 14 と同じノリで、定義 4.3.1 と定義 4.3.6 の同等性を確かめよ (要するに、他の条件の下では条件 (ii) と (v) を入れ替えられることを示せばよい。)

締め切りなど:

締め切りは 2002 年 7 月 15 日 (月) の 15:00 (曜日と時刻の変更に注意!),

提出場所は僕の部屋 (理 1-508) の前の封筒かポスト

用紙はできうる限り A4 の紙を用いる (B5 などの小さい紙は紛れてなくなるかも)

とします.

(いつも通りの) レポートのお約束:

- 友達と相談しても, 本を調べても, 何をやっても良いから, 自分で理解した範囲を書くこと. その際, 参考文献や議論した友達の名前も明記すること (友達と議論したり, 本を見たからと言って悪い点をつける, などと言うことは絶対にしない. 一番大事なのは自分でわかったところを表現することだから, それまでの過程で何をやっても問題ない.)
- なお, 問題の番外編として, 今までの講義内容・講義形態についての感想, 不満, 文句, このように改善すべしとの意見などもできるだけ書いてください. お願いします.

(レポート問題に関連しての補足). なお, ブラウン運動を規定する条件は更に言い換えることが可能なようである.

- B_t や $B_{t+s} - B_s$ が正規分布になること (上の条件 (ii) や (v)) を仮定する代わりに, 独立増分性 (iii) と定常性 (iv) (および $B_{t+s} - B_s$ の期待値や分散) を仮定する手もある. これは直感的にはわかりやすい. と言うのは, 独立増分性を何回も使うと, $B_{s+t} - B_s$ 自身をより細かい時間間隔での独立な増分の和として書ける. 時間間隔をゼロにした極限を考えると, 中心極限定理から $B_{s+t} - B_s$ が正規分布に従うことが結論できる (ここもレポートにしようかと思ったけど, きちんとした問題にするのが大変そうだったので, やめた. 根性のある人は挑戦して見てくだされ.)

—以下, レジユメの続き—

4.4 ブラウン運動の paths の性質

(この節の内容には, 長田先生, 服部先生の講義ノートを参考にさせていただきました. 深く感謝します. ただし, 僕の講義に現れるであろうミスは, 勿論, 僕が全責任を負うべきものです.)

ブラウン運動が本当に数学的に定義できるのか (定義 4.3.1 を満たす確率過程は存在するのか) はひとまずおいておいて, 定義 4.3.1 を満たす確率過程はどのような性質を持っているのか (特に, ブラウン運動の path はどのような性質を持っているのか) を考えていくことにする. 以下ではいくつかの定理を述べるが, その証明は総じて大変なので, 後の方の小小節で分割して述べる. まずは証明に気を取られず, 全体像を掴んで欲しい.

ランダムウォークの例 (数値計算によるグラフ) から, ブラウン運動の path はかなりギザギザのものらしい, と予想できよう.

それをもっと見るには, 以下の事実が役に立つ.

定理 4.4.1 (diffusive scaling) 任意に $c > 0$ を決める. $\{B_t\}_{t \geq 0}$ をブラウン運動とし, 新しい確率過程 B'_t を

$$B'_t(\omega) \equiv \frac{1}{\sqrt{c}} B_{ct}(\omega) \quad (4.4.1)$$

によって定義すると, $\{B'_t\}_{t \geq 0}$ もブラウン運動になっている.

ブラウン運動は(大体)一意に決まる, ことを認めると, この定理は, ブラウン運動の paths を時間方向に $1/c^2$, 空間方向に $1/c$ 倍した新しい B'_t が, もとの B_t と「同じ分布」に従うことを意味している. このことからブラウン運動の path は非常にギザギザで, 傾きが無限大のところが無数にありそうなのが以下のようにしてわかる.

いま, 元の $B_t(\omega)$ のなかで, 傾きが 1 くらいの所があったとしよう. $c \gg 1$ とすると, B'_t の中で対応する箇所の傾きは, $\frac{1}{\sqrt{c}} \times c = \sqrt{c}$ 倍になっている. c は任意だから, c を十分大きくとることで, 対応する傾きをいくらでも大きくすることができる. ところが, B'_t が, もとの B_t と「同じ分布」に従っているのだから, もとの $B_t(\omega)$ 達のなかにも, もともと「傾きが無限大」のところがあったと隠れていたということだ.

次の 2 つの定理は, 上の予想 (path がギザギザなこと) を数学的に表現したものである.

定理 4.4.2 B_t を 1 次元ブラウン運動とすると, ほとんど全ての $\omega \in \Omega$ に対し, t の関数としての $B_t(\omega)$ は, どの $t \geq 0$ でも Lipschitz 連続ではない.

註 4.4.3 関数 $f(t)$ が区間 I で Lipschitz 連続である, とは, 区間 I で決まる適当な定数 C があって, 任意の $s, t \in I$ に対して

$$|f(s) - f(t)| \leq C |s - t| \quad (4.4.2)$$

となることを言う (つまり, 後ででてくる Hölder 連続性で $r = 1$ とした場合).

註 4.4.4 この定理と次の定理では, ω を一つ決めたときのサンプル $B_t(\omega)$ を問題にしている. この決められた $B_t(\omega)$ が全ての t で微分不可能だと言っているのである.

定理 4.4.5 B_t を 1 次元ブラウン運動とすると, ほとんど全ての $\omega \in \Omega$ に対し, t の関数としての $B_t(\omega)$ は, どの t においても微分不可能である.

このようにブラウン運動の path はかなり「ギザギザ」なんであるが, 無茶苦茶ギザギザではない. 以下の定理がそれを示す.

定理 4.4.6 B_t を 1 次元ブラウン運動とし, $0 < r < 1/2$ なる r を任意に固定する. 更に $T > 0$ を勝手にとる. すると, ほとんど全ての $\omega \in \Omega$ に対し, t の関数としての $B_t(\omega)$ は, $t \geq 0$ で r -Hölder 連続である. より詳しく書くと,

$$|B_t(\omega) - B_s(\omega)| \leq C |t - s|^r \quad \forall s, t \in [0, T] \quad (4.4.3)$$

がなりたつ. ここで C は ω, T に依存するかも知れない有限の定数である.

註 4.4.7 (言葉の註) 関数 $f(t)$ が区間 I にて, 指数 r で Hölder 連続 (または r -Hölder 連続) である, とは以下が成り立つこと:

$$\text{適当な定数 } C \text{ でもって, すべての } s, t \in I \text{ に対して } |f(s) - f(t)| \leq C|s - t|^r .$$

実は, 上の定理 4.4.6 よりも強く, 以下が成り立つ. これは, $\frac{1}{2}$ -Hölder 連続性に少しだけ及ばない結果である.

定理 4.4.8 (Lévy) B_t を 1 次元ブラウン運動とし, $c > 1$ を固定する. ほとんど全ての $\omega \in \Omega$ に対し, ω に依存するだろう $\delta > 0$ が存在して,

$$|B_t(\omega) - B_s(\omega)| \leq c \left\{ 2|t - s| \log |t - s|^{-1} \right\}^{1/2} \quad \forall |s - t| < \delta \quad (s, t \in [0, 1]) \quad (4.4.4)$$

がなりたつ.

最後に, 両者の中間とも言える定理を紹介しよう. まず準備として, 関数 $f(t)$ が与えられたとき, f の区間 $[0, T]$ 上での 2 次変分 (quadratic variation) を以下で定義する. まず, 区間 $[0, T]$ を任意に分割し:

$$0 = t_0 < t_1 < t_2 < \dots < t_{n-1} < t_n \quad (4.4.5)$$

この分割を Δ と書くことにする. また,

$$|\Delta| \equiv \max_i |t_i - t_{i-1}| \quad (4.4.6)$$

を 分割の幅 と呼ぶ. この分割 Δ に関する f の 2 次変分を

$$V_\Delta(f) \equiv \sum_{i=1}^n [f(t_i) - f(t_{i-1})]^2 \quad (4.4.7)$$

と定義する.

定理 4.4.9 区間 $[0, T]$ の分割の列 $\Delta_1, \Delta_2, \dots$ で, その幅が

$$\sum_{n=1}^{\infty} |\Delta_n| < \infty \quad (4.4.8)$$

を満たすようなものを考える. B_t を 1 次元ブラウン運動として, sample 毎にその $[0, T]$ 上での 2 次変分 $V_{\Delta_n}(B \cdot(\omega))$ を考えると, ほとんど全ての $\omega \in \Omega$ に対し,

$$\lim_{n \rightarrow \infty} V_{\Delta_n}(B \cdot(\omega)) = T \quad (4.4.9)$$

がなりたつ.

4.4.1 準備: Borel-Cantelli の補題

実は以下の補題はこの講義では使わないことにしようと思っていたのだが, これ無しではどうにもつらいので, 使うことにした. 非常に標準的な定理であるから, 証明は省略する. (この補題を使えば, 第 2 章の議論なども少し簡単になる部分があるのだが ...)

まず, 事象の列 $\{A_n\}$ が与えられたとき,

$$\limsup_{n \rightarrow \infty} A_n \equiv \lim_{m \rightarrow \infty} \bigcup_{n \geq m} A_n = \{\omega \in \Omega \mid \omega \text{ は無限に多くの } A_n \text{ に入っている}\} \quad (4.4.10)$$

$$\liminf_{n \rightarrow \infty} A_n \equiv \lim_{m \rightarrow \infty} \bigcap_{n \geq m} A_n = \{\omega \in \Omega \mid \omega \text{ は十分大きな } m \text{ より先の全ての } A_n \text{ に入っている}\} \quad (4.4.11)$$

と定義する⁷. すると:

補題 4.4.10 (Borel-Cantelli) A_n を事象の列とする.

(i) $\sum_{n=1}^{\infty} P[A_n] < \infty$ ならば,

$$P\left[\limsup_{n \rightarrow \infty} A_n\right] = 0, \quad P\left[\liminf_{n \rightarrow \infty} A_n^c\right] = 1 \quad (4.4.12)$$

である.

(ii) $\sum_{n=1}^{\infty} P[A_n] = \infty$, かつ, A_n が互いに独立 ならば,

$$P\left[\limsup_{n \rightarrow \infty} A_n\right] = 1, \quad P\left[\liminf_{n \rightarrow \infty} A_n^c\right] = 0 \quad (4.4.13)$$

である.

4.4.2 定理 4.4.1 の証明

これは簡単. 要するに定義 4.3.1 の条件を一つ一つチェックすれば良い.

1. $B_0 = 0$ なら, $B'_0 = 0$ だ. c が正で有限の限りは, $B_t(\omega)$ が連続なら $B'_t(\omega)$ も連続だ.

2. 正規分布の式に代入して計算すると, B'_t も分散が t の正規分布に従うことがわかる (各自, チェック).

3, 4. $c > 0$ の限り, 時間の順序は変わらないので, $0 = t_0 < t_1 < t_2 < \dots$ なら $0 = ct_0 < ct_1 < ct_2 < \dots$ である. 従って, B'_t の独立性や定常性が保証される. \square

一番の山場は 2 の条件を満たすように時間と空間をスケールするところだった. 他はほとんどアタリマエやったね.

4.4.3 定理 4.4.2 の証明

$I = [0, 1]$ の区間について証明する. ここができれば, 平行移動不変性から, 他の区間でも同様. 証明はまず, Lipschitz 連続になりそうな点の集合を作って, その集合が測度 0 であることを言う方向で進む.

⁷集合 (事象) に対しての \limsup, \liminf であるが, この由来は, それぞれの事象の indicator が,

$$\limsup_{n \rightarrow \infty} I[A_n] = I[\limsup_{n \rightarrow \infty} A_n], \quad \liminf_{n \rightarrow \infty} I[A_n] = I[\liminf_{n \rightarrow \infty} A_n]$$

を満たしていることによる. なお, $\limsup_{n \rightarrow \infty} A_n$ のことを A_n i.o. と書くことが多い (i.o. は infinitely often の略)

Step 1. 正の整数 n, M に対して,

$$U_n^M \equiv \left\{ \omega \in \Omega \mid I \text{ 内に } t \text{ が存在して } |t-s| < \frac{2}{n} \text{ なる全ての } u \text{ に対して} \right. \\ \left. |B_t(\omega) - B_s(\omega)| \leq M|t-s| \text{ が成り立つ} \right\} \quad (4.4.14)$$

と定義する. つまり, U_n^M は $|t-s| < 2/n$ の時に Lipschitz 条件が定数 M で成り立ってしまう sample ω の全体. 以下, この U_n^M の測度がゼロであることを目指す.

Step 2. U_n^M の定義には I 内の全ての時刻がでてくるので扱いにくい. これを回避するため, I を n 等分して $t_i = i/n$ ($i = 0, 1, 2, \dots, n$) とおき, 分割の幅を $\Delta = 1/n$ と書く. この第 2 ステップの目的は, U_n^M を, t_i 達だけの値で決まる別の集合 V_n^M で押さえ込むことである. 具体的には,

$$Y_{n,k}(\omega) \equiv \max_{j=0,1,2} \left\{ |B_{t_{k+j}}(\omega) - B_{t_{k-1+j}}(\omega)| \right\} \quad (4.4.15)$$

を定義し,

$$V_n^M \equiv \left\{ \omega \in \Omega \mid \exists k, Y_{n,k}(\omega) \leq \frac{4M}{n} \right\} \quad (4.4.16)$$

を導入する. これは, 隣り合った 3 つの区間の両端での B_t の差の最大値が $4M/n$ より小さくなるような ω を集めてきたもの. すると,

$$U_n^M \subset V_n^M, \quad \text{よって} \quad P[U_n^M] \leq P[V_n^M] \quad (4.4.17)$$

である.

(理由) $\omega \in U_n^M$ ならば, $t \in I$ が存在して $|t-s| < 2/n$ である限り $|B_t(\omega) - B_s(\omega)| < M|t-s|$ であった. そこでこの t を挟むように t_k, t_{k+1} をとってやると, $j = -1, 0, 1, 2$ に対して $|t_{k+j} - t| < 2/n$ が成り立つ. あとは三角不等式で,

$$|B_{t_{k+j}} - B_{t_{k+j-1}}| \leq |B_{t_{k+j}} - B_t| + |B_t - B_{t_{k+j-1}}| \leq M|t_{k+j} - t| + M|t - t_{k+j-1}| \leq \frac{4M}{n} \quad (4.4.18)$$

と評価すればよい.

Step 3. 以下, $P[V_n^M]$ の評価に入る. V_n^M は「ある k が存在する」という条件で決められていたので,

$$P[V_n^M] = P\left[\bigcup_k \left\{ Y_{n,k} \leq \frac{4M}{n} \right\} \right] \leq \sum_{k=1}^{n-2} P\left[Y_{n,k} \leq \frac{4M}{n} \right] = (n-2) P\left[Y_{n,1} \leq \frac{4M}{n} \right] \quad (4.4.19)$$

と評価できる(後ろの等式には平行移動不変性を用いた). $Y_{n,1}$ の定義は $B_{t_j} - B_{t_{j-1}}$ ($j = 1, 2, 3$) の 3 つの差に依存しているが, この 3 つの差は独立・同分布であり, かつこの 3 つ全てが $4M/n$ より小さいことが必要なので,

$$= (n-2) \left\{ P\left[B_{1/n} \leq \frac{4M}{n} \right] \right\}^3 \quad (4.4.20)$$

ここまで来れば, 単に一つの時刻 $t = 1/n$ での分布の問題なので, 陽に正規分布を用いて計算できる. 分散が V の正規分布は

$$P[a < Z < b] = \frac{1}{\sqrt{2\pi V}} \int_a^b \exp\left(-\frac{x^2}{2V}\right) dx \quad (4.4.21)$$

であるから⁸,

$$P\left[B_{1/n} \leq \frac{4M}{n}\right] = \sqrt{\frac{n}{2\pi}} \int_{-4M/n}^{4M/n} \exp\left[-\frac{n}{2}x^2\right] dx \stackrel{\text{変数変換}}{=} \frac{1}{\sqrt{2\pi}} \int_{-4M/\sqrt{n}}^{4M/\sqrt{n}} \exp\left[-\frac{y^2}{2}\right] dy \quad (4.4.22)$$

と計算できる. 最後の積分は $n \rightarrow \infty$ でゼロに行くことはすぐに見て取れる (積分範囲がゼロになるから). 従って,

$$\lim_{n \rightarrow \infty} P[U_n^M] \leq \lim_{n \rightarrow \infty} P[V_n^M] = 0 \quad (4.4.23)$$

が結論できた.

4. 以上を $P[U_n^M]$ の言葉に翻訳しよう. U_n^M を規定している条件をよく見ると, U_n^M は n と共に増加する (少なくとも非減少である) ことがわかる. 従って

$$P[U_\infty] = P\left[\bigcup_{n=1}^{\infty} U_n\right] = \lim_{n \rightarrow \infty} P[U_n^M] = 0 \quad (4.4.24)$$

が成り立つ. 従って, M についても極限をとって,

$$P\left[\bigcup_{M=1}^{\infty} U_\infty^M\right] = 0, \quad (4.4.25)$$

すなわち, $B_t(\omega)$ が Lipschitz 連続であるような ω の集合 U_∞ の測度がゼロであることがわかり, 定理は証明された. \square

4.4.4 定理 4.4.5 の証明

定理 4.4.2 を使うと簡単である. 一言で言うと, 微分可能であれば Lipschitz 連続であるから, Lipschitz 連続でさえない関数が微分できることはあり得ない.

4.4.5 定理 4.4.6 と定理 4.4.8 の証明

定理 4.4.6 は定理 4.4.8 が証明できれば良い. 問題は定理 4.4.8 であるが ...

4.4.6 定理 4.4.9 の証明

Step 1. 以下の不等式が成り立つことをまず, 示そう. 任意の分割 Δ に対して,

$$E\left[\left(V_\Delta(B_t) - T\right)^2\right] \leq C |\Delta| T, \quad C \equiv \int_{-\infty}^{\infty} (z^2 - 1)^2 e^{-z^2/2} \frac{dz}{\sqrt{2\pi}} \quad (4.4.26)$$

がなりたつ.

そのために, V_Δ の定義を用いて計算する. 定義から (Δ が N 分割するものだととして)

$$V_\Delta(B_t(\omega)) - T = \sum_{i=1}^N \{B_{t_i}(\omega) - B_{t_{i-1}}(\omega)\}^2 - T = \sum_{i=1}^N \left[\{B_{t_i}(\omega) - B_{t_{i-1}}(\omega)\}^2 - (t_i - t_{i-1}) \right] \quad (4.4.27)$$

⁸以下の式はブラウン運動のスケーリング (定理) を用いて,

$$P\left[B_{1/n} \leq \frac{4M}{n}\right] = P\left[B_1 \leq \frac{4M}{\sqrt{n}}\right]$$

を経由した方が楽だし, 見通しも良いが, 敢えて愚直に定義通りやってみた

と書くことはいつでもできる．さて，右辺にでている和の各項は互いに独立，かつそれぞれの期待値は丁度ゼロである．したがって，(4.4.27) の両辺の 2 乗の期待値をとると，

$$E\left[\left(V_{\Delta}(B_t) - T\right)^2\right] = \sum_{i=1}^N E\left[\left(\{B_{t_i} - B_{t_{i-1}}\}^2 - (t_i - t_{i-1})\right)^2\right] \quad (4.4.28)$$

を得る．ところが，右辺の確率変数 $B_{t_i} - B_{t_{i-1}}$ は期待値 0，分散が $t_i - t_{i-1}$ の正規分布に従う．ので，具体的に

$$\begin{aligned} E\left[\left(\{B_t - B_s\}^2 - (t - s)\right)^2\right] &= \frac{1}{\sqrt{2(t-s)\pi}} \int_{-\infty}^{\infty} \{x^2 - (t-s)\}^2 \exp\left(-\frac{x^2}{2(t-s)}\right) dx \\ &= (t-s)^2 \int_{-\infty}^{\infty} (z^2 - 1)^2 e^{-z^2/2} \frac{dz}{\sqrt{2\pi}} = C(t-s)^2 \end{aligned} \quad (4.4.29)$$

と計算できる．従って，(4.4.28) から

$$E\left[\left(V_{\Delta}(B_t) - T\right)^2\right] = \sum_{i=1}^N C(t_i - t_{i-1})^2 \leq \sum_{i=1}^N C|\Delta|(t_i - t_{i-1}) = C|\Delta|T \quad (4.4.30)$$

となって，(4.4.26) が証明された．

Step 2. 以上で， $V_{\Delta}(B_t) - T$ の分散についての評価が得られたので (今まで何回もやったように) チェビシェフの不等式を用いると，任意の正の ℓ に対して

$$P\left[\left|V_{\Delta}(B_t) - T\right| > \frac{1}{\ell}\right] \leq C|\Delta|T\ell^2 \quad (4.4.31)$$

が得られる．これを $\sum_{n=1}^{\infty} |\Delta_n| < \infty$ を満たす Δ_n について用いると，

$$\sum_{n=1}^{\infty} P\left[\left|V_{\Delta}(B_t) - T\right| > \frac{1}{\ell}\right] \leq \sum_{n=1}^{\infty} C|\Delta_n|T\ell^2 < \infty \quad (4.4.32)$$

が結論できる．従って，Borel-Cantelli の補題 (i) から，

$$P\left[\liminf_n \left\{\left|V_{\Delta}(B_t) - T\right| > \frac{1}{\ell}\right\}^c\right] = 1 \quad (4.4.33)$$

となる．この ℓ についての積事象を考えても，

$$P\left[\bigcap_{\ell=1}^{\infty} \left(\liminf_n \left\{\left|V_{\Delta}(B_t) - T\right| > \frac{1}{\ell}\right\}^c\right)\right] = 1 \quad (4.4.34)$$

が成り立つ．これは「任意の $\ell > 0$ に対して (ω に依存するかも知れない) ある n から先では $\left|V_{\Delta}(B_t(\omega)) - T\right| > \frac{1}{\ell}$ が成り立たない」ことを意味する．逆に言うと (ω に依存するかも知れない) ある n から先では

$$\left|V_{\Delta}(B_t(\omega)) - T\right| \leq \frac{1}{\ell} \quad (4.4.35)$$

になってしまうのである． ℓ は任意に大きくとれるから，これは (確率 1 で)

$$\left|V_{\Delta}(B_t) - T\right| = 0 \quad (4.4.36)$$

であることを意味する．

□

(以下は7月16日に渡したプリントの改訂版である。正直なところ、学期の終わりで疲れが出てきたのか、もう終わった講義のプリント作りに力が入らず、結局あれから余り進歩しなかった。でも最後のほう (tightness) は少し整理したので、興味のある人は読んでみてください。)

7月16日の連絡:

(1) 宣言通り、来週(7月23日)は期末テストです。場所、時間は通常の講義と同じ。なお、以下の要領で持ち込みを認めます。これを機会に勉強してください。

- 持ち込めるものは A4 の大きさの紙の片側に自筆で書いたもの一枚のみ。裏側まで書いたり、コピーしたりしたものは無効である。
- 持ち込んだ紙には名前と学生番号を書いて、答案用紙と共に提出すること (持ち込み無しで試験を受けようと言う人も、白紙に名前と学生番号を書いて提出してください。)

(2) 期末テストの時に、「夏休み特別レポート」を同時に出題するかもしれません。

(3) 修士課程の学生さんは以下の要領で「講義内容要約」を提出してください。4年生の方は提出する義務はありませんが、提出されたものには目を通します。

提出期限: 2002年9月20日(金)一応、午後5時ごろを×切。

提出先: 原の部屋の前の封筒

提出部数: 3部

(以下は教務委員会からの文言を再録したもの)

講義内容要約は「教官がどのような講義をおこなったか」ではなく、「自分が何についてどのような理解・展望に達したか」という視点から講義全体の要点を再構成して書くこと。重要と思われる点、自分が興味を感じた点を強調して書くのも一つの方法である。また、それは講義を受講していない第三者が読んで理解できる形に書くこと。単に項目を箇条書きにただけのもの、あるいは数式だけが連なったものは論外である。一定水準に達しない報告書に対しては再提出を要求する。各科目とも2ページ程度を目安とする。

4.5 ランダムウォークからブラウン運動へ II

さて、この講義の締めくくりとして、ランダムウォークからブラウン運動を構成して Wiener measure の存在を証明する、その粗筋を紹介しよう。これからの部分はまあ「お話」と思って聞いてくれればよい。この小節と次の小節では [Bill-I, p.xx] と言うのは P. Billingsley, *Convergence of Probability Measures* の第一版 (1968) の xx ページへの参照である。同様に、[Bill-II, p.xx] と言うのは第二版 (1999) への参照である。

少し思い出しておく。 S_n をランダムウォークの n 歩目の位置として、これらを折れ線をつないだ $\{\tilde{S}_s\}_{s \geq 0}$ 、および N を大きな正の整数として、

$$W_t^{(N)} \equiv \frac{1}{\sqrt{N}} \tilde{S}_{Nt} \quad (4.5.1)$$

を定義した．それで， $N \rightarrow \infty$ で $W_t^{(N)}$ の集まりが「ブラウン運動に収束」することを示したいわけである．

なお，技術的な困難を少しでも減らすために，構成するブラウン運動 B_t は $0 \leq t \leq 1$ の所だけ考えることにする．ここを作れば，あとはこれを同じ分布に従うものを一杯持ってきて， $1 \leq t \leq 2, 2 \leq t \leq 3, 3 \leq t \leq 4$ とつないでいけば全ての $t \geq 0$ で構成できる．

まず，この収束がどのような意味のものかをはっきりさせるところから始める．区間 $[0, 1]$ 上で定義された連続関数の全体を $C[0, 1]$ と書く（要するに path の空間）．この空間の元 $f, g \in C[0, 1]$ に対してその距離を

$$d(f, g) \equiv \max_{0 \leq t \leq 1} |f(t) - g(t)| \quad (4.5.2)$$

と定義する．この距離を元にして「開集合」が定義され，この開集合族によって $C[0, 1]$ は位相空間になる．さらにこの位相空間の Borel 集合族（全ての開集合を含む最小の σ 加法族）を考えて可測空間とする．これを (S, \mathcal{S}) と書くことにする．

こうすると，上で定義した $W_t^{(N)}$ も， B_t もこの可測空間 (S, \mathcal{S}) 上での測度を定めることになる．この準備の下で， $C[0, 1]$ 上の確率変数 ($W_t^{(N)}$ や B_t を念頭においている) の法則収束を以下のように定める．

定義 4.5.1 $C[0, 1]$ 上の確率変数の列 $X_t^{(N)}$ ($N = 1, 2, 3, \dots$) が確率変数 Y_t に法則収束するとは，任意の $C[0, 1] \rightarrow \mathbb{R}$ の有界な連続関数⁹ F に対して¹⁰，

$$\lim_{N \rightarrow \infty} \langle F(\{X_t^{(N)}\}) \rangle = \langle F(\{Y_t\}) \rangle \quad (4.5.3)$$

が成り立つことを言う．ここで左右両辺の期待値は，それぞれの確率変数 ($X_t^{(N)}, Y_t$) の分布 P_N, P についてのものである．なお，このとき測度について P_N は P に弱収束すると言う．

註 4.5.2 以前に法則収束を「累積分布関数が収束すること」として定義した．しかし，今考えている $W_t^{(N)}$ や B_t は $C[0, 1]$ の元であり，すなわち曲線（またはグラフ）である．こんなものの分布関数など，なかなかうまく定義できない（無限次元空間の中の分布だから）．そこで分布関数そのものを見る代わりに，「確率変数の関数の期待値が収束する」ことを要求することにした（実数値確率変数についての両方の定義の同等性は以前に少しだけ言ったよね...）

定理 4.5.3 上のように定義した $W_t^{(N)}$ ($0 \leq t \leq 1$) は，ブラウン運動 B_t ($0 \leq t \leq 1$) に，定義 4.5.1 の意味で法則収束する．

4.5.1 大まかな方針

証明の大まかな方針は以下の通りである．

⁹位相空間 X から Y への関数 f が連続とは， Y の任意の開集合 V に対し，その逆像 $f^{-1}(V)$ が X での開集合になっていること

¹⁰しつこい註: F は $C[0, 1]$ の関数をつつ決める毎に F の値（実数値）が決まるような関数であるから，標語的には「関数の関数」になっている． F が連続というときに使う「開集合」は勿論，距離 $d(f, g)$ で決まるものを用いる

まず注意すべきは、今考えている確率変数 $W_t^{(N)}$ や B_t はサンプルを一つ決めた毎に $[0, 1]$ 区間上の連続関数 (path) になっていることである。つまり、この path を決めるには $[0, 1]$ 上の全ての t に対する $W_t^{(N)}$ や B_t の値を決める必要がある。だから、path の分布を問題にするのにも、「 $[0, 1]$ 上の全ての t に対する $W_t^{(N)}$ や B_t の分布」を見る必要があるだろう。これは今までに考えてきた問題よりも格段に難しい (今までは例えば S_n や Z_n など、たった一つの数の分布や収束を問題にしていた。今は無限個の数の分布や収束を扱わないといけなくなった!)

と考えると、ちょっと気が遠くなるのだが、諦めるのはマダ早い。確かに、path を完全に決定するには全ての $t \in [0, 1]$ における値を決定する必要がある。しかし、有限個の点を見ても、ある程度のことは言えないか? 例えば、全ての $t \in [0, 1]$ は大変だけど、1以上の整数 n を決めて、 $t = i/n$ ($i = 0, 1, 2, \dots, n$) の所での値だけを見たらどうだろう? 少なくとも必要条件としては、これらの有限個の時刻での値の分布が収束して行くことが必要になりそうだ。問題は逆が言えるかと言うことである。

定理の形にすると、以下のような流れになる。まず、有限次元分布を定義する。

定義 4.5.4 (と言うほど大げさなものではないが) $C[0, 1]$ 上の確率変数 X_t に対し、その有限次元分布とは 任意の $n \geq 1$ と $0 \leq t_1 < t_2 < \dots < t_n \leq 1$ に対する $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ の分布のことである。

しつこいけども、 $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ とは、もともとの X_t のグラフを $t_1 < t_2 < \dots < t_n$ で切った切り口を見ている感じ。

さて、 $X^{(N)}$ が Y に法則収束するならば、 $X^{(N)}$ の有限次元分布が Y の有限次元分布に法則収束することは (ほとんど) 明らかである。我々の使いたいのはこの逆であるが、これは一般には保証されない、つまり、 $X^{(N)}$ の任意の有限次元分布が何かの分布に収束したからと言って、その極限を有限次元分布に持つような確率変数 Y があるとは限らない。

この点を補うには、もっと条件を付け加えないといけない。その代表的なものが tightness と呼ばれる概念である。

定義 4.5.5 (tightness) 測度空間 S, S 上の確率測度の列 $\{P_N\}$ がある。この測度の列が tight であるとは、以下が成り立つことと定義する：

任意の $\epsilon > 0$ に対して S の compact な部分集合 K が存在して、全ての N に対して $P_N(K) > 1 - \epsilon$ が成り立つ。

上の定義では、全ての N に共通に K がとれる必要がある。

勿論、上の定義の P_N としては、 $X_t^{(N)}$ の分布を表すものを想定している。

さて、以上の準備の下で、我々に有用な定理は以下のようになる。

定理 4.5.6 (Bill-II, Theorem 7.1 または Bill-I, Theorem 8.1) (S, S) 上の確率測度の列 $\{P_n\}$ がある。 $\{P_n\}$ が何かの測度 P に弱収束するための十分条件の一つは以下の通り：

- $\{P_n\}$ の有限次元分布が P の有限次元分布に法則収束し,
- かつ, $\{P_n\}$ が定義 4.5.5 の意味で tight である.

要するに(良くわからんけども) tightness + 有限次元分布の収束 が十分条件だと言うわけやね.

註 4.5.7 上の定理の書き方はちょっといい加減である. この定理(と類似の議論)の凄いところは, 有限次元分布と tightness だけで, 行き先の P の存在まで言ってしまうこと(そうでなければ Wiener 測度の構成には使えない). 上の定理の書き方は何となく P の存在を仮定しているように見えるが, それは誤りである. なお, 上では Billingsley の定理そのものを引用しているが, むしろ地の文 (Bill-II p.88) のように議論すべし.

4.5.2 有限次元分布の収束について

Tightness は大変なので, まずは有限次元分布の収束を見ておこう. ほとんど明らかであるが...

まず, 一次元分布を見ることにしよう. B_t の一次元分布(要するに B_t そのものの分布)は平均値 0, 分散が t の正規分布であるから, $W_t^{(N)}$ の分布が ($N \rightarrow \infty$ の極限で) 平均値 0, 分散が t の正規分布に法則収束することを言えばよい. でも, $W_t^{(N)}$ の作り方(スケールの仕方)を思い出すと, これはほとんど¹¹単なる中心極限定理であり, 簡単に証明できる.

次に 2 次元分布であるが, この場合 B_s と B_t を直接みるよりも, B_s と $B_t - B_s$ を見た方が楽である(これらは独立だから). 対応する $W^{(N)}$ の部分も独立であるので, やはり中心極限定理から, 望み通りの法則収束が言える.

3 次元以上の分布も全く同様.

4.5.3 Tightness の十分条件 について

上でブラウン運動の構成方法の概要を説明したが, そこで一番難しいのは tightness を証明することである. この小節では tightness を保証する十分条件の一つを与えて説明することにする. 基本的なものは以下の定理であろう. 以下に出てくる $W_t^{(N)}$ は, いままで考えてきたもの (4.5.1) である.

定理 4.5.8 [Bill-I, Theorem 8.4]

$W_t^{(N)}$ ($N = 1, 2, 3, \dots$) が tight である十分条件の一つは以下の通り: 任意の $\epsilon > 0$ に対して, (大きな) 定数 $\lambda > 1$ と $N_0 > 0$ が存在して, 全ての $N > N_0$ と $n \geq 0$ で

$$P \left[\max_{0 \leq i \leq N} |S_{n+i} - S_n| \geq \lambda \sqrt{N} \right] \leq \frac{\epsilon}{\lambda^2} \quad (4.5.4)$$

が成り立つこと.

この定理は証明抜きで引用する. 興味のある人は Billingsley の本の該当部分を参照されたい. 更に, 上の定理を更に使いやすくしたものとして, 以下の補題もある.

¹¹ 「ほとんど」とわざわざ書いたのは, t が $1/N$ の整数倍でない場合, $W_t^{(N)}$ はランダムウォーク上の 2 点をつなぐ折れ線上にあるので, 少しだけ補正が必要だから

補題 4.5.9 $\{W_t^{(N)}\}_{t \geq 0}$ ($N = 1, 2, 3, \dots$) が tight である十分条件の一つは, 以下の通りである: 定数 $K > 0$ と $a > 1/2$, および K, a に依存するかも知れない $N_0 > 0$ が存在して, 任意の $0 \leq n_1 < n_2$ と全ての $N \geq N_0$ に対して

$$E \left[|S_{n_2} - S_{n_1}|^{4a} \right] \leq K (n_2 - n_1)^{2a} \quad (4.5.5)$$

が成り立つこと.

註 4.5.10 実は上の定理と補題¹²は, 我々の考えているよりも広い状況の下で成り立つ. すなわち, $W^{(N)}$ を定義する元の $\{S_n\}_{n \geq 0}$ は, 単に \mathbb{R} または \mathbb{R}^d に値をとる確率変数であればよい (でも, あまり話を広げると混乱するから, 以下ではいままでどおり, 独立な確率変数の和の場合に限定する.)

以下では定理 4.5.8 を認めた上で (1) 補題 4.5.9 の証明 (2) 我々のブラウン運動の構成が補題 4.5.9 を満たすこと (従って $W^{(N)}$ は tight であり, ブラウン運動が構成できること) を

4.5.4 定理 4.5.8 を認めた上で, 補題 4.5.9 の証明

まず, ランダムウォークから作った $W^{(N)}$ の場合, 元になる S_n は定常性を満たしている. つまり, $S_{n+i} - S_n$ の分布は, 単なる S_i の分布と同じである. 従って, (4.5.4) の条件は,

$$P \left[\max_{0 \leq i \leq N} |S_i| \geq \lambda \sqrt{N} \right] \leq \frac{\epsilon}{\lambda^2} \quad (4.5.6)$$

と同じことである. 以下, 補題 4.5.9 の条件の下で (4.5.6) を示す [大枠は Bill-I, p. 63 による]. まず,

$$\max_{0 \leq i \leq N} |S_i| \geq \lambda \sqrt{N} \quad (4.5.7)$$

であるためには,

$$|S_N| \geq \frac{\lambda \sqrt{N}}{2} \quad (4.5.8)$$

または

$$|S_N| \leq \frac{\lambda \sqrt{N}}{2} \quad \text{かつ} \quad |S_1|, |S_2|, \dots, |S_{N-1}| \text{ のどれかが } \lambda \sqrt{N} \text{ より大} \quad (4.5.9)$$

である必要がある. (4.5.9) は更に, どれが $\frac{\lambda \sqrt{N}}{2}$ より大きいかで分類するつもりになると,

$$(4.5.9) \subset \bigcup_{i=1}^{N-1} \left\{ |S_N| \leq \frac{\lambda \sqrt{N}}{2} \quad \text{かつ} \quad |S_i| \geq \lambda \sqrt{N} \quad \text{かつ} \quad |S_1|, |S_2|, \dots, |S_{i-1}| < \lambda \sqrt{N} \right\} \quad (4.5.10)$$

¹²補題に類似した条件は, 自分の仕事関連でより広い条件の下で 10 年以上前に用いた. しかし, この補題 4.5.9 自身が広い条件の下で成り立つかどうかは, 今回, 改めてチェックする時間がなかった (以下の証明では一カ所, S_n と $S_{n+i} - S_n$ の独立性を用いているところがあるので, そのままでは独立性をはずした状況には広げられない.) と言うわけで, 鵜呑みにしないで下さいね (多分大丈夫だと思うし, 期末テストなどが全て終わればチェックしますが, 今はテスト問題の方が優先事項)

のような事象の和の部分集合になっていることがわかる．よって，

$$P\left[\max_{0 \leq i \leq N} |S_i| \geq \lambda\sqrt{N}\right] \leq P\left[|S_N| \geq \frac{\lambda\sqrt{N}}{2}\right] + \sum_{i=1}^{N-1} P\left[|S_N| \leq \frac{\lambda\sqrt{N}}{2} \text{ かつ } |S_i| \geq \lambda\sqrt{N} \text{ かつ } |S_1|, |S_2|, \dots, |S_{i-1}| < \lambda\sqrt{N}\right] \quad (4.5.11)$$

が得られた．さて，(4.5.5) から，チェビシエフの不等式の証明と同じようにして

$$P\left[|S_N| \geq \frac{\lambda\sqrt{N}}{2}\right] = P\left[|S_N|^{4a} \geq \left(\frac{\lambda\sqrt{N}}{2}\right)^{4a}\right] \leq \left(\frac{\lambda\sqrt{N}}{2}\right)^{-4a} \times E\left[|S_N|^{4a}\right] \leq \left(\frac{\lambda\sqrt{N}}{2}\right)^{-4a} \times KN^{2a} = \frac{K'}{\lambda^{4a}} \quad (4.5.12)$$

となる ($K' = K2^{4a}$ は定数) ．

また， $|S_N| < \frac{\lambda\sqrt{N}}{2}$ かつ $|S_i| \geq \lambda\sqrt{N}$ ならば $|S_N - S_i| > \frac{\lambda\sqrt{N}}{2}$ であるから，

$$P\left[|S_N| \leq \frac{\lambda\sqrt{N}}{2} \text{ かつ } |S_i| \geq \lambda\sqrt{N} \text{ かつ } |S_1|, |S_2|, \dots, |S_{i-1}| < \lambda\sqrt{N}\right] \leq P\left[|S_N - S_i| > \frac{\lambda\sqrt{N}}{2} \text{ かつ } |S_i| \geq \lambda\sqrt{N} \text{ かつ } |S_1|, |S_2|, \dots, |S_{i-1}| < \lambda\sqrt{N}\right] = P\left[|S_N - S_i| > \frac{\lambda\sqrt{N}}{2}\right] \times P\left[|S_i| \geq \lambda\sqrt{N} \text{ かつ } |S_1|, |S_2|, \dots, |S_{i-1}| < \lambda\sqrt{N}\right] \quad (4.5.13)$$

も成立する (最後の所では S_i までと $S_N - S_i$ が独立なことを用いて確率を分けた) ．やはり (4.5.5) から，チェビシエフの不等式の証明と同じようにして

$$P\left[|S_N - S_i| > \frac{\lambda\sqrt{N}}{2}\right] \leq \left(\frac{\lambda\sqrt{N}}{2}\right)^{-4a} \times K(N-i)^{2a} = \frac{K'}{\lambda^{4a}} \frac{N-i}{N^{4a}} \leq \frac{K'}{\lambda^{4a}} \quad (4.5.14)$$

が成り立つので，

$$(4.5.11) \leq \frac{K'}{\lambda^{4a}} \left(1 + \sum_{i=1}^{N-1} P\left[|S_i| \geq \lambda\sqrt{N} \text{ かつ } |S_1|, |S_2|, \dots, |S_{i-1}| < \lambda\sqrt{N}\right]\right) \quad (4.5.15)$$

が得られた．

さて，(4.5.15) の右辺に出ている事象は，「 S_1, S_2, \dots, S_{N-1} のどれかが $\lambda\sqrt{N}$ より大きい」と言う事象を， $\lambda\sqrt{N}$ より大きい $|S_i|$ で分解して書いたものである．だから，

$$\sum_{i=1}^{N-1} P\left[|S_i| \geq \lambda\sqrt{N} \text{ かつ } |S_1|, |S_2|, \dots, |S_{i-1}| < \lambda\sqrt{N}\right] \leq 1 \quad (4.5.16)$$

が満たされており，最終的に

$$P\left[\max_{0 \leq i \leq N} |S_i| \geq \lambda\sqrt{N}\right] \leq \frac{K'}{\lambda^{4a}} = \frac{K'}{\lambda^{4a-2} \lambda^2} \quad (4.5.17)$$

が得られた．補題の仮定から $4a > 2$ であるので， ϵ に応じて λ を十分に大きくとると (4.5.17) の右辺を ϵ/λ^2 より小さくすることができ，定理 4.5.8 の条件が満たされていることがわかった．□

4.5.5 ランダムウォークから作った $W^{(N)}$ が tight であることの確認

$a = 1$ ととった場合の補題 4.5.9 を確かめれることにする.

まず, 元になる S_n は定常性を満たしており, $S_{n+i} - S_n$ の分布は, 単なる S_i の分布と同じである. 従って, (4.5.5) の条件は ($a = 1$ ととる),

$$E\left[|S_n|^4\right] \leq K n^2 \quad (4.5.18)$$

と同じことである. 以下, (4.5.18) を示す.

そこで, (4.5.18) の左辺を計算するのであるが, ランダムウォークの場合, $S_n = X_1 + X_2 + X_3 + \dots + X_n$ と書いたときの X_j が独立・同分布であるから,

$$\begin{aligned} E\left[|S_n|^4\right] &= E\left[\left(\sum_{j=1}^n X_j\right)^4\right] = \sum_{0 < i, j, k, l \leq n} E\left[X_i X_j X_k X_l\right] \\ &= \sum_{i=1}^n E\left[(X_i)^4\right] + 3 \sum_{i \neq j} E\left[(X_i)^2 (X_j)^2\right] \\ &\leq n_2 \langle X_1^4 \rangle + (n_2)^2 \langle X_1^2 \rangle^2 \leq K (n_2)^2, \quad (K \equiv \langle X_1^2 \rangle^2 + \langle X_1^4 \rangle) \end{aligned} \quad (4.5.19)$$

と計算できて, 確かに条件 (4.5.18) を満たしていることがわかった. □