shinshu.tex, 服部哲弥
信州大学集中講義 1998/7/14–17, 90 分 5 回
19980527–31;0601–04;09–16;19–21

1　準備．(火曜日１)

- 確率論の初等的・基礎的な考え方をいくつか紹介します.
- 興味はあるけれども，知識のあまりない人 (高校の確率論程度) にさらに興味を持ってもらうこと (大学中級程度まで) が目標です.
- 一般論は事実の列挙にとどめ，それを応用した計算例を示します．系統だった知識や目新しい話はありません.
  (1) 切符売り場の行列は一列にすべきでしょうか？(分散)
  (2) さいころの目が５種類出そろうのに平均何回投げればよいでしょうか？(期待値の線形性)
  (3) 公式 $\displaystyle\lim_{n\to\infty}\left(\sum_{k=n}^{\infty}\frac{n^k x^k}{k!}\right)^{1/n}=e\cdot x$ を導けますか？ (極限定理)
  (4) 「文豪ザル」は TOKYOTO と KYOTOFU どちらがお好き？(マルチンゲール)
  (5) すごろくのコマの動きを遠くから眺めるとどう見えるでしょう？(確率過程)
  (6) 酔っぱらいの足どり， Poisson 過程と Wiener 過程，マルコフ性
- この講義のレベルの教科書：
  (1) 佐藤坦，初めての確率論　測度から確率へ，共立出版，1994.
  (2) 西尾真喜子，確率論，実教出版，1978.
  (3) 小谷真一，測度と確率１，２，岩波講座現代数学の基礎，1997.
  (4) フェラー，確率論とその応用 (I 上, 下,II 上, 下)，紀伊国屋書店,
  (5) D. Williams, Probability with martingales, Cambridge University Press, 1991.
  (6) R. Durrett, Probability: Theory and Examples, Duxbury Press, 1991.

  そのさらに先をのぞき見たい人のための日本語の入門的解説書：
  (1) 舟木直久，確率微分方程式，岩波講座現代数学の基礎，1997.
  (2) 楠岡成雄，確率と確率過程，岩波講座応用数学，1993.

  ある比較的新しい応用から見た確率論の日本語の入門的解説書：
  (1) Ya. G. Sinai, 確率論入門コース，シュプリンガー東京，1995.
  (2) 樋口保成，パーコレーション　ちょっと変わった確率論入門，遊星社，1992.

1.1　例：切符売場の行列．−確率論の見方は常識で見落としやすい−

私の知る限り国内，外国を問わずどこでも，スーパーでレジが複数あるときはレジごとに別々に並ぶ．電車や娯楽施設の自動券売機も機械ごとに別々に並ぶことが多いように思う．1980 年代までの日本では，たいていのところで，一つ一つの窓口や電話ボックスなどの前に別々に並んでいた．これに対して，現在では，銀行のキャッシュコーナーや飛行機のチェックインカウンターなどを中心に，窓口が複数ある場所で，それぞれの窓口の前に別々に並ぶのではなく，一列に並んで空いた窓口を順に利用する「一列待ち」が行われている場合がある（別紙図参照）.

---

0 例題の配置

		LDP
	V[·], E[·]	極限定理
martingale	ABRACADABRA, MSD	LIL, $T_Z=1$

1980 年代後半だったと思うが,「一列待ち」の導入が話題になったことがあって，しばらくは「一列待ち」がいろいろなところで試みられた．銀行のキャッシュコーナー，特にスペースのゆったり取れる利用客が多い支店，では「一列待ち」が定着したように思うが，それ以外のところでは必ずしも定着しなかった．

ＪＲのＹ駅の指定券前売り窓口は，昔は「並列待ち」だったのを 1990 年代の初めに「一列待ち」に変えた．ところがＹ駅は一旦始めた「一列待ち」をまもなくやめて「並列待ち」に戻した．意見箱を利用して理由を問い合わせたところ，Ｙ駅の意見は

(1) 平均時間は変わらない,
(2) 列が長く見えてＹ駅は混んでいると思われてしまう,
(3) 誘導人員が確保できない,

ということであった．第２点は心理学的な問題，第３点は経済的な問題，が中心だが，第１点は平均値という数学的（しかも確率論的）な問題である.「平均が変わらないから一列待ちには（数学的にも）意味がない」という主張にみえる．

しかし，これは適切な理由とは言えない．Ｙ駅がこれを理由にあげたということは，確率論のものの見方が常識として定着していないことを示す一つの典型例である．

待ち時間の平均は変わらなくても,「一列待ち」には意味がある

ことを論じよう．

最初に，平均値に関するＹ駅の主張は正しいことを確認しておく．

分かりやすくするため，かつ，議論に曖昧さが出ないように，いくつか条件をおく[1]．

(1) 窓口の数を定数 M とし，待っている人数を N，即ち自分が N 人目（自分が到着する直前 $N-1$ 人が待っていた）とする．窓口 M に比べて人数 N がいつも十分大きいとする．特に，窓口に空き時間はない（全ての窓口がいつも仕事をしている）とする．
(2) 客 i $(i = 1, 2, \cdots, N)$ の処理時間 S_i は確率変数であり，$\{S_i\}$ は独立（客の間に相関はない）とする[2]．窓口は全て同じ処理能力で，各々の客の処理時間の分布は等しい（見ただけでは時間がかかりそうかどうか区別できない）とする．特に平均処理時間 $\tau = \mathrm{E}[\,S_i\,]$ は客によらず一定である．実際に測定してもらえれば直ちに分かるが，客によって券の購入に必要な時間はばらばらである．このことを S_i が確率変数であるとしてとらえる[3]．確率変数 $X_1,\,X_2,\,\cdots$, が独立とは，$\mathrm{Prob}[\,X_1 \in A_1,\ X_2 \in A_2,\ \cdots\,] = \prod_i \mathrm{Prob}[\,X_i \in A_i\,]$ ということ．
(3) 「一列待ち」と「並列ならび」は並び方以外の条件は変わらないとする．例えば，どの窓口があいたか判断して窓口まで歩くのに要する時間は無視する．

上記条件から従うことだが,「並列待ち」の場合どの窓口も均等に並ぶとする．どの窓口が早いか分からないときは人々は自然に均等に並ぶ傾向がある．即ち「並列待ち」は M 個の窓口それぞれに N/M 人ずつ客が並び，ある窓口が空けばその窓口に並んでいた客が順に処理を受け,「一列待ち」は N 人の客が長い一列を作り，M 個の窓口のどれかが空き次第，順に客が空いた窓口へ行って処理を受ける[4]．

さて，一人の客の処理をするのに平均 τ [5] の処理時間を要するとする $(\tau = \mathrm{E}[\,S_i\,])$．自分が N 人目だとすると，自分の分が終わるまでに窓口がしなければならない仕事の総量（処理時間の合計，のべ時間）は平均 $N\tau$ である．窓口が一つしかかなければ平均 $N\tau$ だけの時間がかかるという意味.「並列待ち」でも「一列待ち」でも常時 M 個の窓口が常に処理を続けているから,

---

[1] 自然科学では，このことを「本質を損なわない範囲で簡単なモデルを考える」と表現する．

[2] 確率変数，独立，平均値，などの概念は確率論の概念で，きちんとした定義があるが，それは次回にふれることにして，ここでは，高校で習った程度の意味で理解しておけばよい．

[3] 現実に処理時間は分布しているか？1993 年初め：窓口 A:3,4, B:2.5,2, C:8 (分)．分布している．

[4] 別紙図ではでこぼこしているが，このようなことは考えない．本題とは関係ないが，レジのようにうまい下手があり，また，時間がかかるかどうか買い物かごの中身からある程度予測がつくときは，ある程度でこぼこに並んだ方がいい場合がある．私の観測では，毎日回に来る人々は驚くほど正確に，待ち時間が一番短い列に並ぶ．この問題についてはここでは扱わない．

[5] タウ，ギリシャ文字の一つ． t や T は時刻を表すために頻繁に使われるので，文字が足りなくなると τ もよく使われる．

自分の番が終わるまでの自分の平均待ち時間 t_1 は $t_1 = N\tau/M$ となり，並び方によらない．即ち平均値に関してはＹ駅の主張は正しい．

しかし,「一列待ち」と「並列待ち」は実際に体験してみると,（体験するまでもなく）たしかに「何かが違う」．どこに違いが出るかを示すために，次の点に注目する．

前売り券を買いに来る人は，次の約束までの時間や仕事の休み時間を利用して窓口に来るか，あるいは，乗る直前に来て発車までの時間に前売り券を買いたい，と考えている.（前売り券を買うためにわざわざ一日中並ぶ覚悟，というのは特殊な場合であろう.）そこで，電車の発車までにあるいは次の約束・用事までに前売り券が買えるかという問題を考える．次の用事までに t_0 の時間的余裕があるとき，それが平均待ち時間 t_1 に比べて長ければ $(t_0 > t_1)$，間に合いそうだと判断して，前売り券を買うために並ぶ.（t_0 と t_1 の大小は判断できることは仮定する.）実際の待ち時間 T は平均 t_1 の周りにばらつく確率変数である．たまたま前の人が長引いて結局買えない場合もあるし，無事買えることもある．$T < t_0$ が実現すれば券を買えるが，窓口が自分より前の客に手間取って $T > t_0$ となってしまったならば待っている間に時間切れとなって券を買わずに次の仕事や約束に向かわなければならない.（発車に間に合わなかった場合は計画を変更しないといけない.）確率 $\mathrm{Prob}[\,T > t_0\,]$ で前売り券を買い損なうので，この確率が小さいほど「望ましい窓口」ということになる．図 2.3 のように，平均待ち時間 t_1 が同じでも待ち時間 T の分布の形によって「前の客に手間取ったための不運な買い損ない」$\mathrm{Prob}[\,T > t_0\,]$ の大きさが変わりうる．この値は，客の処理時間 S_i の分布の具体形が分からないと計算できないが，目安として分布の **分散** を考えることができる．図 2.3 の 2 つの分布を比べると，分布が広がっているほど $\mathrm{Prob}[\,T > t_0\,]$ は大きい（但し $t_0 > t_1$）．分布の広がりの目安が分散であるから，分散が小さいほど買い損ないが少なく望ましい，と想像される[6]．

「並列待ち」と「一列待ち」の待ち時間をそれぞれ $T^{(1)}$ および $T^{(2)}$ と置くと,

$$T^{(1)} = \sum_{i=1}^{N/M} S_{j_i}\,,\ \text{（並列待ちの待ち時間）},$$
$$T^{(2)} = \frac{1}{M}\sum_{i=1}^{N} S_i\,,\ \text{（一列待ちの待ち時間）},$$

と書ける.（S_i は客 i の処理時間．$j_1, j_2, \cdots, j_{N/M}$, は並列並びについて自分と同じ列に並ぶ客.）既に見たように，$\mathrm{E}[\,T^{(1)}\,] = \mathrm{E}[\,T^{(2)}\,] = t_1$，即ち待ち時間の期待値（平均値）は等しい．

客の処理時間の分布は独立でかつ客によらない（独立同分布）という仮定から，客一人当たり窓口当たりの処理時間の分散 $\mathrm{V}[\,S_i\,]$ は客によらず一定なので，これを $\mathrm{V}[\,S_i\,] = \sigma^2$ とおく．σ はシグマと読む．これもギリシャ文字で，標準偏差（分散の平方根）などを表すのに良く用いる．

分散についての次の公式を利用する．

(1) $\{S_i\}$ が独立ならば，$\mathrm{V}[\,S_1 + S_2 + \cdots + S_n\,] = \mathrm{V}[\,S_1\,] + \mathrm{V}[\,S_2\,] + \cdots + \mathrm{V}[\,S_n\,]$.
(2) a が定数ならば（確率変数でないならば）$\mathrm{V}[\,a\,X\,] = a^2\,\mathrm{V}[\,X\,]$.

二つ目の公式で右辺 a の 2 乗が出ることが鍵となるが，この公式の導出は別の機会にまわす．これらの公式と $T^{(1)}, T^{(2)}$ の定義から

$$\mathrm{V}[\,T^{(1)}\,] = \frac{N}{M}\,\sigma^2\,,\ \text{（並列待ちの場合）},$$
$$\mathrm{V}[\,T^{(2)}\,] = \frac{N}{M^2}\,\sigma^2\,,\ \text{（一列待ちの場合）}.$$

[6] この想像は一般的には成り立たない．分布の形が違いすぎると $\mathrm{Prob}[\,T > t_0\,]$ の大小と分散の大小が一致しなくなる場合がある．本当は $\mathrm{Prob}[\,T > t_0\,]$ を比較すべきである．今回は計算のしやすい分散を用いて話をする．正規分布などの具体的な分布を仮定すれば $\mathrm{Prob}[\,T > t_0\,]$ も計算可能になる．

$\mathrm{V}[\,T^{(2)}\,] = \dfrac{N}{M^2}\,\sigma^2$ を証明しておこう．分散についての二つの公式を順に用いて

$$\begin{aligned}\mathrm{V}[\,T^{(2)}\,] &= \mathrm{V}[\,M^{-1}\,S_1 + M^{-1}\,S_2 + \cdots + M^{-1}\,S_N\,]\\ &= \mathrm{V}[\,M^{-1}\,S_1\,] + \mathrm{V}[\,M^{-1}\,S_2\,] + \cdots + \mathrm{V}[\,M^{-1}\,S_N\,]\\ &= M^{-2}\,(\mathrm{V}[\,S_1\,] + \cdots + \mathrm{V}[\,S_N\,])\\ &= \frac{N}{M^2}\sigma^2\,.\end{aligned}$$

たしかに成り立つ．

$M > 1$ だと $\mathrm{V}[\,T^{(1)}\,] > \mathrm{V}[\,T^{(2)}\,]$ となるから，図 2.3 で言えば，並列待ちが分散の大きい左側の図，一列待ちが右側の図に対応する．一列待ちの方が並列待ちに比べて，待ち時間（並び始めてから買えるまでの時間）の分散，即ち散らばり具合が小さい．それ故一列待ちのほうが処理し損なう確率（前売り券を買い損なう確率）が低いと思われる.（実際この結論は正しい.）

従って次の約束までの余裕（並んでいられる時間）が限られているときに，買い損なう恐れが小さい．

安易な直感は誤りに陥る場合があること，確率論には（他の自然科学同様に），根本まで立ち返って考えることによって正しい結論を目指す役割があること，の一例として一列待ちの数学的意味を考えた．

確率論は，曖昧な言い方をすれば「ばらつく量」を扱う数学である．数学ということは，矛盾なく推論及び計算ができるということである．ばらつく量と言ったのは，応用上はいろんなケースを念頭に置いている．

- 測定実験では測定のたびに値が異なるのが普通．正しい値は一つのはずだが，制御できない撹乱要因に由来する誤差のために，値が散らばる．
- 天気予報の場合は，観測地点の数などの観測能力の限界から雨の量などの予測値が不確実である．これも予測できる精度より細かいことは制御できないばらつきとみなす．この場合の制御とは限られた予算の中での最善の予測ということ．
- 賭事の場合はルーレットの回り具合やトランプの切り具合など全てを測定しては賭事にならないだろう．社会的な約束事のために要因をわざわざ制御しないケースである．
- 偏差値や工場で生産する製品の性能の場合は，個別に詳しく調べれば違いがはっきりしていても，一人に先生を一人つけたり，一つ一つの製品を手で丁寧に作っていたりするのは経済的に不可能なので，可能な範囲で制御して，より細かい個別の違いは制御不能なばらつきと考える．この場合の制御とは，大学をどれくらい作るか，や，工場で製造機械の精度や工場内の環境をどれくらい厳しく管理するかなど．そうみたときの能力の分布をはかるのが偏差値である．能力にばらつきがなければ能力測定型の入学試験は無意味になる．分布が入学試験という社会現象の原因になっている．

実験や天気予報では値が散らばることは望ましくないので，測定装置を良くするなど制御能力を増す努力が行われる．しかし，誤差は決してなくならない．賭事で百発百中は無理な相談だ．そこで，不規則性を不可避として，なおかつ規則性を見い出そうというのが確率論の役割になる．

一見直感に反することを正しいと言うことや正しい直感を示すことは簡単ではなく，一見直感の利かない準備を必要とする．

今回いくつかの用語や公式を定義や証明をせずに用いた．これについて次に触れよう．

問題．

ＪＲのＹ駅の指定券前売り窓口では一旦開始した「一列待ち」をまもなくやめた．問い合わせに対する回答によると，やめた理由は

(1) 平均時間は変わらない,
(2) 列が長く見えてＹ駅は混んでいると思われてしまう,
(3) 誘導人員が確保できない,

ということであった.

問１. 第１点については講義でみたように適切な理由とは言えない．この観点から,「一列待ち」に戻すよう駅の担当者を説得する文書を作れ.

問２. 第２点は前売り券購入客の理解が必要である．この点もふまえて，ＪＲは「一列待ち」について客の理解を得るためにどうＰＲすればよいか，ＪＲの広報担当官の立場になって，広告説明文を提案せよ.

駅の担当者及び前売り券購入客（一般の人たち）は，確率論の専門家ではないし，この講義を聞いてもいない．その前提で分かりやすく，しかし誤解のないように説明せよ.

1.2 確率空間.

不規則性（量のばらつき）が当然に持っていると考えられている性質を矛盾なく導くための数学が**確率論**である．矛盾なく，かつ，統一的に導くために，**測度論**と呼ばれる数学の上に確率論を定義する.（測度論が数学として確立しているので，それに基づいて確率論を定義すれば，矛盾なく種々の量が計算できることが保証される.）

1.2.1 測度論としての確率論.

数学的には確率論の基礎の部分は測度論（積分論）そのものである.

確率論		測度論	定義
全事象	Ω	全体集合	
事象	$A \in \mathcal{F}$	可測集合	$\mathcal{F}$ は σ 加法族
確率	P	$P(\Omega)=1$ なる測度	$P: \mathcal{F} \to [0,1]$, σ 加法性
確率変数	X	可測関数	$X:\ \Omega \to \mathbf{R}$; $\{X \le a\} \in \mathcal{F}$ [7] , $a \in \mathbf{R}$
期待値	$\mathrm{E}[\ \cdot\ ]$	積分	$\mathrm{E}[\ X\ ] = \int_\Omega X(\omega)\, P(d\omega)$
分散	$\mathrm{V}[\ \cdot\ ]$		$\mathrm{V}[\ X\ ] = \mathrm{E}[\ (X - \mathrm{E}[\ X\ ])^2\ ]$

さいころや硬貨投げに関連する確率の計算は高校時代までにも習ったかも知れない．直感通りと思った諸君もいるかも知れないが，それは正しい．さいころや硬貨投げのように確率変数の取りうる値（さいころの目や硬貨の表裏で決まる値）が有限個しかない場合は，確率の計算は直感通りであると言ってほぼよい．基本的には

ある事象の起きる確率 ＝ その事象の起こる場合の数 ÷ 全ての場合の数

で全て計算できる．場合の数が有限だから,「数えればいい」のである.

取りうる値が無数にある場合は気を付けないといけない. 取りうる値が無数にある簡単な分布の例として区間 $[0,1]$ 上の**一様分布**があるこの分布は A を区間 $[0,1]=\{x \mid 0 \le x \le 1\}$ の部分集合（事象）とするとき，事象 A が起きる確率 $P(A)$ が区間 A の長さに等しいものを言う.

長さ 1cm の線分を針でランダム（でたらめ）に刺したとき，刺した場所が左端から何 cm にあるかを測ってその値を X としたとき，X の分布が一様分布である．例えば $\mathrm{Prob}[\ 0 \le X \le 0.1\ ] = 0.1$ などとなる．$0 \le x \le 1$ を満たす任意の実数 x を値として取りうるので，取りうる値に無数の可能性がある.

一様分布を例にとって，安易な計算 (?) が矛盾を引き起こすことを示そう．$[0,1]$ の上の一様分布 P では１点 x $(0 \le x \le 1)$ が生じる確率 $P(\{x\})$ は，１点の長さは 0 だから， $P(\{x\})=0$ のはずである．点 x と点

[7] 確率論では集合 (事象) の定義 $\{\omega \in \Omega \mid X(\omega) \le a\}$ を書くのに，要素を省略して $\{X \le a\}$ と書くことがある．そのほうが「らしい」ので個人的に好き.

y が異なるならば，事象 $\{x\}$ と事象 $\{y\}$ は**排反事象だから，和事象の確率は個々の事象の確率の和**なので，$P(\{x,y\})=0$ である．これをとことん突き詰めると，$[0,1]$ は $0\leq x\leq 1$ を満たす点の和集合であるから，つぎのように考えたくなる：

$$P(\bigcup_{x\in[0,1]}\{x\})=\sum_{x\in[0,1]}P(\{x\})=\sum_{x\in[0,1]}0=0 \quad (?)$$

ところが，$\bigcup_{x\in[0,1]}\{x\}=[0,1]$ だから，

$$0=P(\bigcup_{x\in[0,1]}\{x\})=P([0,1])=1$$

となり，$0=1$ となって矛盾する．1 点の確率 $P(\{x\})=a>0$ としても同様の理由から $1=\infty$ となってやはり矛盾する．

この矛盾は，排反事象の和事象の確率は個々の事象の確率の和である，という規則を無限個の和事象 $\bigcup_{x\in[0,1]}\{x\}$ に適用するときに論理的限界があることを示している．実数区間の点は $n=1,2,3,\cdots$ のように番号を付けて並べることができない（自然数と対応がつかない）ことが知られている．このことを実数区間の点の**濃度が非可算**であるという．濃度とは個数という言葉を無限個の場合に拡張した概念．確率の定義では**自然数で番号づけられるときに限り**，排反事象の和の確率 = 各事象の確率の和，という公式を正しいと認めている．自然数で番号づけられる「無限個」を（濃度が）**可算**であるという．

$$P(\bigcup_{x\in[0,1]}\{x\})=\sum_{x\in[0,1]}P(\{x\}) \quad (?)$$

は，区間 $[0,1]$ の中の点 x の個数が非可算なので，**許されない**変形である．

現代確率論は抽象的な準備の定義から始まるが，これは有限の場合には素朴な定義に合うように，また，無限の場合にも矛盾が起きないように，そして，矛盾が起きない限りできるだけ広いケースに確率が定義できるように，ぎりぎりまで一般化した結果である．過去の数学者の工夫の結果，現代での決定版となった確率の定義は初めて見ると意味が分かりにくいかもしれないが，人智の一つの到達点として一見に値する．

Ω [8] を集合とする（考察対象の空間）．Ω の集合族（部分集合からいくつか選んだ集まり）$\mathcal{F}$ が σ 加法族（シグマかほうぞく）であるとは以下を満たすことと定義する．

空集合と全体集合． $\emptyset\in\mathcal{F}$, $\Omega\in\mathcal{F}$.

補集合． $A\in\mathcal{F} \ \Rightarrow\ A^c\in\mathcal{F}$,

可算和． $A_i\in\mathcal{F}$, $i=1,2,3,\cdots \ \Rightarrow\ \bigcup_{i=1}^{\infty}A_i\in\mathcal{F}$,

組 $(\Omega,\mathcal{F})$ を可測空間と呼ぶ．

$P:\ \mathcal{F}\to[0,1]$ （$\mathcal{F}$ に属する集合を与えると 0 以上 1 以下の実数値が決まる関数，という意味），

即ち，σ 加法族 $\mathcal{F}$ の上で定義された $[0,1]$ に値をとる関数 P が，$A_i\in\mathcal{F}$, $i=1,2,3,\cdots$, が互いに素（すなわち，$j\neq k$ ならば必ず $A_j\cap A_k=\emptyset$ である）ならば，必ず $P(\bigcup_{n=1}^{\infty}A_n)=\sum_{n=1}^{\infty}P(A_n)$ を満たすとき，P は確率測度であるという．この性質，即ち，排反事象の和の確率 = 各事象の確率の和，が**可算個**の和集合の場合に成立すること，を σ 加法性と呼ぶ．

$\mathcal{F}$ の要素 $A\in\mathcal{F}$ を**事象**，$P(A)$ を事象 A の**確率**という．集合 Ω（全体集合），Ω の部分集合たち (全てとは限らない) を要素とする σ 加法族 $\mathcal{F}$，$\mathcal{F}$ を定義域とする測度 P の 3 つ組 $(\Omega,\mathcal{F},P)$ が与えられた集合 (空

[8] オメガ，ギリシャ文字

間) Ω を**確率空間**と呼ぶ．定義を満たす 3 つ組は何であっても確率空間と呼んで良い，ということ．例えば下記の例のようなものは最も簡単な代表例たちである．

全体集合 Ω とは，制御できない撹乱の結果の全体である．Ω の要素は見本と呼ばれ,「さいころを振ったとき出た個々の目」を表す．一様分布の反例で見たように，Ω が無限集合の時は Ω の各要素（見本）の起きる確率だけでは事象の確率は求められない．そこで，見本（要素）ではなく，事象（部分集合）ごとに確率を定める，と考える．ところが，全ての部分集合に確率を定義しようとすると，無限集合の場合に，確率が持っていてほしい性質に矛盾することが起こりうる（上述の反例より少しこみいってくる）．勝手な部分集合を全て事象とは呼べない（実際に利用する場合もそんなに奇妙な部分集合を考察の対象にはしないので必要がない）．事象の集まりを σ 加法族と呼んで定義し，確率を σ 加法族の上の関数と考えるのはこのような考察に基づく．

確率という言葉の定義はこれだけ（つまり,「どんなさいころについても共通に成り立つ性質」が全て出てくると考えている）である．この定義を満たす限り矛盾は生じないことが分かっている（σ 加法性が**可算個**の和集合の場合にのみ保証されているところがみそ）．できるだけ多くの場合を統一的に扱い，かつ，できるだけ短い言葉で定義する，という一般化と抽象化のために一見「確率らしさ」が見えないかも知れないが,これだけの定義からもいろいろな確率らしい性質が証明できる．

空集合と全体集合． $P(\emptyset) = 0$, $P(\Omega) = 1$.

有限加法性． $A_i \in \mathcal{F}$, $i = 1, 2, 3, \cdots, N$, が互いに素のとき $P(\bigcup_{n=1}^{N} A_n) = \sum_{n=1}^{N} P(A_n)$.

補集合． $P(A^c) + P(A) = 1$.

単調性． $A_1 \subset A_2$ のとき $P(A_1) \leq P(A_2)$，特に，$P(A) \leq 1$.

劣加法性． $P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$．$A_n$ たちが有限個 $(A_n,\ n = 1, 2, \cdots, N)$ のときも同様の性質（有限劣加法性）が成立．

和． $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

単調収束性（確率の連続性）． (1) $A_1 \supset A_2 \supset A_3 \cdots$ のとき $\lim_{n\to\infty} P(A_n) = P(\bigcap_{n=1}^{\infty} A_n)$.

(2) $A_1 \subset A_2 \subset A_3 \cdots$ のとき $\lim_{n\to\infty} P(A_n) = P(\bigcup_{n=1}^{\infty} A_n)$.

問題．

講義中に述べた確率の定義から上記最初の性質 $P(\emptyset) = 0$ を証明せよ.（確率の定義のうち，σ 加法性において，全ての A_n を空集合 $\emptyset$ においてみよ.）

1.2.2　測度空間の例．

具体的な計算をするためには具体的な確率が必要である．確率論の一般論の成果を利用するためには，具体例は以上の定義を満たしている必要がある．素朴な（さいころや硬貨投げのような）確率は上記の定義を満たしているので，今まで（高校までで習った）通り確率を計算してよい．一様分布のように実数（あるいはその区間）の上の確率を考えるときは積分で定義すれば自動的に上記の定義を満たしていることが分かっている．

σ 加法族の例．

Ω を集合とすると，

- $\{\emptyset, \Omega\}$ は σ 加法族である．これが最小の（集合の数が一番少ない）σ 加法族.
- Ω の部分集合を全て集めた集合族を 2^Ω と書くと，2^Ω は σ 加法族．特に $\emptyset \subset \Omega$, $\Omega \subset \Omega$ だから $\emptyset \in 2^{Omega}$, $\Omega \in 2^{Omega}$.

例１．さいころを n 回投げる試行．

$$\Sigma = \{1,2,3,4,5,6\},\ \Omega_n = \Sigma^n,\ \mathcal{F}_n = 2^{\Omega_n},\ A \subset \Omega_n \quad \text{に対して } P(A) = 6^{-n}\sharp A = \frac{\sharp A}{\sharp \Omega},$$

とおくと，$(\Sigma^n, \mathcal{F}_n, P)$ が求める確率空間．例えば $n=3$ はさいころを３回投げる試行を考える場合であり，さらに，$(\omega_1, \omega_2, \omega_3) = (1,4,3)$ はその３回のさいの目が順に $1,4,3$ だったことを表す．そしてそれが起きる確率は $P(\{(1,4,3)\}) = 6^{-n}$.

例２．さいころを永久に投げ続ける試行．

この講義では $\mathbf{N} = \{1,2,3,\cdots\}$ は自然数の集合とする．

$$\Sigma = \{1,2,3,4,5,6\},\ r = \sharp\Sigma = 6,\ \Omega = \{\omega = (\omega_1, \omega_2, \cdots) \mid \omega_i \in \Sigma\ (i \in \mathbf{N})\},$$
$$\mathcal{F}_n = \sigma[\{\{\omega \in \Omega \mid \omega_i = v_i,\ i = 1,2,\cdots,n\} \mid (v_1, \cdots, v_n) \in \Sigma^n\}]\ \mathcal{F} = \sigma[\bigcup_{n \in \mathbf{N}} \mathcal{F}_n],$$

$P : \mathcal{F} \to \quad [0,1]$ を $n \in \mathbf{N}$, $A \in \mathcal{F}_n$ に対して $P(A) = r^{-n}\sharp A$ を満たすように定義された確率測度とする．

集合族 $\mathcal{A}$ に対して $\sigma[\mathcal{A}]$ は $\mathcal{A}$ を含む最小の σ 加法族．任意の有限回の試行についてあらゆる情報の確率が（残りの無限回をみないで判断できる集合ならば）計算できる，ということである．しかし，かってな無限列の集合 $A \subset \Omega$ を持ってきたときには $P(A)$ が計算できるとは限らない．これが確率空間になっている (P が $\mathcal{F}$ 上に確率測度として定義できる) ことを証明するのは次の例同様，若干の工夫を必要とするが，省略する．

例３．一様分布（ルベーグ測度）．

「長さ」という常識的，素朴な概念を厳密化した，という意味ではもっとも基本的な確率空間（測度空間）の例．$\Omega = [0,1]$，$\mathcal{B}$ は Ω の部分閉区間 $[a,b]$ $(0 \le a < b \le 1)$ を含む最小の σ 加法族とする．$P : \mathcal{B} \to [0,1]$ を $P([a,b]) = b - a$ を満たす測度とすると，そのような P は存在し，しかも，以上の条件だけで一つに決まることが知られている.（この事実の証明は若干時間がかかるので，この講義では省略する.）一般に任意の有限区間 $[a,b]$ 上の一様分布も同様に定義できる．

集合族 $\{(-\infty, x] \mid x \in \mathbf{R}$ を含む最小の σ 加法族を（１次元）ボレル集合族と呼び，$\mathcal{B}$ と書くことが多い．σ 加法族の定義から，$\mathcal{B}$ は開区間 (a,b)，閉区間 $[a,b]$，片側開区間 $[a,b)$，さらには $\mathbf{R}$ の開集合，閉集合を全て含むことが分かる．従って上記の閉区間を含む最小の σ 加法族はボレル集合族である．

例４．実数上の分布．

確率の定義は，現代的な積分の一般論（測度論とは実は現代的な積分論の一般論のことである）の定義と本質的に同じである．だから，確率が積分で定義されていれば自動的に確率の定義を満たす.（積分の現代的な定義は次の講義で述べる.）

一般に ρ [9] を，負にならない ($\rho(x) \geq 0$) 関数であって，$\displaystyle\int_\Omega \rho(x)\,dx = 1$ を満たすものとするとき，

$$P(A) = \int_A \rho(x)\,dx$$

で定義される確率を考えることができる．$\int_A$ は積分範囲が集合 A であるという意味．ρ をこの確率の**密度**と呼ぶ．Ω は実数全体や実数の部分集合を考えうる．例えば，区間 $[0,1]$ 上の一様分布は，$P(A)$ が集合 A の長さに等しいと定義したが，これは積分

$$P(A) = \int_A 1\,dx\,, \quad A \subset [0,1],$$

で確率を定義したということと同じである．つまり密度関数 $\rho(x) = 1$, $x \in [0,1]$, で定義される分布である．

問題（カントール集合）．

$[0,1] \supset A_1 \supset A_2 \supset A_3 \cdots$ を満たす実数の集合の列 $A_1, A_2, \cdots,$ を次のように定義する．A_1 は $[0,1]$ を 3 等分して真ん中の区間だけを取り除いたもの，即ち $A_1 = [0,1/3] \cup [2/3,1]$ で，2 つの区間の和集合．A_2 は A_1 の 2 つの区間を各々 3 等分して，各々真ん中の区間だけを取り除いたもの，即ち $A_2 = [0,1/9] \cup [2/9,3/9] \cup [6/9,7/9] \cup [8/9,1]$ で，4 つの区間の和集合．以下同様に，A_n は A_{n-1} の 2^{n-1} 個の（各長さ 3^{-n+1} の）区間を各々 3 等分して，各々真ん中の区間だけを取り除いた，2^n 個の長さ 3^{-n} の区間の和集合とする．A_n の長さを求めよ．

$A = \displaystyle\bigcap_{n=1}^{\infty} A_n$ （全ての A_n に含まれる点の集合）で定義される集合をカントール集合と呼ぶ．カントール集合 A の長さ（区間 $[0,1]$ の上の一様分布に従うときの $P(A)$）を上記列挙した最後の性質（確率の連続性）を利用して求めよ．(A は非可算個の点からなる集合である，つまり,「実数と同じくらい個数が多い」ことが知られている．カントール集合は長さ（ルベーグ測度）0 の非可算集合の簡単な例として有名である．)

1.3 確率変数．

$(\Omega, \mathcal{F}, P)$ を確率空間とし，Ω 上の実数値関数 $X : \Omega \to \mathbf{R}$ を考える．$A \subset \mathbf{R}$ に対して $X^{-1}(A) = \{X \in A\} = \{\omega \in \Omega \mid X(\omega) \in A\}$ で実数上の集合関数 X^{-1} を定義する．実数値関数 $X : \Omega \to \mathbf{R}$ が**確率変数**であるとは，どんな実数 x に対しても $X^{-1}((-\infty, x]) == \{X \leq x\} = \{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$ であることを言う．$\mathcal{F}$ を明示したいときは $\mathcal{F}$–可測関数とも呼ぶ．

通常考えるような関数は全て確率変数になる．つまり，数学的な定義を大雑把に言うと確率変数とは関数のことである．関数をわざわざ確率変数と呼ぶのは，確率論の「気持ち」では，Ω は制御不能な要因の全体を表し，Ω 上の関数は制御不能な要因によって値が変わりうる（不規則性を生じる）量と感じるからである．

既に注意したように，勝手な集合 $A \subset \Omega$ の確率が全て存在するとは限らない．確率空間を定義するとき，確率を定義できる集合，つまり事象，の集まりとして σ 加法族 $\mathcal{F}$ を与える．同様に，勝手な関数 $X : \Omega \to \mathbf{R}$ を全て考慮の対象にできない．そこで，現代確率論では，確率変数の定義にも若干の条件をおく．それが上述の確率変数の定義である．「気持ち」は，最小限の要求として $P(X \leq x)$ が計算できることが期待されている．即ち $\{X \leq x\} \in \mathcal{F}$.

数学（特に解析学）では極限を扱う．確率変数列 X_n, $n = 1, 2, \cdots,$ の極限も無条件に確率変数になるほうが定理がきれいになるので，確率変数の値として $\pm\infty$ を許すのが測度論では普通である．$\pm\infty$ を含む演算についての注意は，

- $-\infty < x < \infty$ が任意の実数 x に対して成り立つとする．

[9] ロー，ギリシャ文字

- 加法や乗法は

$$\infty + x = \infty + \infty = \infty \times |x| = \infty \times \infty = \infty,$$
$$-\infty + x = -\infty + (-\infty) = -\infty \times |x| = \infty \times (-|x|) = -\infty \times \infty = -\infty$$

  など，常識的なものである．注意すべき点は $\infty - \infty$ が**定義されていない**，つまり，そのような演算が出てきてはいけない，という点である．

- $\pm\infty$ を含む以上の演算の定義は，積分における $\pm\infty$ の取り扱いでも同様である．

確率変数についての上記の定義は（$\pm\infty$ を考慮に入れれば）この場合も有効である．即ち，（$\pm\infty$ を許す意味の）実数値関数 $X : \Omega \to \mathbf{R} \cup \{\pm\infty\}$ が確率変数であるとは，どんな実数 x に対しても $X^{-1}([-\infty, x]) = \{-\infty \le X \le x\} \in \mathcal{F}$ であることを言う．

知られている性質．

(1) 確率変数の和，積，差，定数倍，は確率変数である．(商は $1/0$ が定義されないので 0 をとらない，という条件が必要.)

(2) X_n, $n \in \mathbf{N}$, が確率変数の列ならば，$\displaystyle\inf_{n\in\mathbf{N}} X_n$, $\displaystyle\sup_{n\in\mathbf{N}} X_n$, $\displaystyle\liminf_{n\in\mathbf{N}} X_n$, $\displaystyle\limsup_{n\in\mathbf{N}} X_n$ は確率変数である．ここで，

$$\liminf_{n\in\mathbf{N}} X_n = \lim_{n\to\infty} \inf_{k\ge n} X_k = \sup_{n\ge 1} \inf_{k\ge n} X_k$$

は，値として $\pm\infty$ を許せば必ず存在する． $\limsup$ も同様.

(3) よく知られているように，数列の極限は常に存在するとは限らないので，極限 $\displaystyle\lim_{n\to\infty} X_n(\omega)$ は全ての ω で存在するとは限らない．しかし，極限が存在するような ω の集合は可測集合である（従って，X_n が収束する確率 $P(\exists \displaystyle\lim_{n\to\infty} X_n)$ が計算できる）：$\{\omega \in \Omega \mid \exists \displaystyle\lim_{n\to\infty} X_n(\omega)\} \in \mathcal{F}$．全く収束しない関数列を作ることは容易だが，$\emptyset \in \mathcal{F}$ だからその場合も除外されていない．必ず収束する例も同様.

以上の性質の証明は難しくはないが，知らなければ全く明らかとはいえない．

例えば和については

$$\{X + Y \le a\} = \bigcup_{q\in\mathbf{Q}} (\{X \le q\} \cap \{Y \le a - q\}) \in \mathcal{F}\,,$$

（$\mathbf{Q}$ は有理数の集合）だから $X + Y$ も確率変数になる．下限については

$$\{\inf_{n\ge 1} X_n \le a\} = \bigcup_{n\ge 1} \{X_n \le a\} \in \mathcal{F}\,,$$

だから $\displaystyle\inf_n X_n$ も確率変数になる．詳しく言うと，仮定より， X_n は確率変数だからどんな $a \in \mathbf{R}$ に対しても $\{X_n \le a\} \in \mathcal{F}$ となり，$\mathcal{F}$ は σ 加法族なので，可算和に関して閉じていることから，$\displaystyle\bigcup_{n\ge 1} \{X_n \le a\} \in \mathcal{F}$ を得る．最初の変形は下限 inf の定義から従う．

例．さいころを永久に投げ続ける試行．

$$\Sigma = \{1,2,3,4,5,6\},\ r = \sharp\Sigma = 6,\ \Omega = \{\omega = (\omega_1, \omega_2, \cdots) \mid \omega_i \in \Sigma\ (i \in \mathbf{N})\},$$
$$\mathcal{F}_n = \sigma[\{\{\omega \in \Omega \mid \omega_i = v_i,\ i = 1,2,\cdots,n\} \mid (v_1,\cdots,v_n) \in \Sigma^n\}]\ \mathcal{F} = \sigma[\bigcup_{n\in\mathbf{N}} \mathcal{F}_n],$$
$$n \in \mathbf{N},\ A \in \mathcal{F}_n \text{ に対して } P(A) = r^{-n}\sharp A\,.$$

数列の，どれでもいいから n 個の項の値を指定すると，それらの値がその通り生じる確率は r^{-n} になる，という意味である．

任意の自然数 n に対して

$$X_n(\omega) = \begin{cases} 1, & \omega_n = 1, \\ 0, & \omega_n \ne 1, \end{cases}$$

は確率変数である．従ってその和 $S_n = X_1 + X_2 + \cdots + X_n$ も確率変数である．S_n は明らかに n 回さいころを投げたときそのうち 1 の目が出た回数を表す．他の目の出現回数を表す確率変数も同様に定義できる．

問題．

(1) 確率変数についての性質のうち， $\inf_{n\in\mathbf{N}} X_n$ が確率変数になることの証明を参考にして，$\sup_{n\in\mathbf{N}} X_n$ が確率変数になることを証明してみよ．

(2) 確率変数の和が確率変数になることの証明で有理数 $q \in \mathbf{Q}$ について和をとっているが，なぜ実数の範囲の和とせず，わざわざ有理数の集合を考えるのか？

(3) $\{\omega \in \Omega \mid \exists \lim_{n\to\infty} X_n(\omega)\} \in \mathcal{F}$ を証明してみよ.（ヒント：収束の定義からこの集合は

$$\{\exists \lim_{n\to\infty} X_n\} = \{\limsup_{n\in\mathbf{N}} X_n < \infty\} \cap \{\liminf_{n\in\mathbf{N}} X_n > \infty\} \cap \{\limsup_{n\in\mathbf{N}} X_n - \liminf_{n\in\mathbf{N}} X_n = 0\}$$

と書ける.）

(4) さいころを永久に投げ続ける試行の例の X_n が確率変数であることを証明せよ．

(5) X^{-1} が集合算を全て保存すること，特に，$X^{-1}\left(\bigcup_n A_n\right) = \bigcup_n X^{-1}(A_n)$ と $X^{-1}(A^c) = (X^{-1}(A))^c$ を示せ.（次の話題で用いる.）

1.3.1 関数が生成する σ 加法族（省略可能？）．

$\{X_n\}$ について確率が計算できるためには，P の定義域 $\mathcal{F}$ が $\{X_n \le a\}$, $a \in \mathbf{R}$, を含んでいることが必要十分である．そこで，$\{X_n \le a \mid a \in \mathbf{R},\ n \in \mathbf{N}\}$ を含む最小の σ 加法族を $\sigma(X_n,\ n \in \mathbf{N})$ などと書く．$\{X_n\}$ について確率が計算できるためには，最初から $\mathcal{F} = \sigma(X_n,\ n \in \mathbf{N})$ と考えればよい．

この考え方だと，Ω で定義された関数（たち） $X_n : \Omega \to \mathbf{R} \cup \{\pm\infty\}$ に対して σ 加法族 $\mathcal{F} = \sigma(X_n,\ n \in \mathbf{N})$ を定義し，その上で $P : \mathcal{F} \to [0,1]$ を与えることになる．

ボレル集合族　（省略可能？）．

確率変数の定義を言い換えよう．$X : \Omega \to \mathbf{R}$ に対して，上の問題と σ 加法族の定義から確かめられることだが， $\mathcal{B}_X = \{A \subset \mathbf{R} \mid X^{-1}(A) \in \mathcal{F}\}$ は（$\mathbf{R}$ を全体集合とする）σ 加法族になる．言い換えると，X が確率変数であるとは $(-\infty, x] \in \mathcal{B}_X$ が全ての実数 x に対して成り立つことである．

$\mathcal{B}_X$ は σ 加法族であったから，ボレル集合族 $\mathcal{B}$ の最小性から $\mathcal{B} \subset \mathcal{B}_X$ となるので，$X^{-1} : \mathcal{B} \to \mathcal{F}$ となる．即ち，任意のボレル集合 $A \in \mathcal{B}$ に対して X の値が A に入る確率 $P(X \in A)$ が定義される（存在する，計算できる）ことになる．これが X が確率変数であることの別の言い方である．

$X : \Omega \to \mathbf{R} \cup \{\pm\infty\}$ に対しても，集合族 $\{[-\infty, x] \mid x \in \mathbf{R}$ を含む最小の σ 加法族を $\mathcal{B}[-\infty, \infty]$ と書けば，X が確率変数であるとは $X^{-1} : \mathcal{B}[-\infty, \infty] \to \mathcal{F}$ となることである．

$A \in \mathcal{B}$ に対して，X の値が A に含まれる確率 $Q(A) = P(X \in A) = P(X^{-1}(A))$ が定義されるから，これによって $\mathcal{B}$ 上の関数（集合関数） $Q = P \circ X^{-1}$ を定義すれば，Q は $\mathcal{B}$ 上の確率測度になることが分かる．$(\mathbf{R}, \mathcal{B})$ 上の確率測度 $Q = P \circ X^{-1}$ を確率変数 X の**分布**，または**法則**と呼ぶ．例えば，X が正規分布に従うと言うときは，$Q = P \circ X^{-1}$ が正規分布であるという意味で使う．

確率を実際に計算するときは分布の具体形が問題であって，確率変数という概念はあまり意識しない．これに対して，確率変数は，現実に調べたい問題を，確率の問題として自然に定式化（式に書き下すなどして，曖昧さなく問題を明らかにこと）するのに有効なことが多い．次回以降見るように，複数のばらつく量を調べたいとき確率変数は重要である．

2　期待値．(火曜日２)

確率変数の定義の一つの重要な数学的意味は，期待値 $\mathrm{E}[\,X\,]$ が積分として定義できることである．

2.1　期待値．

最初に定義せずに使った期待値 $\mathrm{E}[\,X\,]$ と分散 $\mathrm{V}[\,X\,]$ の定義及びいくつかの性質を取り上げる．

$(\Omega,\mathcal{F},P)$ を確率空間とし，$X,\,Y$ を（$(\Omega,\mathcal{F},P)$ の上の）確率変数とする．

期待値は**積分**である，と定義すると素朴な気持ちに合うことがわかっている．確率空間上の確率変数の積分，即ち期待値

$$\mathrm{E}[\,X\,]=\int_\Omega X(\omega)\,dP(\omega)=\int_\Omega X\,dP$$

を定義する現代的な積分論を述べておく．いろんな確率（確率空間）を統一的に（一般論で）扱うために，高校時代に習う積分（リーマン積分）以外のものでも，以下の定義に当てはまるものは全てまとめて積分と呼ぶことにする．

$A\in\mathcal{F}$ の**定義関数**とは $\mathbf{1}_A:\Omega\to\mathbf{R}$;

$$\mathbf{1}_A(\omega)=\begin{cases}1, & \omega\in A,\\ 0, & \omega\notin A\,.\end{cases}$$

定義関数の線形結合を**単関数**と呼ぶ：$X=a_1\mathbf{1}_{A_1}+\cdots+a_N\mathbf{1}_{A_N}$, $a_i\in\mathbf{R}$, $A_i\in\mathcal{F}$, $i=1,2,\cdots,N$, $N\in\mathbf{N}$. [10]　階段状の関数をイメージすればよいが，可測集合 A_i たちはカントール集合のように無数の「破片」からなっているかもしれないので,「高さ方向」は階段だが,「水平方向」は複雑かもしれない．

$A\in\mathcal{F}$ における確率変数 $X:A\to\mathbf{R}$ の**期待値 (即ち積分)**

$$\mathrm{E}[\,X;A\,]=\int_A X\,dP=\int_{\omega\in A}X(\omega)\,P(d\omega)$$

の定義．$X=X_+-X_-$; $X_+=\max\{X,0\}$ (≥ 0), $X_-=-\min\{X,0\}$ (≥ 0), と分解することにより，非負確率変数 $(X(\omega)\ge 0,\,\omega\in A)$ について定義しておいて，一般の X については $\mathrm{E}[\,X;A\,]=\mathrm{E}[\,X_+;A\,]-\mathrm{E}[\,X_-;A\,]$ とすれば（$\mathrm{E}[\,X_+;A\,]=\mathrm{E}[\,X_-;A\,]=\infty$ でない限り）よいことが分かる．括弧内の制限条件は実際には重要．関数列の極限と積分の交換に関して多数の定理があるが，それは結局，X_n が $n\to\infty$ で $\pm\infty$ に発散する場合の微妙さか，さもなければ括弧内の事態について，問題が起きないための便利な十分条件をいろいろ用意しているにすぎない．

非負値確率変数 $X:\Omega\to\mathbf{R}_+$ の期待値 (積分) は

$$\mathrm{E}[\,X\,]=\sup\{a_1P(A_1)+\cdots+a_NP(A_N)\mid a_i\in\mathbf{R},\ i=1,2,\cdots,N,\ \bigcup_{i=1}^N A_i=\Omega,\ a_1\mathbf{1}_{A_1}+\cdots+a_N\mathbf{1}_{A_N}\le X\}$$

で定義する．定義から特に $X=a_1\mathbf{1}_{A_1}+\cdots+a_N\mathbf{1}_{A_N}$ が単関数のとき

$$\mathrm{E}[\,X\,]=a_1\,P(A_1)+\cdots+a_N\,P(A_N)\,.$$

$A\in\mathcal{F}$ における期待値は $\mathrm{E}[\,X;A\,]=\mathrm{E}[\,X\mathbf{1}_A\,]$ である．例えば $\mathrm{E}[\,X;X\ge a\,]=\mathrm{E}[\,X\mathbf{1}_{\{X\ge a\}}\,]=\mathrm{E}[\,X\mathbf{1}_{\{\omega\in\Omega\mid X(\omega)\ge a\}}\,]$ は証明や計算で頻出のテクニック．

以上の定義は確率変数 X の値域に ∞ が含まれるかどうかによらない．定義から非負値確率変数 X に関して $P(X=\infty)>0$ ならば $\mathrm{E}[\,X\,]=\infty$ である．

非負とは限らない場合は $\mathrm{E}[\,X_\pm\,]<\infty$ の時のみ積分可能 (可積分) と呼び，上の定義で積分を定義する．この条件は非負値関数 $|X|$ の積分 $\mathrm{E}[\,|X|\,]<\infty$ と同値である．$\mathrm{E}[\,X_\pm\,]$ の一方のみが ∞ のときはそれぞ

[10] $A\subset\Omega$ で定義された関数 Y, z に対して $X=Y+Z$, $X=Y\cdot Z$ はそれぞれ $X(\omega)=Y(\omega)+Z(\omega)$, $X(\omega)=Y(\omega)\,Z(\omega)$ で定義される関数，さらに $a\in\mathbf{R}$ に対して $X=aY$ は $X(\omega)=a\,Y(\omega)$ で定義される関数．また，$X\ge 0$, $X\ge Y$ はそれぞれ $X(\omega)\ge 0$, $X(\omega)\ge Y(\omega)$, $(\omega\in\Omega)$ を意味することにする．

れ $\mathrm{E}[\,X\,]=\pm\infty$ と書く．両方 ∞ のときは積分は定義されない．(広義リーマン積分のうち $\dfrac{1}{x}\sin x$ の積分のように正の部分と負の部分の積分がそれぞれ発散するが，積分範囲を有界区間 $[-M,M]$ に切っておいて $M\to\infty$ で定義すると収束する積分はルベーグ積分不可能である.)

2.2 期待値の性質．

2.2.1 非負確率変数の期待値の性質．

上に述べた期待値の定義は測度論に基づく出来るだけ短い定義である．実際に使いやすい (種々の定理を証明しやすい) 定義や性質を導くのに必要な性質を列挙しておこう．以下，$(\Omega,\mathcal{F},P)$ を確率空間，X, X_n などは Ω 上の**非負値**確率変数とする．

(1) $X\le Y$ ならば $0\le \mathrm{E}[\,X\,]\le \mathrm{E}[\,Y\,]$.
(2) 正定数 a, b に対して $\mathrm{E}[\,aX+bY\,]=a\mathrm{E}[\,X\,]+b\mathrm{E}[\,Y\,]\le\infty$ （期待値の線型性）.
(3) $\mathrm{E}[\,X\,]=$ ならば $P(X>0)=0$, $P(X=0)=1$.
(4) $\{X_n\}$ が各 $\omega\in\Omega$ 毎に単調増加で X に収束するならば $\mathrm{E}[\,X_n\,]$ は (単調増加して) $\mathrm{E}[\,X\,]$ に収束する（単調収束定理）．関数や積分が ∞ でも構わない.
(5) $E\in\mathcal{F}$, $P(E)=0$ ならば $\mathrm{E}[\,X;E\,]=0$．$X=\infty$ でも構わない.
(6) $X=Y$, a.e., ならば $\mathrm{E}[\,X\,]=\mathrm{E}[\,Y\,]$．$P(X=Y)=1$ であることを「ほとんど至る所 $X=Y$」と表現し，$X=Y$, a.e., などと書く.
(7) $X_n\uparrow X$, a.e., ならば $\mathrm{E}[\,X_n\,]\uparrow \mathrm{E}[\,X\,]$．たいていの収束定理は前提条件で「全ての $\omega\in\Omega$ について」を a.e., におきかえても成り立つ.
(8) $P(\liminf_n X_n)\le \liminf_n P(X_n)$ （Fatou の補題）.
(9) $c\,P(X\ge c)\le \mathrm{E}[\,X\,]$ $(c>0)$.
(10) $c\ge 0$ かつ X^2 が可積分ならば $c^2P(|X-\mathrm{E}[\,X\,]|>c)\le \mathrm{V}[\,X\,]$ （Chebyshev の不等式）.
(11) $\displaystyle\sum_{n=1}^{\infty}\mathrm{E}[\,X_n\,]<\infty$ ならば $\displaystyle\lim_{n\to\infty}X_n=0$, a.e., 即ち，ほとんど確実に収束する （Borel）–Cantelli 第 1 定理）.

(3)(4) の証明はいずれも確率の連続性を用いる．(3) は $\{X>0\}=\displaystyle\lim_{n\to\infty}\{X>1/n\}$ を用いれば明らかだが，(4) の証明は少し手間を要する．関数 $\alpha^{(r)}:[0,\infty]\to[0,\infty]$ を

$$\alpha^{(r)}(x)=\begin{cases}0, & x=0\\ (i-1)2^{-r}, & (i-1)\,2^{-r}<x\le i\,2^{-r}\le r,\\ r, & x>r,\end{cases}$$

で定義する (左連続なので単調増加関数列 $X_n\uparrow X$ に対して $\alpha^{(r)}\circ X_n\uparrow\alpha^{(r)}\circ X$)．単関数近似列 $Y^{(r)}=\alpha^{(r)}\circ X$ を利用すると，単関数列 Z_n, $n=1,2,3,\cdots$ であって各点で X に収束 ($\displaystyle\lim_{n\to\infty}Z_n(\omega)=X(\omega)$) し，期待値が $\mathrm{E}[\,X\,]$ に収束 ($\mathrm{E}[\,X\,]=\displaystyle\lim_{n\to\infty}\mathrm{E}[\,Z_n\,]$) する増大列 $(Z_1(\omega)\le Z_2(\omega)\le\cdots)$ が取れる．次に二重数列の一般論から $Y_n\uparrow X$ となる任意の単関数列に対しても期待値の列が同じ値に収束する (近似列によらない) ことが言える．これを $\alpha^{(r)}\circ X_n$ に適用する.

(9) は後述の Markov の不等式の例と見ることもできるが，線型性と正値性を用いて

$$\mathrm{E}[\,X\,]=\mathrm{E}[\,X;X\ge c\,]+\mathrm{E}[\,X;X<c\,]\ge \mathrm{E}[\,X;X\ge c\,]\ge \mathrm{E}[\,c;X\ge c\,]=cP(X\ge c)$$

と，直接証明できる．この種の不等式は，簡単だがいろいろな公式を証明するのに有用である.

(11) は，線型性と単調収束定理から $\mathrm{E}[\,\displaystyle\sum_{n=1}^{\infty}X_n\,]=\sum_{n=1}^{\infty}\mathrm{E}[\,X_n\,]\le\infty$ となるので，既に注意したように $\mathrm{E}[\,X\,]<\infty$ ならば $P(X=\infty)=0$ と合わせると結論を得る.

2.2.2 確率変数の期待値の性質．

非負値とは限らない確率変数 X の期待値は $X_+=\max\{X,0\}$, $X_-=-\min\{X,0\}$ とおけば $X=X_+-X_-$, $|X|=X_++X_-$ となる．

X が $(P$–$)$ 積分可能 (可積分，期待値が存在する) とは非負値確率変数 $|X|$ に対して $\mathrm{E}[\,|X|\,]<\infty$ なること．このとき $\mathrm{E}[\,X\,]=\mathrm{E}[\,X_+\,]-\mathrm{E}[\,X_-\,]$ で期待値を定義する．

このことを $X\in\mathcal{L}^1(\Omega,\mathcal{F},P)$ と書くこともある．(関数解析で $X\in L^1(\Omega,\mathcal{F},P)$ と書くと，$X=Y$, a.e., なる Y は X と同一視するが，確率論では同一視するとは限らない．$\mathcal{L}$ を使ったのはその細かい違いにこだわった.)

以下，X, X_n を (非負とは限らない一般の) 確率変数とする．

(1) $|\mathrm{E}[\,X\,]|\le\mathrm{E}[\,|X|\,]$.
(2) 狭義 (有限区間で非積分関数が有界な場合) リーマン積分はここでの一様分布による期待値に等しい.
(3) X, Y が可積分ならば，定数 a, b に対して $\mathrm{E}[\,aX+bY\,]=a\mathrm{E}[\,X\,]+b\mathrm{E}[\,Y\,]$ が成り立つ　（期待値の線型性）.
(4) $n\ne m$ ならば $A_n\cap A_m=\emptyset$ のとき，$\displaystyle\sum_{n=1}^{\infty}\mathrm{E}[\,X;A_n\,]=\mathrm{E}[\,X;\bigcup_{n=1}^{\infty}A_n\,]$ が成り立つ　（積分範囲に関する加法性）.
(5) 各点で $\displaystyle\lim_{n\to\infty}X_n=X$，かつ，非負値可積分 $(\mathrm{E}[\,Y\,]<\infty)$ 確率変数 Y があって，$|X_n|\le Y$ が各点かつ全ての $n\in\mathbf{N}$ に対して成り立つならば，$\displaystyle\lim_{n\to\infty}\mathrm{E}[\,|X_n-X|\,]=0$ であり，従って特に $\displaystyle\lim_{n\to\infty}\mathrm{E}[\,X_n\,]=\mathrm{E}[\,X\,]$ （Dominated convergence theorem）.
(6) X, X_n が可積分で $\displaystyle\lim_{n\to\infty}X_n=X$, a.e., とする．このとき，$\displaystyle\lim_{n\to\infty}\mathrm{E}[\,|X_n-X|\,]=0$ と $\displaystyle\lim_{n\to\infty}\mathrm{E}[\,|X_n|\,]=\mathrm{E}[\,|X|\,]$ は同値である　（Scheffé の補題）.
(7) $g:\mathbf{R}\to[0,\infty]$ は単調非減少な $\mathcal{B}$ 可測関数，即ち，$g^{-1}([0,a])\in\mathcal{B}$ が全ての $a\ge0$ に対して成り立つ.（$\mathcal{B}$ は実数の全ての区間を含む最小の σ 加法族.）このとき $\mathrm{E}[\,g\circ X\,]\ge\mathrm{E}[\,g\circ X;X\ge c\,]\ge g(c)P(X\ge c)$ （Markov (Chernov) の不等式）.
(8) X が可積分ならば $P(X\ge c)\le\exp(-\theta\,c)\,\mathrm{E}[\,\exp(\theta\,X)\,]$ $(c\in\mathbf{R},\ \theta>0)$.
(9) $I=(a,b)\subset\mathbf{R}$ （無限区間を許す），$g:I\to\mathbf{R}$ は convex（下に凸）とし，$\mathrm{E}[\,|X|\,]<\infty$, $P(X\in I)=1$, $\mathrm{E}[\,|g\circ X|\,]<\infty$ を満たすならば，$\mathrm{E}[\,g\circ X\,]\ge g(\mathrm{E}[\,X\,])$　（Jensen の不等式）.
(10) $|\mathrm{E}[\,XY\,]|^2\le\mathrm{E}[\,|XY|\,]^2\le\mathrm{E}[\,X^2\,]\mathrm{E}[\,Y^2\,]$　（Schwarz の不等式）.

(8) は Markov の不等式 (7) で $g(x)=\exp(\theta x)$ としたもの．(10) は関数（確率変数）の集合を考え，$\mathrm{E}[\,|X|^p\,]$ をノルムと見る立場で自然な不等式．他に，Hölder 不等式，三角不等式 (Minkowski 不等式)，ノルムの p についての単調性，ピタゴラスの定理，中線連結定理，などがある．

$A\in\mathcal{F}$ における期待値は $\mathrm{E}[\,X;A\,]=\mathrm{E}[\,X\mathbf{1}_A\,]$ である．このとき，$\mathrm{E}[\,X;A\,]=\int_A X|_A\,dP|_A$ となる．ここで $X|_A$ は X の A への制限，$P|_A$ は P で定義域 $\mathcal{F}$ を $\{E\in\mathcal{F}\mid E\subset A\}$ に制限したもの．つまり，$\mathrm{E}[\,X;A\,]$ の定義は，期待通り，積分範囲を A に制限した積分に一致する．

問題．

特に，$P(A)=\mathrm{E}[\,\mathbf{1}_A\,]$, $P(A\cap B)=\mathrm{E}[\,\mathbf{1}_A\cdot\mathbf{1}_B\,]$, など．では，$P(A\cup B)$ を $\mathrm{E}[\,\cdot\,]$ を用いて表すと？

2.3 例．

2.3.1 さいころの目．

$$\Sigma=\{1,2,3,4,5,6\},\ \Omega_n=\Sigma^n,\ \mathcal{F}_n=2^{\Omega_n},\ A\subset\Omega_n\ \text{ に対して }P(A)=6^{-n}\sharp A=\frac{\sharp A}{\sharp\Omega},$$

n 回目までに平均何種類の目 $\mathrm{E}[\,N\,]$ が出るか？

数え上げる方法：$n=1$: $N=1$ 種類． $n=2$ 確率 $1/6$ で $N=1$, $5/6$ で $N=2$ だから $\mathrm{E}[\,N\,]=11/6$. どんどん計算が大変になる．

期待値の線型性を使う方法：

$$X_i = \begin{cases} 1, & \exists k;\ 1\le k\le n,\ \omega_k = i,\ (\text{即ち，目 } i \text{ が出るとき}), \\ 0, & (\text{目 } i \text{ が出ないとき}), \end{cases}$$

とおくと $N=\sum_{i=1}^{6} X_i$ である．どの i についても $\mathrm{E}[\,X_i\,]=P(X_i=1)=1-P(X_i=0)$ であり，$P(X_i=0)$ は目 i が n 回のさいころ投げの間中一度も出ない確率だから，$P(X_i=0)=6^{-n}$ となるので，

$$\mathrm{E}[\,N\,]=\sum_{i=1}^{6}\mathrm{E}[\,X_i\,]=6(1-6^{-n})\,.$$

すんなりできてしまう．異なる目が出る事象は排反でも独立でもないが，期待値を求めるだけなら別々に計算すればよいところが期待値の線型性の有効性である．

2.3.2 誕生日休業制度と最大労働力．

ある会社で従業員の誕生日を休日とすることにした．1 年 365 日，誕生日以外の休日はなし，二人以上の誕生日が重なっても振り替え休日等はない，とする．このとき，平均延べ年間労働力 L (従業員数 n と労働日数の積) を最大にするには何人の従業員を雇えばいいか．なお，誕生日は独立で一様分布する，即ち，例えば一月一日が n 人のうちだれの誕生日でもない確率は $\left(\frac{364}{365}\right)^n$ である，と仮定する．

誕生日が重なる可能性があるので L は定数ではなく確率変数である．

数え上げる方法：$n=1$ のときは誕生日は重ならないから $a_1=\mathrm{E}[\,L\,]=364$. $n=2$ のとき，$P(L=364\times 2)=\dfrac{1}{365}$, $P(L=363\times 2)=\dfrac{364}{365}$, だから

$$a_2=\mathrm{E}[\,L\,]=\frac{2\times 364}{365}\times(363+1)=2\times\frac{364^2}{365}>a_1\,.$$

$n=3$ のときは $L=364\times 3$, $L=363\times 3$, $L=362\times 3$, の 3 通りがある．どんどん大変になっていく．

期待値の線型性を使う方法：

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 日が労働日のとき}, \\ 0, & \text{第 } i \text{ 日が休日のとき}, \end{cases}$$

とおくと $L=n\sum_{i=1}^{365} X_i$ である．どの i についても $\mathrm{E}[\,X_i\,]=P(X_i=1)$ は第 i 日が従業員の誕生日にならない確率だから，仮定よりその値は $\left(\frac{364}{365}\right)^n$ となるので，

$$a_n=\mathrm{E}[\,L\,]=365\left(\frac{364}{365}\right)^n n\,.$$

よって

$$a_{n+1}-a_n=\left(\frac{364}{365}\right)^n(364-n)$$

となるので $a_1<a_2<\cdots<a_{364}=a_{365}>a_{366}>a_{367}>\cdots$ を得て，364 人または 365 人雇うのが良いということになる．

以上の例のように，サンプルの数が有限個だと数え上げでも計算できるが，確率論のものの見方を用いると簡単になることがある．確率変数の個数が非可算無限個（確率過程）になると，確率論のものの見方を用いなければ定義すらできなくなる．

2.4　素朴な期待値との関係　（省略可能？）．

X の期待値 $\mathrm{E}[\,X\,]$ とは素朴には X のとりうる値の（確率に応じた）平均値，である．例えば整数値をとる確率変数（さいころの目，など）X に対して，

$$\mathrm{E}[\,X\,] = \sum_n n\,P(X=n) = \sum_n X\,P(X=n)$$

となる．とりうる値が実数（非可算種類）だと和では書けない．素朴には $X=x$ における X の値の分布密度 ρ_X があるときは

$$\mathrm{E}[\,X\,] = \int x\rho_X(x)\,dx$$

と計算する．

$A\in\mathcal{B}$ に対して $Q(A)=P(X\in A)=P(X^{-1}(A))$ で定義された Q が $(\mathbf{R},\mathcal{B})$ 上の確率になり，X の分布 (法則) と呼ぶ，と先ほど説明した．Q は X の値の出る頻度を与えるのだから，X の期待値は Q で計算できるはずである．実際，Q の密度が ρ_X，すなわち，$P(X\in A)=Q(A)=\int_A \rho_X(x)\,dx$, $A\in\mathcal{B}$, のとき，素朴な計算式

$$\mathrm{E}[\,X\,] = \int_{\mathbf{R}} x\,\rho_X(x)\,dx\,,$$

が示せる．Q は X の値の分布だから，例えば X が整数値しか取らない確率変数ならば，密度は持たない．その場合でも一般論は変わらない．例えば n が整数のとき $P[X=n]=q_n$ $(\sum_n q_n = 1)$ とすると，同様に次式を得る：

$$\mathrm{E}[\,X\,] = \sum_{n=-\infty}^{\infty} n\,q_n\,,$$

X の期待値が元の確率空間 $(\Omega,\mathcal{F},P)$ に関係なく，値の空間（X の**状態空間**と呼ぶ）の上の確率 $(\mathbf{R},\mathcal{B},Q)$ だけで決まることにも注目してほしい．例えば，測定実験を行うことを考えると，実験データ X を集めることで，計器の値の分布 Q が実験的に求められ，これに基づいて次の実験の予測が可能になる．データがばらつく本当の原因 $(\Omega,\mathcal{F},P)$ は必要ない．

確率変数という言葉を使うと，確率変数 X が一様分布に従うとは，A を区間 $[0,1]$ の部分集合で $A\in\mathcal{B}$ とするとき，$X\in A$ が起きる確率 $P(X\in A)$ が区間 A の長さに等しいことを言う．

これまでの講義で定義をしないまま使った確率変数，期待値，分布と言った用語は積分論に基づいてこのようにきちんと定義できる．

3　独立性．(水曜日)

3.1　分散．

確率変数の分散は

$$\mathrm{V}[\,X\,] = \mathrm{E}[\,(X-\mathrm{E}[\,X\,])^2\,]$$

で定義される．期待値の線型性から直ちに次の性質が分かる．

(1) $\mathrm{V}[\,X\,] = \mathrm{E}[\,X^2\,] - \mathrm{E}[\,X\,]^2$,
(2) a が定数のとき $\mathrm{V}[\,aX\,] = a^2\mathrm{V}[\,X\,]$,
(3) $\mathrm{V}[\,X+Y\,] = \mathrm{V}[\,X\,] + \mathrm{V}[\,Y\,] + 2\mathrm{E}[\,(X-\mathrm{E}[\,X\,])(Y-\mathrm{E}[\,Y\,])\,]$,

第 1 の性質を証明するとき，期待値 $\mathrm{E}[\,X\,]$ は定数（期待値を取ると確率変数ではなくなる）であること，従って，線型性から

$$\mathrm{E}[\,\mathrm{E}[\,X\,]\,] = \mathrm{E}[\,X\,]\,\mathrm{E}[\,1\,] = \mathrm{E}[\,X\,]$$

となること，を用いる．第 2 の性質は一列並びの問題で解決の鍵になった性質である．最後の性質は，一般には分散には加法性がないことを表す．

$$\begin{aligned}
&\mathrm{V}[\,X+Y\,] = \mathrm{E}[\,(X+Y-\mathrm{E}[\,X\,]-\mathrm{E}[\,Y\,])^2\,]\\
&= \mathrm{E}[\,(X-\mathrm{E}[\,X\,])^2\,] - 2\mathrm{E}[\,X-\mathrm{E}[\,X\,]\,]\mathrm{E}[\,Y-\mathrm{E}[\,Y\,]\,] + \mathrm{E}[\,(Y-\mathrm{E}[\,Y\,])^2\,]\\
&= \mathrm{V}[\,X\,] + \mathrm{V}[\,Y\,] - 2(\mathrm{E}[\,XY\,] - \mathrm{E}[\,X\,]\mathrm{E}[\,Y\,])\,.
\end{aligned}$$

分散が加法性を持つのは，X と Y が独立な場合である．

3.2 独立性．

測度論でない確率論らしさ．独立な確率変数を考えるということは測度論の目で見ると直積測度という特別な測度を考えることに対応する．特に確率変数列の極限など無限個の独立確率変数を考えれば，無限次元空間上の確率測度を考えていることに自動的になるので，測度論から確率論が離れて，特別な場合の深い考察に向かうことになる．

確率空間 $(\Omega, \mathcal{F}, P)$ 上の n 個の確率変数 $X_i : \Omega \to \mathbf{R}$, $i = 1, \cdots, n$, が**独立**とは，

$$P(X_1 \le a_1, \cdots, X_n \le a_n) = \prod_{i=1}^{n} P(X_i \le a_i), \qquad a_i \in \mathbf{R},\ i = 1, 2, \cdots, n,$$

が成り立つことと定義する．σ 加法族の性質から X_i, $i = 1, \cdots, n$, が独立ならば

$$P(X_1 \in A_1, \cdots, X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i), \qquad A_i \in \mathcal{B},\ i = 1, 2, \cdots, n,$$

が成り立つ．$\mathcal{B}$ は全ての区間を含む最小の σ 加法族．さらにこのことから $f_i : \mathbf{R} \to \mathbf{R}$, $i = 1, \cdots, n$, が $(\mathbf{R}, \mathcal{B})$ 上の可測関数，即ち，

$${f_i}^{-1}((-\infty, a]) \in \mathcal{B},\ a \in \mathbf{R}$$

が成り立つとき，$f_i \circ X_i$ たちも独立になる．

期待値との関連では，次のことが証明できる．$X_i : \Omega \to \mathbf{R}$, $i = 1, \cdots, n$, が独立で X_i たちが可積分ならば，これらの積も可積分で，

$$\mathrm{E}[\,X_1\,X_2\,\cdots\,X_n\,] = \mathrm{E}[\,X_1\,]\,\mathrm{E}[\,X_2\,]\,\cdots\,\mathrm{E}[\,X_n\,]\,.$$

即ち，**独立な確率変数の積の期待値は期待値の積に等しい．**（X_i たちが独立ならば $f_i \circ X_i$ たちも独立になるから，これらの積の期待値も期待値の積に等しくなる．）

証明は $X_i = X_{i+} - X_{i-}$ と非負値確率変数の差で書くことで，先ず非負値確率変数たちについての問題に帰着させる．そして，単関数 $Y_i^{(r)} = \alpha^{(r)} \circ X_i$ たちやその積については期待値は定義によって $\{X_i \le \beta_i\}$ という形の可測集合やその共通部分の確率の線形結合に書けるから，定義に戻って共通部分の確率を $P(X_i \le \beta_i)$ の積に変形することで

$$\mathrm{E}[\,Y_1^{(r)}\,Y_2^{(r)}\,\cdots\,Y_n^{(r)}\,] = \mathrm{E}[\,Y_1^{(r)}\,]\,\mathrm{E}[\,Y_2^{(r)}\,]\,\cdots\,\mathrm{E}[\,Y_n^{(r)}\,]$$

を言う．

実際 $\alpha^{(r)} \circ X = \sum a_i \mathbf{1}_{A_i}$ および $\alpha^{(r)} \circ Y = \sum b_j \mathbf{1}_{B_j}$ とおけば $A_i \in \sigma(X)$, $B_j \in \sigma(Y)$, かつ，$\sigma(X) \perp \sigma(Y)$ だから

$$\mathrm{E}[\,\alpha^{(r)} \circ X\,\alpha^{(r)} \circ Y\,] = \sum\sum a_i b_j P(A_i \cap B_j) = \sum\sum a_i b_j P(A_i)\,P(B_j) = \mathrm{E}[\,\alpha^{(r)} \circ X\,]\,\mathrm{E}[\,\alpha^{(r)} \circ Y\,]\,.$$

ここで $r \to \infty$ とすれば単調収束定理より結論を得る．

特に，このことから，X と Y が独立ならば，$\mathrm{E}[\,XY\,] = \mathrm{E}[\,X\,]\mathrm{E}[\,Y\,]$ だから，

$$\mathrm{V}[\,X+Y\,] = \mathrm{V}[\,X\,] + \mathrm{V}[\,Y\,] + 2\,\mathrm{E}[\,X-\mathrm{E}[\,X\,]\,]\,\mathrm{E}[\,Y-\mathrm{E}[\,Y\,]\,] = \mathrm{V}[\,X\,] + \mathrm{V}[\,Y\,]$$

を得る．即ち，**独立な確率変数の和の分散は分散の和に等しい．**または，**独立な確率変数に対しては分散は加法的である．**

確率変数の無限列，即ち確率変数が無限個 $X_i,\, i=1,2,\cdots,$ ある場合には，$\{X_i\}$ が独立とは，任意のの有限部分列が独立，即ち，任意の n に対して，$X_i,\, i=1,\cdots,n,$ が独立ということ，と定義する．

確率変数列が独立ならば任意の部分列は独立になる．これは元の列が有限列でも無限列でも正しい．しかし，有限列の場合，X と Y が独立，Y と Z が独立，Z と X が独立で，X,Y,Z が独立でない例が知られている．無限列の場合はその有限部分列が全て独立であることで全体の独立性を**定義する**.

独立性と分布　（省略可能？）．

分布の言葉でいうと次のようになる．$(X_1,\cdots,X_n)$ の結合分布（まとめて n 次元実数値確率変数としてみたときの n 次元空間上の分布）を Q，X_i の分布を Q_i とすると，$X_i,\, i=1,\cdots,n,$ が独立とは,

$$Q[A_1\times\cdots\times A_n]=Q_1[A_1]\,Q_2[A_2]\,\cdots\,Q_n[A_n]$$

が全ての $A_i\in\mathcal{B},\, i=1,2,\cdots,n,$ に対して成り立つこと．ここで，左辺は直積集合の確率，右辺は確率の積である．

3.3　分布関数．

確率空間 $(\Omega,\mathcal{F},P)$ 上の確率変数 $X:\Omega\to\mathbf{R}\cup\{\pm\infty\}$ に対して $F_X(x)=P(X\le x)=P(\{\omega\mid X(\omega)\le x\})$ を X の分布関数と呼ぶ.

分布関数の性質．F が何かの確率変数の分布関数ならば以下が成立．逆に以下が成立すれば何かの確率空間の何かの確率変数の分布関数になっている．

(1) $F:\mathbf{R}\cup\{\pm\infty\}\to[0,1]$,
(2) $F\ \uparrow$,
(3) $F(\infty)=1,\, F(-\infty)=0$　（X が ∞ をとる確率が 0 ならば $\displaystyle\lim_{x\to\infty}F(x)$ など）,
(4) F は右連続：$\displaystyle\lim_{y\downarrow x}F(y)=F(x)$.

分布関数と期待値の関係．F を X の分布関数とし，f をボレル可測関数（$(\mathbf{R},\mathcal{B})$ 上の可測関数）とすると

$$\mathrm{E}[\,f(X)\,]=\int_{\mathbf{R}}f(x)\,dF(x)\,.$$

本当は右辺は Lebesgue–Stieltjes 積分である．今まで述べてきた期待値の定義を F が定義する測度に適用すれば定義を得るが，正確な説明は省略する．証明は f が単関数のときを証明して単調収束定理を用いる．詳しいことは省略する．分布関数 F が微分可能な場合は

$$\mathrm{E}[\,f(X)\,]=\int_{\mathbf{R}}f(x)\,\frac{dF}{dx}(x)\,dx\,.$$

あとでこれを用いる.

3.3.1　分布関数で書いた独立性．

$F_X,\,F_Y,\,X+Y$ を $X,\,Y,\,X+Y$ の分布関数とする．$X,\,Y$ が独立ならば,

$$F_{X+Y}(x)=\int_{-\infty}^{\infty}F_Y(x-y)\,dF_X(y)\,.$$

右辺を F_X と F_Y の**畳み込み**と呼び,

$$f*g(x)=\int_{-\infty}^{\infty}g(x-y)\,df(y)$$

と書く．右辺は Lebesgue–Stieltjes 積分．分布関数が微分可能な場合は

$$f * g(x) = \int_{-\infty}^{\infty} \frac{df(y)}{dy}\, g(x-y)\, dy$$

と書ける．

この公式は形式的には次のようにして「導く」ことができる．δ を Dirac のデルタとして，

$$\int \left[\frac{d}{dz} \mathbf{1}_{X \le z} \right] \mathbf{1}_{Y \le x-z}\, dz = \int \delta(z-X)\, \mathbf{1}_{Y \le x-z}\, dz = \mathbf{1}_{X+Y \le x}\,.$$

両辺の期待値をとって左辺で X と Y が独立であることを使い，さらにパラメータ微分と期待値の交換を許す（一般に可能）と，

$$\int \frac{d}{dz} \mathrm{E}[\, \mathbf{1}_{X \le z} \,] \mathrm{E}[\, \mathbf{1}_{Y \le x-z} \,]\, dz = \mathrm{E}[\, \mathbf{1}_{X+Y \le x} \,] = P(X+Y \le x) = F_{X+Y}(x)\,.$$

左辺も同様に変形すると $\displaystyle\int F_x'(z)\, F_Y(x-z)\, dz$ を得る．(Fubini に相当する性質が必要なので，準備なしで正しく証明することはできない．)

たたみ込みの性質．

$f_i : \mathbf{R} \to [0,1]$, $i = 1,2,3$, が分布関数，即ち単調増加右連続で，$\displaystyle\lim_{x \to -\infty} f_i(x) = 0$, $\displaystyle\lim_{x \to \infty} f_i(x) = 1$, $i = 1,2,3$, を満たすとする．

(1) $f_1 * f_2 = f_2 * f_1$.
(2) $(f_1 * f_2) * f_3 = f_1 * (f_2 * f_3)$.
(3) $(f_1 * f_2)(x+y) \ge f_1(x) f_2(y)$, $x, y \in \mathbf{R}$.

f_i が微分可能なときの証明．(1) は部分積分と $y' = x - y$ で

$$f_1 * f_2(x) = \int_{-\infty}^{\infty} f_1'(y)\, f_2(x-y)\, dy = \int_{-\infty}^{\infty} f_2'(x-y)\, f_1(y)\, dy = f_2 * f_1(x)\,.$$

(2) は

$$\begin{aligned} &(f_1 * f_2) * f_3(x) = f_3 * (f_1 * f_2)(x) = \int\!\!\int f_3'(x-y)\, f_1'(z)\, f_2(y-z)\, dy \\ &= \int\!\!\int f_3'(x-z)\, f_1'(y)\, f_2(z-y)\, dy = \int\!\!\int f_1'(y)\, f_2(x-z-y)\, f_3'(z)\, dy \\ &= f_1 * (f_2 * f_3)(x)\,. \end{aligned}$$

(3) は独立確率変数 X_1, X_2 を用意して $f_i(x) = P(X_i \le x)$ および $(f_1 * f_2)(x) = P(X_1 + X_2 \le x)$ に注目するのが分かりやすい．実際，

$$(f_1 * f_2)(x+y) = P(X_1 + X_2 \le x+y) \ge P(X_1 \le x,\ X_2 \le y) = P(X_1 \le x) P(X_2 \le y) = f_1(x) f_2(y)\,.$$

Superadditivity．

X_n, $n \in \mathbf{N}$, を i.i.d.，即ち，独立確率変数で分布 (関数) がどれも等しいものたちとし，その分布関数を $G(x) = P(X_n \le x)$ とおく．独立なので和 $\displaystyle\sum_{k=1}^{n} X_k$ の分布関数は G の n 回 convolution に等しい：$G^{*n}(x) = P\left(\displaystyle\sum_{k=1}^{n} X_k\right)$.

さて，$x \in \mathbf{R}$ を固定し，実数列 a_n, $n \in \mathbf{N}$, を $a_n = \log G_n(nx)$ で定義すると，(3) から，任意の自然数 n, m に対して

$$a_{n+m} = \log G_{n+m}((n+m)x) \ge \log G_n(nx) + \log G_m(mx) = a_n + a_m$$

が成り立つ．等号なら加法性と呼ばれる性質なので，この，不等号の性質を superadditivity と呼ぶ．

Superadditive な数列 $\{a_n\}$ は a_n/n の $n\to\infty$ の極限が存在する．より詳しく，次のことが成り立つ．

$$\lim_{n\to\infty}\frac{1}{n}a_n=\sup_{n\geq 1}\frac{1}{n}a_n\,.$$

証明は，確率論は関係なく単なる初等解析である．任意の $\ell, n\in\mathbf{N}$ に対して $n=q\ell+r$ となる整数たち $q\geq 0,\ 1\leq r\leq \ell$ がとれるので superadditivity から

$$\frac{1}{n}a_n\geq\frac{q}{n}a_\ell+\frac{1}{n}a_r\,.$$

q, r のとりかたから $\dfrac{q}{n}\geq\dfrac{1}{\ell}-\dfrac{1}{n}$ なので，$c=\displaystyle\min_{1\leq r\leq\ell}a_r-a_\ell$ とおくと $c<\infty$ は n によらない．ゆえに

$$\frac{1}{n}a_n\geq\frac{1}{\ell}a_\ell+\frac{c}{n}\,.$$

従って，

$$\liminf_{n\to\infty}\frac{1}{n}a_n\geq\frac{1}{\ell}a_\ell\,.$$

これが任意の $\ell\geq 1$ に対して成り立つので証明が終わる．

3.4 例：ある公式の証明．－大偏差値原理－

極限を考えることで種々の公式を導くことができることは数列などで周知の通り．微積分は数列の極限で定義されていて，その威力は計り知れない．

取り上げる例に即して言えば数列や級数の極限に関する公式として自然対数の底

$$e=\sum_{n=0}^{\infty}\frac{1}{n!}=\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$$

や Stirling の公式

$$n!=\sqrt{2\pi n}n^n e^{-n+\theta/(12n)},\ 0<\exists\theta=\theta(n)<1,$$

などがある．

数列と同様に，確率変数の列 X_n の極限を考えると新しい結果（公式など）が得られる．ここでは分布関数を用いて，関数列の極限に関する公式

$$\lim_{n\to\infty}\left(\sum_{k=n}^{\infty}\frac{n^k x^k}{k!}\right)^{1/n}=\begin{cases}e\cdot x, & 0\leq x<1,\\ e^x, & x\geq 1,\end{cases}$$

を証明する．

$$G_n(x)=\begin{cases}e^{-x}\displaystyle\sum_{k=n}^{\infty}\frac{x^k}{k!}=1-e^{-x}\sum_{k=0}^{n-1}\frac{x^k}{k!}, & x\geq 0,\\ 0, & x<0,\end{cases}$$

とおくと，求める公式は

$$\lim_{n\to\infty}G_n(nx)^{1/n}=J(x),\ x\geq 0,$$

である．ここで

$$J(x)=\begin{cases}xe^{1-x}, & 0\leq x\leq 1,\\ 1, & x>1\,.\end{cases}$$

結果を知っていれば初等的にもできるが，ここではこの式と確率論の関係を述べたい．より一般的な結果の中の特別な場合としてとらえられる．

$x=0$ の場合は当然なので，以下 $x>0$ とする．

3.4.1　由来．－二分木の大きさ－

二分木：

(1) node と directed edge からなる平面連結図形．edge は下向き．
(2) edge は 2 つの node を結ぶ．行き先の node を子，手前の node を親と呼ぶ．
(3) どの node から出る edge も 2 つ以内．
(4) 唯一の root を除き，どの node に入る edge も 1 つ．root に入る edge はない．
(5) loop (自分に戻る path (edge による通り道) またはある node から別の node に行く 2 つの path) はない．

を満たすもの．root からの path の長さで node の level を定義する．root は level 0．2 つの子に親から見て左と右を assign．level の小さい方からの辞書式順序で一つの level 内の node を左から右に整列する．

level の順についで左から右に並べることで，存在する node に root から順に自然数 i を対応させる．各 node にデータが対応する．$x(i)$, $i = 1, 2, \cdots, n$, をデータとすると，

(1) $x(1)$ は root (node 1) に対応させる．
(2) $x(i)$ まで対応が済んだとき，$x(i+1)$ は root から始めて，

- その node に対応しているデータより $x(i+1)$ が大きければ右の子に進み，判断をくり返す，
- その node に対応しているデータより $x(i+1)$ が大きくなれば左の子に進み，判断をくり返す，
- その node にデータが未対応なら $x(i+1)$ を対応させて判断を終える．

このように対応すると，各 node の左の sub-tree は大きいデータから，右の sub-tree は大きくないデータからなるから，recursive にソートできる．

n データから作られた n node の二分木の高さ（node の level の最大値）を H_n とする（$H_0 = H_1 = 0$, $H_n \geq [\log_2 n]$）．データの順序はランダム S_n 通りの並び方が等確率で生じるとして確率空間を定義すれば H_n は確率変数になる．

これについて次のことが知られている[11]．

3.4.2　定理．

H_n は次の意味で n が大きいとき $c \log n$ のように振る舞う．

(1) 任意の $\epsilon > 0$ に対して $\displaystyle\lim_{n\to\infty} P(|\frac{H_n}{c\log n} - 1| > \epsilon) = 0$．このことを $\dfrac{H_n}{c\log n}$ は 1 に確率収束すると言う．
(2) 任意の $p > 0$ に対して $\mathrm{E}[\ \left(\dfrac{H_n}{c\log n}\right)^p\ ] = 1$．

ここで c は $2J(c^{-1}) = 1$ の唯一の正根．

c の値が，最初に掲げた公式とこの問題の関係を暗示する．実際，以上のことが公式を用いて証明される．但し，類似の，より詳しい確率論の議論が必要になるので，この話はしない．

どこが難しそうか？

公式に戻る．まず x が 1 より大きいか小さいかで大きく関数形が変わる．このほか，

- $\displaystyle\lim_{n\to\infty} G_n(x) = 0$, $x \geq 0$,
- $\displaystyle\lim_{x\to\infty} G_n(x) = 1$, $n \geq 0$,

は定義から容易に分かる．従って，x と n を同時に大きくしたらどうなるかが微妙であることが分かる．

[11] L. Devroye, A note on the height of binary search trees, Journ. Assoc. Comput. Machinery, **33** (1986) 489–498.

3.4.3 ある解き方．ー大偏差値原理ー

G_n は分布関数の持つべき性質を満たしていることに注意．実際，$G = G_1$ は指数分布 $P(X \in A) = \int_{A\cap[0,\infty)} e^{-x}\,dx$ の分布関数である．このとき

$$P(X \le x) = (1-e^{-x})\mathbf{1}_{x\ge 0} = G(x)\,.$$

さらに帰納法で $G_{n+1} = G * G_n = G_n * G$ が言える．実際，

$$G * G_n(x) = \int_{0\le y\le x} e^{-(x-y)} e^{-y} \sum_{k=n}^{\infty} \frac{y^k}{k!}\,dy = e^{-x}\sum_{k=n}^{\infty}\frac{x^{k+1}}{(k+1)!} = G_{n+1}(X)\,.$$

このことは，G_n がそれぞれ指数分布に従う独立な確率変数 X_i, $i = 1, 2, \cdots, n$, の和の分布関数になっていること

$$G_n(x) = G * G_{n-1}(x) = P(\sum_{i=1}^{n} X_i \le x)$$

を表す．そこでこの事実を鍵にして，問題の公式を独立確率変数列に関する極限定理という立場で解こう．

初めに上と同じ X に対して

$$K(\theta) = \log \mathrm{E}[\ e^{-\theta X}\ ],\ \theta > 0,$$

とおく．上と同じ $\{X_i\}$，即ち，X と同分布の i.i.d.，に対して，

$$e^{nK(\theta)} = \mathrm{E}[\ \exp\left(-\theta\sum_{i=1}^{n} X_i\right)\ ] = \int_{\mathbf{R}} e^{-\theta\, x}\frac{dG_n}{dx}(x)\,dx$$

が期待値と独立性，分布関数の関係から言えることに注意．

J との関係は次の事実に集約される．

$$J(x) = \inf_{\theta>0} \mathrm{E}[\ e^{-\theta\,(X-x)}\ ] = \exp\left(\inf_{\theta>0}(K(\theta)+\theta x)\right),\ x \ge 0\,.$$

証明は

$$f(\theta) = \log \mathrm{E}[\ e^{-\theta\,(X-x)}\ ] = \theta x - \log(\theta+1)$$

の増減表を書けば

$$\inf_{\theta>0}\exp(f(\theta)) = \begin{cases} \exp(f(x^{-1}-1)) = xe^{1-x}, & 0 \le x < 1, \\ \exp(f(+0)) = 1, & x \ge 1\,. \end{cases}$$

次の主張が導ければ公式が導けることになった：

$$\psi(x) = \lim_{n\to\infty}\frac{1}{n}\log G_n(nx),$$

とおくとき，

$$\psi(x) = \inf_{\theta>0}(K(\theta)+\theta\, x),\ x > 0\,.$$

Superadditivity の議論から，ψ が存在し，しかも

$$\psi(x) = \sup_{n\ge 1}\frac{1}{n}\log G_n(nx)$$

となる．$x > 0$ なら $G(x) > 0$ だから $(\infty >)\psi(x) > -\infty$．さらに ψ が上に凸であることも言える．実際 $\alpha+\beta = 1$ なる正数 α, β と $N \in \mathbf{N}$ に対して $n = N\alpha$, $m = N\beta$ とおいて superadditivity を使えば

$$\frac{1}{N}\log G_N(N(\alpha x+\beta y)) \ge \alpha\frac{1}{n}\log G_n(nx) + \beta\frac{1}{m}\log G_m(mx)$$

となるので，$N \to \infty$ を取ればよい．

$\psi(x) \le \inf_{\theta>0}(K(\theta)+\theta\, x)$ は容易．実際 Markov の不等式から任意の $\theta>0$, $n\in\mathbf{N}$, $x\ge 0$, に対して

$$\begin{aligned} K(\theta)+\theta x &= \frac{1}{n}\log\left[\mathrm{E}[\ \exp(n\theta\,(x-\frac{1}{n}\sum_{i=1}^{n}X_i))\ ]\right] \\ &\ge \frac{1}{n}\log\left[P(x\ge \frac{1}{n}\sum_{i=1}^{n}X_i)\right] = \frac{1}{n}\log G_n(nx) \end{aligned}.$$

$n\to\infty$ として右辺は $\psi(x)$ になる．そして $\theta>0$ に関して極小をとれば求める不等式になる．

$\psi(x) \ge \inf_{\theta>0}(K(\theta)+\theta\, x)$ を証明する．$n\in\mathbf{N}$ をとる．最初に注意したことと部分積分から

$$\begin{aligned} e^{nK(\theta)} &= \int_{\mathbf{R}} e^{-\theta\, x}\frac{dG_n}{dx}(x)\,dx = \theta\int_{\mathbf{R}} e^{-\theta\, x}G_n(x)\,dx = \theta\int_0^\infty e^{-\theta\, x}G_n(x)\,dx \\ &\le \theta\int_0^\infty \exp\left(n\psi(\frac{x}{n})-\theta\, x\right)\,dx = n\theta\int_0^\infty \exp(n\psi(y)-n\theta\, y)\,dy\,. \end{aligned}$$

$G\le 1$ なので $\psi\le 0$，よって

$$\int_b^\infty \exp(n\psi(y)-n\theta\, y)\,dy \le \int_b^\infty \exp(-n\theta\, y)\,dy \le (n\theta)^{-1}\exp(-nb\theta)\,.$$

他方，$\psi\uparrow$ だから

$$\int_c^{c'} \exp(n\psi(y)-n\theta\, y)\,dy \ge (c'-c)\exp(n\psi(c)-n\theta\, c')\,.$$

$c'>c$ ならば各 $\theta>0$ 毎に十分大きい b に対して n に無関係に

$$2\int_c^{c'} \exp(n\psi(y)-n\theta\, y)\,dy \ge \int_b^\infty \exp(n\psi(y)-n\theta\, y)\,dy$$

とできる．従って，ある $b\ge 0$ に対して

$$e^{nK(\theta)} \le 2n\theta\int_0^b \exp(n\psi(y)-n\theta\, y)\,dy \le 2n\theta b\exp[n\sup_{y\ge 0}(\psi(y)-\theta\, y)]$$

となる．(b で切ったのは $\int_0^b 1\,dy=b$ を有限にするためだけ.）b は n によらないから

$$K(\theta) \le \sup_{y\ge 0}(\psi(y)-\theta y),\ \theta>0\,.$$

$x>0$ で ψ は有限で上に凸．$y\ne x$ に対して y と x を結ぶ直線の傾きは y について単調減少だからその $y\uparrow x$, $y\downarrow x$ の極限として $\psi'(x-0)$, $\psi'(x+0)$ がそれぞれ有限確定する．ψ は単調増加だから $\psi'(x)\ge 0$, $x>0$, しかもいま述べたことから（広義）単調減少．与えられた $x>0$ に対して $\psi'(x-0)\ge\theta_x\ge\psi'(x+0)$ なる $\theta_x\ge 0$ が取れる．

$\theta_x>0$ ならば

y	0		x		∞
$\psi'(y)-\theta_x$	$(\le\infty)$	$+$		$-$	$(\ge -\theta_x)$
$\psi(y)-\theta_x y$		$\nearrow$	$\psi(x)-\theta_x x$	$\searrow$	

故に

$$K(\theta_x) \le \sup_{y\ge 0}(\psi(y)-\theta_x y) \le \psi(x)-\theta_x x\,.$$

$\theta_x=0$ でも $K(\theta)$ が $\theta=0$ で連続なので

$$K(0)=\lim_{\theta\downarrow 0}K(\theta) \le \lim_{\theta\downarrow 0}\sup_{0\ge 0}(\psi(y)-\theta y)\,.$$

$\psi'(x-0)\ge\theta_x=0\ge\psi'(x+0)\ge 0$ だから

y	0		x		∞
$\psi'(y)$	$(\le\infty)$	$+$		0	
$\psi(y)$		$\nearrow$	$\psi(x)$	$\rightarrow$	

故に $\psi(y) \le \psi(x)$, $y \ge 0$, となって $K(\theta_x) = K(0) \le \lim_{\theta\downarrow 0} \sup_{y\ge 0}(\psi(x) - \theta y) = \psi(x)$.

いずれにしても $\inf_{\theta>0}(K(\theta) + \theta x) \le K(\theta_x) + \theta_x x \le \psi(x)$ だから，証明は終わった．

iid のサンプル平均 $\dfrac{1}{n}\sum_{i=1}^{n} X_i$ の平均値 $\mathrm{E}[\,X\,]$ からのずれ $G_n(nx) = P(\dfrac{1}{n}\sum_{i=1}^{n} X_i \le x)$ を調べるのに，X の母関数 K の θ についての極小値を求める問題に帰着させている．この解法（を無限次元の場合に拡張したもの）を大偏差値原理と呼ぶ．中心極限定理によれば，収束するのは $\dfrac{1}{\sqrt{n}}(\sum_{i=1}^{n} X_i - \mathrm{E}[\,X\,])$ だから，n が大きいとまれにしか起こらない事象を調べていることからこの名前が付いた．

ところで，公式から，より簡単な事実として，

$$\lim_{n\to\infty} G_n(nx) = \begin{cases} 0, & 0 \le x < 1, \\ 1, & x \ge 1, \end{cases}$$

を得る．これが確率論的には大数の法則に対応する．

問題．

Stirling の公式などを利用して公式を直接証明してみよ．

3.5 極限定理のいろいろ．

先ほどの公式は

$$\psi(x) = \lim_{n\to\infty} \frac{1}{n} \log G_n(nx) = \inf_{\theta>0}(K(\theta) + \theta\, x)\,,\ x \ge 0\,.$$

に帰着されることを見た．指数分布に従う確率変数 X と等しい分布を持つ独立同分布確率変数列 X_i, $i \in \mathbf{N}$, に対して $G_n(x) = P\left(\sum_{i=1}^{n} X_i \le x\right)$ および $K(\theta) = \log \mathrm{E}[\,e^{-\theta\,X}\,]$ であったから，$S_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ とおくと，

$$\lim_{n\to\infty}[P(S_n/n \le x)]^{1/n} = \inf_{\theta>0} \mathrm{E}[\,e^{\theta\,(x-X)}\,]$$

から導かれたことになる．この式は X が指数分布以外でも広く成り立つ．大偏差値原理と呼ばれる手法の原型になる公式である．独立同分布確率変数列の和 $S_n = \sum_{i=1}^{n} X_i$ （あるいは平均 S_n/n）は実験や観測データから本来の分布を推定したり，逆に複雑な現象の累積効果を予測する場合など，非常に重要な応用上の意味もある．

大偏差値原理は一般論のうちで精密な難しいほうの結果である．よりやさしい結果が先に知られている．(有名なので，ここでは省略してきた．結果のみ.）

以下 X_i, $i \in \mathbf{N}$, は独立確率変数列で可積分な X と同じ分布とする．$X - \mathrm{E}[\,X\,]$ を考えればよいので，$\mathrm{E}[\,X\,] = 0$ としておく．また， $S_n = \sum_{i=1}^{n} X_i$ とする．

大数の法則． $\mathrm{E}[\,|X|\,] < \infty$ ならば $\lim_{n\to\infty} \dfrac{1}{n} S_n = 0$, a.s., かつ in $\mathcal{L}^1$.

中心極限定理． $\sigma^2 = \mathrm{V}[\,X\,] < \infty$ ならば， $\lim_{n\to\infty} P(\dfrac{1}{\sigma\sqrt{n}} S_n \le x) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-y^2/2)\, dy$.

重複対数の法則． $\sigma^2 = \mathrm{V}[\ X\ ] = 1$ ならば,

$$\limsup_{n\to\infty} \frac{1}{\sqrt{2n\log\log n}} S_n = +1,\ a.s.,\ \text{かつ}\ \liminf_{n\to\infty} \frac{1}{\sqrt{2n\log\log n}} S_n = -1,\ a.s..$$

(Strassen)． $\sigma^2 = \mathrm{V}[\ X\ ] = 1$ とする．S_n を線型内挿で連続時刻 S_t に拡張して，$Z_n(t) = \dfrac{1}{\sqrt{2n\log\log n}} S_{nt}$, $0 \le t \le 1$, とおく．

集合値確率変数 K を，$t \in [0,1]$ で $f(t)$ に一様収束する $Z_n(t)$ の部分列がとれる関数 f からなる集合とする（path の極限形の集合）．$K' = \{f \in C[0,1] \mid \exists h;\ f(t) = \int_0^t h(s)\,ds,\ \int_0^1 h(s)^2 \le 1\}$ とすると，$P(K = K') = 1$.

$\lim_{n\to\infty} Y_n = X$, a.s., は $P(\{\omega \in \Omega \mid \lim_{n\to\infty} Y_n(\omega) = X(\omega)\}) = 1$ のこと．括弧内が可測集合であることは既に述べた．$\lim_{n\to\infty} Y_n = X$, in $\mathcal{L}^1$, は $\lim_{n\to\infty} \mathrm{E}[\ |Y_n - X|\ ] = 0$. 大数の法則は $K = \sup_{n\ge 0} \mathrm{E}[\ {X_n}^4\ ] < \infty$ のとき証明が容易 (同分布の必要もない)．$\mathrm{E}[\ {X_n}^4\ ]$ が可積分だから $\mathrm{E}[\ {X_n}^2\ ]$,$\mathrm{E}[\ {X_n}^3\ ]$ も可積分．Schwarz から $(\mathrm{E}[\ {X_n}^2\ ])^2 \le \mathrm{E}[\ {X_n}^4\ ] \le K$. 以上と独立性と $\mathrm{E}[\ X\ ] = 0$ から $\mathrm{E}[\ \sum_{n=1}^{\infty} \left(\frac{S_n}{n}\right)^4\ ] \le 3K \sum_{n=1}^{\infty} n^{-2} < \infty$. 故に $\sum_{n=1}^{\infty} \left(\frac{S_n}{n}\right)^4 < \infty$, a.s.. よって $\lim_{n\to\infty} S_n/n = 0$, a.s..

中心極限定理のように分布関数が不連続点を除いて収束することを弱収束と呼ぶ．この定理を証明するには，特性関数 $\mathrm{E}[\ \exp(\frac{1}{\sigma\sqrt{n}} S_n \le x)\ ]$ を考える．

重複対数の法則は大偏差値の領域である．Stroock は Brown 運動の場合に Strassen の定理を (再) 証明した．

3.6　独立性の implication　（省略可能？）．

独立性は σ 加法族の独立性の定義に帰着される．

(1) σ 加法族の列 $\mathcal{F}_k$, $k = 1,2,\cdots,n$, が独立とは任意の部分列 $i_1, i_2, \cdots, i_m$ と任意の $F_{i_k} \in \mathcal{F}$, $k = 1,2,\cdots,m$, に対して $P\left(\bigcap_{k=1}^{m} F_{i_k}\right) = \prod_{k=1}^{m} P(F_{i_k})$ が成り立つこと．このとき，確率変数 X_k, $k = 1,2,\cdots,n$, が独立ということと $\sigma(X_k) = \{\{X_k \le x\} \mid x \in \mathbf{R}\}$, $k = 1,2,\cdots,n$, が独立ということは同値.

(2) 事象 $E_k \in \mathcal{F}$, $k = 1,2,\cdots,n$, が独立とは任意の部分列 $i_1, i_2, \cdots, i_m$ に対して $P\left(\bigcap_{k=1}^{m} E_{i_k}\right) = \prod_{k=1}^{m} P(E_{i_k})$ が成り立つこと．このことと $\{\emptyset, E_k, {E_k}^c, \Omega\}$, $k = 1,2,\cdots,n$, が独立であることは同値.

(3) E_k, $k \in \mathbf{N}$, が独立で $\sum P(E_n) = \infty$ ならば $P(E_n,\ i.o.) = P(\limsup E_n) = 1$　（第 2 Borel–Cantelli Lemma）.

(4) 確率変数列 X_n, $n \in \mathbf{N}$, に対して $\bigcap_{n=1}^{\infty} \sigma(X_{n+1}, X_{n+2}, \cdots)$ に含まれる集合を末尾事象と呼び，末尾事象の作る σ 加法族に関して可測な関数を末尾関数と呼ぶ．独立確率変数列でなくて良い.

末尾事象の例：$\{\exists \lim_{n\to\infty} X_n\}$, $\{\sum_{n=1}^{\infty} X_n < \infty\}$.

末尾関数の例：$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k$.

(5) 独立確率変数列 X_n, $n \in \mathbf{N}$, の末尾 σ 加法族を $\mathcal{T}$ とすると，$F \in \mathcal{T}$ ならば $P(F) = 0$ または $P(F) = 1$. ξ が末尾関数ならば $P(\xi = c) = 1$ となる定数 $c \in [-\infty, \infty]$ がある．即ち，決定論的になる.（Kolmogorov の 0–1 法則）.

(6) BP の注意 Williams §4.11.

4 マルチンゲール．(木曜日)

4.1 条件付き期待値．

$(\Omega, \mathcal{F}, P)$ ：確率空間，$X : \Omega \to \mathbf{R}$ ：可積分 $(\mathrm{E}[\,|X|\,] < \infty)$ な確率変数，$\mathcal{B} \subset \mathcal{F}$ ：σ 加法族．

定義によれば，条件付き期待値とは関数（確率変数）をならした関数（確率変数）を考えること．例えば，2 変数関数 $f(x,y)$ において y について積分した関数 $g(x) = \int f(x,y)\,dy$ を考えることである．

確率変数をならした確率変数を考える理由は部分的な情報を取り入れる，ということである．基本的に確率論で X の値が定数でない（Ω の関数である）理由は制御不能な攪乱によるものであった．ある意味で前もって予測できるのは期待値 $\mathrm{E}[\,X\,]$ しかない．もっと一般に $\mathrm{E}[\,X^n\,]$ や分布 $P(X \in A)$ も分かるが，実際にどの値が実現するかは分からない．

ところが，トランプゲームで一部のカードを表にする場合や，時々刻々変化する確率現象（確率過程と呼ぶ）においてある時刻まで観測した場合は，部分的に実現した内容についての情報がある．実際に実現すればそれは確定したとしてそこから進めばよいが，前もって，そのときの状況に対処する作戦を立てる場合には，実現した部分までの状況についての関数を考えておくべきである．残りの部分は未定だから，残りの部分に関して期待値をとったもの，つまり，実現部分 $\mathcal{B}$ について可測な関数であって，残りに関してならしたものを考えたい．

定義を正確に言うと，g を改めて $\mathbf{R^2}$ 上の関数，ただし，y について定数関数，と見たものを，f の $\mathcal{B} \times \mathbf{R}$ に関する条件付き期待値と呼ぶ．$\mathcal{B}$ は残った変数 $x \in \mathbf{R}$ の空間の σ 加法族．このような,「ならされた関数」は直感的にはいつでもありそうに見えるが，実際，非常に一般的に存在することがきちんと証明されている．数学的にどのようなことが保証されているか見ておこう．

4.1.1 定理．

X の $\mathcal{B}$ に関する条件付き期待値，即ち，可積分な $\mathcal{B}$–可測確率変数 $Y : \Omega \to \mathbf{R}$ $(\mathrm{E}[\,|Y|\,] < \infty,\ \{\{Y \le x\} \mid x \in \mathbf{R}\} \subset \mathcal{B})$ で，任意の $B \in \mathcal{B}$ に対して $\mathrm{E}[\,Y; B\,] = \mathrm{E}[\,X; B\,]$ を満たすもの，が存在する．$\tilde{Y}$ も条件付き期待値ならば $Y = \tilde{Y}$, a.s., という意味で一意的である．

条件付き期待値（の一つ）を $\mathrm{E}[\,X \mid \mathcal{B}\,]$ と書く．即ち，条件付き期待値 $\mathrm{E}[\,X \mid \mathcal{B}\,]$ は，$\mathrm{E}[\,\mathrm{E}[\,X \mid \mathcal{B}\,]; A\,] = \mathrm{E}[\,X; A\,]$, $A \in \mathcal{B}$, と，$\{\mathrm{E}[\,X \mid \mathcal{B}\,] \le x\} \in \mathcal{B}$ で特徴づけられる確率変数である．

すぐ分かること.（いずれも二つの性質を満たすことを直接確かめられる.）

(1) $\mathrm{E}[\,X \mid \{\emptyset, \Omega\}\,] = \mathrm{E}[\,X\,]$．即ち，定数関数になる．
(2) $\mathrm{E}[\,X \mid \mathcal{F}\,] = X$．

注．

(1) 標準的な証明法は Radon–Nikodým によるもの．Williams は §9.5 で $\mathcal{L}^2$ 理論に基づく証明を与えている．

(2) 直積測度空間 $(\Omega, \mathcal{F}, P) = (\Omega_1 \times \Omega_2, \mathcal{B}_\infty \times \mathcal{B}_\in, P_1 \times P_2)$ に対して，$\mathcal{B} = \{B_1 \times \Omega_2 \mid B_1 \in \mathcal{B}_\infty\} \subset \mathcal{F}$ に関する条件付き期待値は本質的に Fubini の定理である．即ち，$\mathrm{E}[\,X \mid \mathcal{B}\,](\omega) = \int_{\Omega_2} X(\omega_1, \omega_2)\,dP_2(\omega_2)$．$X$ が $\mathcal{B}$–可測関数のとき右辺が $\mathcal{B}_\infty$–可測関数になることは Fubini の定理の一部．これを Ω 上の関数（ω_2 に関して変化しない）と見れば $\mathcal{B}$–可測関数になる．
条件 $\mathcal{B}_\infty$ を情報とみなすと，条件付き期待値は $\mathcal{B}_\infty$ に関する X の情報のみを残して残り (Ω_2) についてならした関数である．

(3) 直積測度空間上で密度（非負可測で積分が 1）ρ を持っている確率測度

$$(\Omega, \mathcal{F}) = (\Omega_1 \times \Omega_2, \mathcal{B}_\infty \times \mathcal{B}_\in),\ P(A) = \int_A \rho(\omega_1, \omega_2)\,dP_1(\omega_1) dP_2(\omega_2)$$

の場合も Fubini の定理である．即ち $\mathrm{E}[\,X \mid \mathcal{B}\,] = \int_{\Omega_2} X(\omega_1, \omega_2)\,\rho(\omega_1, \omega_2)\,dP_2(\omega_2)$．

(4) 分割 $\Omega = \sum_{i=1}^{N} B_i,\ B_i \in \mathcal{F}$, が生成する σ 加法族 $\mathcal{B} = \sigma(\{B_1, \cdots, B_N\})$ の場合．$\omega \in B_i \in \mathcal{B}$ のとき $\mathrm{E}[\,X \mid \mathcal{B}\,](\omega) = \dfrac{\mathrm{E}[\,X; B_i\,]}{P(B_i)}$ で与えられる関数になる．即ち，各 B_i 上で定数の単関数である．

いずれも条件付き期待値になることは直接確かめられる．

4.1.2 独立確率変数による条件付き期待値．

$\mathrm{E}[\,X \mid Y\,] = \mathrm{E}[\,X \mid \sigma(Y)\,]$ で確率変数に関する条件付き期待値を定義する．

$\{X_i \mid i = 1, 2, \cdots, n\}$ と $(\mathbf{R^n}, \mathcal{B}^n)$–可測関数 f に対して $g(x) = \mathrm{E}[\,f(x, X_2, \cdots, X_n)\,]$ とおくと $g(X_1) = \mathrm{E}[\,f(X_1, X_2, \cdots, X_n) \mid X_1\,]$．つまり，$X_1$ だけ残して，残りについて積分し，結果を X_1 で書いたものが条件付き期待値.

証明は，$\{X_i\}$ の分布が直積測度に関して密度を持っている場合（独立な場合を典型例として）は Fubini. 一般には f が矩形集合の支持関数の場合に計算し，あとは与式が成り立つ集合族を考える．

例．さいころ 1 回投げ.

$$\Omega = \{(\omega) \mid \omega \in \{1, 2, \cdots, 6\}\},\ \mathcal{F} = 2^{\Omega},\ P(A) = 6^{-1}\sharp A\,.$$

それぞれ 1, 2 が出る事象の支持関数

$$X_1(\omega) = \begin{cases} 1 & \omega = 1 \\ 0 & \omega \neq 1 \end{cases},\quad X_2(\omega) = \begin{cases} 1 & \omega = 2 \\ 0 & \omega \neq 2 \end{cases}.$$

$Z = \mathrm{E}[\,X_2 \mid X_1\,] = \mathrm{E}[\,X_2 \mid \sigma(X_1)\,]$ を計算してみる．$\sigma(X_1) = \{\emptyset, B, B^c, \Omega\}$; $B = \{1\}$，は $\{X_1 \le x \mid x \in \mathbf{R}\} = \{\emptyset, B^c, \Omega\}$ から．

Z は $\sigma(X_1)$–可測なので B 上定数（$\sigma(X_1) = \sigma(\{B, B^c\})$ で $\{B, B^c\}$ が分割であることに注意）．$\dfrac{1}{6}Z(1) = \mathrm{E}[\,Z; B\,] = \mathrm{E}[\,X_2; B\,] = \dfrac{1}{6}X_2(1) = 0$, $\dfrac{1}{6}(Z(1) + 5Z(2)) = \sum_{i=1}^{6} \dfrac{1}{6}Z(i) = \mathrm{E}[\,Z\,] = \mathrm{E}[\,X_2\,] = \dfrac{1}{6}$, だから,

$$Z(\omega) = \mathrm{E}[\,X_2 \mid X_1\,](\omega) = \begin{cases} 0 & \omega = 1 \\ \frac{1}{5} & \omega \neq 1 \end{cases} = \frac{1}{5}(1 - X_1(\omega))\,.$$

よって X_1 に関する条件付き期待値は X_1 で書ける：

$$\mathrm{E}[\,X_2 \mid X_1\,] = \frac{1}{5}(1 - X_1)\,.$$

4.1.3 性質．

(1) $\mathrm{E}[\,\mathrm{E}[\,X \mid \mathcal{B}\,]\,] = \mathrm{E}[\,X\,]$.

(2) $\{X \le x \mid x \in \mathbf{R}\} \subset \mathcal{B}$（即ち，$X$ が $\mathcal{B}$ 可測）ならば $\mathrm{E}[\,X \mid \mathcal{B}\,] = X$, a.s.. 特に X が定数なら $\mathrm{E}[\,X \mid \mathcal{B}\,] = X$, a.s..

(3) $\mathrm{E}[\,aX + bY \mid \mathcal{B}\,] = a\mathrm{E}[\,X \mid \mathcal{B}\,] + b\mathrm{E}[\,Y \mid \mathcal{B}\,]$, a.s.　（線型性）．

(4) $x \ge 0$ ならば $\mathrm{E}[\,X \mid \mathcal{B}\,] \ge 0$　（正値性）．

(5) 単調収束定理，Fatou の補題（$X_n \ge 0$ のとき），dominated convergence theorem，Jensen，は全て期待値を条件付き期待値に置き換えても成立．但し，後者は確率変数なので，a.s. に成立，と書き換える．特に Jensen と (2) から $\|\mathrm{E}[\,X \mid \mathcal{B}\,]\|_p \le \|X\|_p$.

(6) $\mathcal{C} \subset \mathcal{B}$ も σ 加法族ならば $\mathrm{E}[\,\mathrm{E}[\,X \mid \mathcal{B}\,] \mid \mathcal{C}\,] = \mathrm{E}[\,X \mid \mathcal{C}\,]$, a.s.　（tower property）．

(7) Z が $\mathcal{B}$–可測で有界 ($|Z| \le M$, a.s.,) ならば $\mathrm{E}[\,ZX \mid \mathcal{B}\,] = Z\mathrm{E}[\,X \mid \mathcal{B}\,]$, a.s.　（「既知情報は定数扱い」）．

$p > 1$, $p^{-1} + q^{-1} = 1$, $X \in \mathcal{L}^p(\Omega, \mathcal{F}, P)$, $Z \in \mathcal{L}^q(\Omega, \mathcal{B}, P)$, ならば同じ結果が成り立つ．

$X \geq 0$ が $\mathcal{F}$–可測，$Z \geq 0$ が $\mathcal{B}$–可測，$\mathrm{E}[\ X\ ] < \infty$, $\mathrm{E}[\ ZX\ ] < \infty$, ならば同じ結果が成り立つ．

(8) $\mathcal{C} \perp \sigma(\sigma(X), \mathcal{B})$ ならば $\mathrm{E}[\ X \mid \sigma(\mathcal{B}, \mathcal{C})\ ] = \mathrm{E}[\ X \mid \mathcal{B}\ ]$, a.s.　（「独立性との関係」）．特に X が $\mathcal{C}$ と独立ならば $\mathrm{E}[\ X \mid \mathcal{C}\ ] = \mathrm{E}[\ X\ ]$, a.s.

例．

$\{X_i \mid i = 1, 2, \cdots, n\}\perp$ が X と同分布の i.i.d. で $\mathrm{E}[\ |X|\ ] < \infty$ とし，$S_n = x_1 + \cdots + X_n$ とおいて，$\mathcal{B}_n = \sigma(S_n, S_{n+1}, S_{n+2}, \cdots) = \sigma(S_n, X_{n+1}, X_{n+2}, \cdots)$ とおく．

$\sigma(X_{n+1}, X_{n+2}, \cdots) \perp \sigma(X_1, S_n) \subset (X_1, \cdots, X_n)$ なので (8) より $\mathrm{E}[\ X_1 \mid \mathcal{B}_n\ ] = \mathrm{E}[\ X_1 \mid S_n\ ]$．対称性から $\mathrm{E}[\ X_1 \mid S_n\ ] = \cdots = \mathrm{E}[\ X_n \mid S_n\ ]$ なので線型性から $\mathrm{E}[\ X_1 \mid \mathcal{B}_n\ ] = \frac{1}{n}\mathrm{E}[\ S_n \mid S_n\ ] = \frac{1}{n}S_n$．最後の変形で (2) を用いた．

4.1.4　条件付き確率　（省略可能？）．

$P(A \mid \mathcal{B}) = \mathrm{E}[\ \mathbf{1}_A \mid \mathcal{B}\ ]$．確率変数である．

$\omega \in B \in \mathcal{B}$ のとき，$P(A \mid \mathcal{B})(\omega) = \mathrm{E}[\ \mathbf{1}_A \mid \mathcal{B}\ ](\omega)$．

分割 $\Omega = \sum_{i=1}^{N} B_i$, $B_i \in \mathcal{F}$, が生成する σ 加法族 $\mathcal{B} = \sigma(\{B_1, \cdots, B_N\})$ の場合，$\omega \in B_i \in \mathcal{B}$ のとき $P(A \mid \mathcal{B})(\omega) = \dfrac{\mathrm{E}[\ \mathbf{1}_A; B_i\ ]}{P(B_i)} = \dfrac{P(A \cap B_i)}{P(B_i)} = P(A \mid B_i)$ で素朴な条件付き確率を再現する．即ち，条件付き確率 $P(A \mid \mathcal{B})(\omega)$ は各 B_i 上で定数の単関数であり，その定数は $\mathcal{B}$ を生成する分割 $\{B_i\}$ の中で ω で起きた事象 B_i の下での素朴な条件付き確率 $P(A \mid B_i)$ である．

Version の問題があるので，$P(\cdot \mid \mathcal{B})$ が確率測度値確率変数になる（ω を固定したとき，可算加法性を全面的に満たす version がある）とは限らない．あるときに regular conditional probability と呼ぶ．極めて technical な内容で省略.

4.2　Filtration と adapted な確率変数列．

これまで確率空間 $(\Omega, \mathcal{F}, P)$ の上の確率変数 $X : \Omega \to \mathbf{R}$ とは $\mathcal{F}$–可測な関数，即ち，$\{X \leq x\} \in \mathcal{F}$, $x \in \mathbf{R}$, を満たす関数，のことを考えてきた．条件付き期待値のように部分 σ 加法族–可測な関数，即ち，$\mathcal{B} \subset \mathcal{F}$ なる σ 加法族 $\mathcal{B}$ に対して $\{Y \leq x\} \in \mathcal{B}$, $x \in \mathbf{R}$ なる関数 Y を X と同時に（対にして）考える場合がある．もっと一般に，ここから先は，σ 加法族の列 $\{\mathcal{F}_n\}$ を考える．イメージは時刻 $n = 0, 1, 2, \cdots$ に対して，時刻 n までで知り得た情報を $\mathcal{F}_n$ とおく．

このイメージに合うのは情報が増大していく場合，即ち，$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}$ が成り立つ場合である．このようになっている σ 加法族の列を filtration と呼ぶ．通常 σ 加法族の列を考察する場合は filtration を考える．以下でも断らなければ $\{\mathcal{F}\}$ は filtration とする．Filtration に対して「最終的な情報」を $\mathcal{F}_\infty = \bigcup_{n=0}^{\infty} \mathcal{F}_n$ とおく．条件付き期待値のように filtration を考えるのは通常それに対応した確率変数の列（chain, 離散時間確率過程 discrete time process）を考えるためである．確率変数列 X_n, $n = 0, 1, 2, \cdots$, が filtration $\{\mathcal{F}_n\}$ に adapted とは，各 n に対して X_n が $\mathcal{F}_n$–可測，即ち，$\{X_n \leq x \mid x\mathbf{R}\} \subset \mathcal{F}_n$ であることを言う．

典型的なイメージとしては，賭事（ゲーム）において X_n が n 回目の試合の結果としての累積配当（$X_n - X_{n-1}$ が単位掛け金あたりの配当になっている）を表し，$\mathcal{F}_n$ は n 回目終了までのゲームの流れ全体を指定した事象（$n+1$ 回目以降のいろんな可能性を全部集めたものという意味で集合）の集合である．累積配当の和という概念に実質的な意味はないが，X_i, $i = 1, 2, \cdots, n$, が時刻 n までのゲームの全記録の情報を持っているならば条件を付ける上では等価である．ゲームに永久に参加するなら，自分の全財産を X_n とおいても同じ情報だが，後で，ゲームから離脱する場合を考えるので，上のように置いた．

明らかに，$\{\sigma(X_0,\cdots,X_n)\mid n=0,1,2,\cdots\}$ は filtration（配当記録だけを情報としている σ 加法族）であり，$\{X_n\}$ はこれに adapted である．

一般に n 変数 $\mathcal{B}_n$–可測関数 f_n たちに対して $W_n = f_n(X_0,\cdots,X_{n-1})$ とおけば，$\{W_n\}$ も adapted. Filtration が与えられたとき，これに adapted な確率変数列としてはこのようなものだけを考えておけばほぼ十分である．

4.3 マルチンゲール．

確率変数列の例としては独立な確率変数列を既に定義した．独立確率変数列の和 S_n や平均の動向 (n dependence) を知りたいことが多いが，$\{S_n\mid n\in\mathbf{Z}_+\}$ は独立確率変数列ではない．しかし，独立確率変数列の足し算で定義されているということの簡単さがこれを取り扱う方法を提供する．このような対象を数学的に扱うのに便利な概念としてマルチンゲールがある．

確率空間 $(\Omega,\mathcal{F},P)$，と filtration $\{\mathcal{F}_n\mid n\in\mathbf{Z}_+\}$ が与えられているとする．確率変数列 $\{X_n\mid n\in\mathbf{Z}_+\}$ がマルチンゲールであるとは

(1) adapted,
(2) 可積分 $\mathrm{E}[\,|X_n|\,]<\infty,\ n\in\mathbf{Z}_+$,
(3) $\mathrm{E}[\,X_n\mid\mathcal{F}_{n-1}\,]=X_{n-1}$,

を満たすことを言う．もちろん最後の性質が最大の特徴である．同様に，確率変数列 $\{X_n\mid n\in\mathbf{Z}_+\}$ が優マルチンゲールであるとは，adapted，可積分，かつ，$\mathrm{E}[\,X_n\mid\mathcal{F}_{n-1}\,]\le X_{n-1}$, を満たすことを言い，確率変数列 $\{X_n\mid n\in\mathbf{Z}_+\}$ が劣マルチンゲールであるとは，adapted，可積分，かつ，$\mathrm{E}[\,X_n\mid\mathcal{F}_{n-1}\,]\ge X_{n-1}$, を満たすことを言う．

条件付き期待値の性質により，劣マルチンゲールならば，$\mathrm{E}[\,X_n\,]\ge\mathrm{E}[\,X_{n-1}\,]$，即ち，期待値は増える．ほかについても同様.

Maritngale ならば $\mathrm{E}[\,X_n-X_{n-1}\mid\mathcal{F}_{n-1}\,]=0$ であるが，これは公平な賭，というイメージを表している．X_n-X_{n-1} が n 回目のゲームの配当率を表す．劣 martingale ならば $\mathrm{E}[\,X_n-X_{n-1}\mid\mathcal{F}_{n-1}\,]\ge 0$.

例：$\{X_n\}$ $\perp$, が可積分で期待値 0，即ち，$\mathrm{E}[\,X_n\,]=0$，ならば $S_n=\displaystyle\sum_{i=1}^{n}X_n$ は martingale，${S_n}^2$ は submartingale.

例２：ξ が可積分とし，$X_n=\mathrm{E}[\,\xi\mid\mathcal{F}_n\,],\ n=0,1,2,\cdots$, とおくと，$\{X_n\}$ は martingale.

テレビの映像手法として，ピンぼけした画面から次第にピントを合わせて最後にくっきりした映像にしたり，web page の装飾として，モザイクをかけた映像から次第にモザイクを細かくしていって最後に本当の絵を表示する方法（インターレース）がある．いろんな絵をランダムに表示させる（例えば日替わりで，あるいは，クイズの連続番組で）ことにすると，平均でぼかしている場合は X_n を表示していることになる．途中の絵の様子から最後の絵を当てる問題などに適用できるだろう．

4.4 マルチンゲール変換．

$\{X_n\}$ が賭やゲームにおける累積配当率を表すとしよう．つまり，X_n-X_{n-1} が n 回目のゲーム終了時の単位掛け金あたりの配当である．

確率変数列 $C_n,\ n=1,2,3,\cdots$, が predictable (previsible) とは C_n が $\mathcal{F}_{n-1}$–可測であること，即ち，$\{C_n\le x\}\in\mathcal{F}_{n-1},\ x\in\mathbf{R}$. C_0 は定義されない．

例えば $\{X_n\}$ が adapted のとき $Y_n=X_{n-1}$ とおけば prdictable になるから，数学的には定義だけの問題とも言える．しかし，気持ちには意味がある．X_n-X_{n-1} を n 回目のゲームの賭の配当率，C_n を n 回目のゲームへの掛け金とすると，掛け金は $n-1$ 回目終了時までのゲームに関する情報 $\mathcal{F}_{n-1}$ で決めるしかないから，C_n は $\mathcal{F}_{n-1}$–可測，即ち predictable である．

このとき，時刻 n （ゲーム終了時点）の全財産を Y_n と書けば，$Y_n = Y_{n-1} + C_n(X_n - X_{n-1})$ となり，adapted な確率変数列になる．しかも， $\{X_n\}$ が martingale で，Y_n が可積分（になるような C_n ）ならば $\{Y_n\}$ も martingale になるし，$\{X_n\}$ が submartingale で $C_n \geq 0$, a.e., $n \in \mathbf{N}$, で可積分性が保証されるならば $\{Y_n\}$ も submartingale である．$Y_0 = 0$ のとき，この Y_n を $Y_n = \sum_{k=1}^{n} C_k(X_k - X_{k-1}) = (C \bullet X)_n$ とおいて，X の C による martingale 変換と呼ぶ．Martingale 変換は確率積分 $\int C\,dX$ の離散類推である．

Martingale の martingale 変換が martingale になる十分条件を正確に書いておく．

定理．

(1) C を非負有界（$0 \leq C_n(\omega) \leq K$, $n \in \mathbf{N}$, $\omega \in \Omega$）な predictable process とする．X が（優；劣）マルチンゲールならば $C \bullet X$ も時刻 0 で 0 の（優；劣）マルチンゲール．

(2) C を有界（$|C_n|(\omega) \leq K$, $n \in \mathbf{N}$, $\omega \in \Omega$）な predictable process とする．X がマルチンゲールならば $C \bullet X$ も時刻 0 で 0 のマルチンゲール．

(3) 上で有界の仮定を $C_n \in \mathcal{L}^2$, $n \in \mathbf{N}$, と $X_n \in \mathcal{L}^2$, $n \in \mathbf{Z}_+$, に置き換えても良い．

可積分性を保証する条件に神経質になるのは，勝ち逃げという反例を除く必要があるからである．Optional stopping で論じる．

4.5　停止時刻．

値に無限大を許す非負整数値確率変数 $T : \Omega \to \mathbf{Z}_+ \cup \{\infty\}$ が stopping time であるとは

$$\{T \leq n\} \in \mathcal{F}_n,\ n \leq \infty,$$

を満たすことを言う．これは

$$\{T = n\} \in \mathcal{F}_n,\ n \leq \infty,$$

と同値．

直感的には T が stopping time とは，ゲームからの離脱時刻ということ．時刻 n 終了時に離脱するかどうかは n までのゲームの情報 $\mathcal{F}_n$ で決まるという意味で $\{T = n\} \in \mathcal{F}_n$ であるべきである．

典型例は，adapted process X とボレル集合 $B \in \mathcal{B} = \sigma(\{[-\infty, a] \mid a \in \mathbf{R}\})$ に対する hitting time $T = \inf\{n \geq 0 \mid X_n \in B\}$．但し， $\inf \emptyset = \infty$．具体的な stopping time としては殆ど hitting time しかない．他方，全ての stopping time についてあることが成り立つと，$\{X_n\}$ に関してある性質が成り立つ，といった非構成的な使い方も多い．

4.6　Doob optional stopping 定理．

補題．

X を (super)martingale, T を stopping time とすると $X^T = \{X_{T\wedge n} \mid n \in \mathbf{Z}_+\}$ も (suer)martingale．特に， $\mathrm{E}[\, X_{T\wedge n}\,] \leq (=) \mathrm{E}[\, X_0\,]$, $n \in \mathbf{Z}_+$．

この補題の重要な点は martingale ということに implicit な条件を除いて可積分性関係の特別な条件を必要としないという点である．

証明は $C_n = \mathbf{1}_{n \leq T} = \begin{cases} 1, & n \leq T, \\ 0, & \text{otherwise}, \end{cases}$ ，即ち，時刻 T まで 1 単位ずつ賭けてそのあとゲームから離脱することを考えると $(C \bullet X)_n = X_{T\wedge n} - X_0 = {X^T}_n - X_0$ となるので，前節の結果から補題を得る．

定理　（省略可能？）.

$\mathrm{E}[\ X_T\ ] = \mathrm{E}[\ X_0\ ]$ は，たとえ $P(T < \infty) = 1$ であっても無条件には保証されない.（つまり補題で $n \to \infty$ と $\mathrm{E}[\ \cdot\ ]$ は無条件では交換できない.）これが成り立つ十分条件を与えるのが optional stopping theorem.

気にしているのは勝ち逃げ戦略．つまり $T = \inf\{n \mid X_n \geq 1\}$ とすると $X_T \geq 1$ on $\{T < \infty\}$ なので，$T < \infty$, a.s., ならば $\mathrm{E}[\ X_T\ ] \geq 1$ が $X_0 = 0$ かつ X_T マルチンゲールでも成り立つので補題 $\mathrm{E}[\ X_{T\wedge n}\ ] = 0$ と比べると交換が保証されないことが分かる．具体的には硬貨投げ (random walk) で X_n を n 回投げた時点で表の数から裏の数を引いた値とすると，表に 1 単位ずつ賭け続けて全財産が正になったらやめる戦略と同じ．これを許せば公平な賭けでも儲けることができる．但し T が有界でないのは現実的ではないので面白くないことが多い.

X を (super)martingale, T を stopping time とすると次のいずれかが成り立てば X_T は可積分でかつ $\mathrm{E}[\ X_T\ ] \leq\ (=)\,\mathrm{E}[\ X_0\ ]$：

(1) T が有界 $(T(\omega) \leq K)$.
(2) $T < \infty$, a.s., かつ X が有界 $(|X_n(\omega)| \leq K)$.
(3) $\mathrm{E}[\ T\ ] < \infty$，かつ $|X_n(\omega) - X_{n-1}(\omega)| \leq K$, $\omega \in \Omega$, $n \in \mathbf{N}$.
(4) supermartingale の場合で $X \geq 0$ かつ $T < \infty$, a.s..

X を martingale, $|X_n(\omega) - X_{n-1}(\omega)| \leq K$, $\omega \in \Omega$, $n \in \mathbf{N}$ とする．C が predictable かつ有界 $(|C(\omega)| \leq K)$, T を可積分 $(\mathrm{E}[\ T\ ] < \infty)$ な stopping time とする．このとき $\mathrm{E}[\ (C \bullet X)_T\ ] = 0$. と $X^T = \{X_{T\wedge n} \mid n \in \mathbf{Z}_+\}$ も (suer)martingale．特に， $\mathrm{E}[\ X_{T\wedge n}\ ] \leq\ (=)\,\mathrm{E}[\ X_0\ ]$, $n \in \mathbf{Z}_+$.

補題　（省略可能？）.

（$\mathrm{E}[\ T\ ] < \infty$ の十分条件.）
T が $\exists N \in \mathbf{N}, \exists \epsilon > 0; (\forall n \in \mathbf{N})\ P(T \leq n+N \mid \mathcal{F}_n) > \epsilon$, a.s., を満たす stopping time ならば $\mathrm{E}[\ T\ ] < \infty$.
これは猿が ABRACADABRA とタイプするまでの時間について $\mathrm{E}[\ T\ ] < \infty$ を意味する.

4.7　例.

4.7.1　ランダムフラクタル.

$[0,1]$ を 3 つの区間に分け，両端の区間を残す．残す区間 I_{11}, I_{12} の長さ $|I_{11}|$, $|I_{12}|$ は，両方ともある確率空間 $(\Omega, \mathcal{F}, P)$ で定義された $[0,1/2)$ に値をとる確率変数で，その分布はそれぞれあらかじめ与えられた確率変数 C_1 と C_2 に等しいとする.

同様に繰り返すことで，第 n 回目の分割の直後に 2^n 個の区間 I_{ni}, $i = 1,\cdots,2^n$ が $[0,1]$ の中に残る．各小区間 I_{ni} を同様に分割して両端の 2 区間 $I_{n+1,2i-1}$, $I_{n+1,2i}$ を残すとき，その長さと $|I_{ni}|$ の比は

$$|I_{mi}|,\ i = 1,2,\cdots,2^m,\ m = 1,\cdots,n,$$

と独立な，$(0,1/2)$ に値をとる確率変数で，その分布は C_1 と C_2 に等しいとする.

$\mathcal{F}_n = \sigma(\{I_{mi} \mid i = 1,2,\cdots,2^m,\ m = 1,\cdots,n\})$ とおけば，これは filtration であり，$|I_{ni}|$ は $\mathcal{F}_n$ 可測で，$\dfrac{|I_{(n+1)(2i-1)}|}{|I_{ni}|}$ は $\mathcal{F}_n$ と独立である.

$1 > \alpha > 0$ を $\mathrm{E}[\ {C_1}^\alpha + {C_2}^\alpha\ ] = 1$ を満たす実数とする．$0 < C_i < 1/2 < 1$ なので左辺は α について単調減少連続関数で，$\alpha = 0$ ならば 2， $\alpha = 1$ ならば $\mathrm{E}[\ C_1 + C_2\ ] < 1$ なので，そのような α がただ一つある. $X_n = \displaystyle\sum_{i=1}^{2^n} |I_{ni}|^\alpha$ とおくと，$\{X_n\}$ は $\mathcal{F}_n$ に関して adapted.

$$\mathrm{E}[\ X_{n+1} \mid \mathcal{F}_n\ ] = \sum_{i=1}^{2^n} \mathrm{E}[\ |I_{(n+1)(2i-1)}|^\alpha + |I_{(n+1)(2i)}|^\alpha \mid \mathcal{F}_n\ ]$$

と書けば,

$$\begin{aligned}\mathrm{E}[\,|I_{(n+1)(2i-1)}|^{\alpha}\mid \mathcal{F}_n\,] &= |I_{ni}|^{\alpha}\mathrm{E}[\,\left(\frac{|I_{(n+1)(2i-1)}|}{|I_{ni}|}\right)^{\alpha}\mid \mathcal{F}_n\,]\\ &= |I_{ni}|^{\alpha}\mathrm{E}[\,\left(\frac{|I_{(n+1)(2i-1)}|}{|I_{ni}|}\right)^{\alpha}\,] = |I_{ni}|^{\alpha}\mathrm{E}[\,{C_1}^{\alpha}\,]\end{aligned}$$

などから

$$\mathrm{E}[\,X_{n+1}\mid \mathcal{F}_n\,] = \sum_{i=1}^{2^n}|I_{ni}|^{\alpha}\mathrm{E}[\,{C_1}^{\alpha}+{C_2}^{\alpha}\,] = \sum_{i=1}^{2^n}|I_{ni}|^{\alpha} = X_n$$

となるので, $\{X_n\}$ はマルチンゲールであることが分かる.(可積分性は $\mathrm{E}[\,|X_n|\,] = \mathrm{E}[\,X_n\,] = \mathrm{E}[\,X_0\,] = 1$ から明らか.)

実はもっと積極的に,劣マルチンゲール収束定理というのがあって, $\sup_{n\ge 1}\mathrm{E}[\,|X_n|\,] = 1 < \infty$ となることから, $X = \lim_{n\to\infty} X_n$, a.s., となる X が存在する(即ち,概収束する).

マルチンゲールの重要な数学的役割は(期待値の何らかの一様性と合わせることによって)確率変数列の極限の存在を示すことである.

4.7.2 文豪ザル.

あるサルがアルファベット大文字をタイプし続ける.毎回 26 字全てを等確率で打つとするとき,平均何回タイプを打てば ABRACADABRA という 11 文字の文字列が出てくるか?

最初の 11 文字がこの文字列になる確率は 26^{-11} だが,だからといって初めてこの文字列が出てくるまで平均 26^{11} 回タイプする,と答えるのは,悪い予想ではないが,正解ではない!

原理的には数え上げで計算できるはずだが,実行すれば大変なことになる.ここは martingale を用いた解法を紹介しよう.

さいころを永久に投げ続ける試行の確率空間のマネをしてタイプをでたらめに永久に打ち続ける確率空間を考える.

$$\begin{aligned}&r = 26,\ \Sigma = \{A, B, \cdots, Z\},\\ &\Omega = \{\omega = (\omega_1, \omega_2, \cdots) \mid \omega_i \in \Sigma\ (i \in \mathbf{N})\},\\ &\mathcal{F}_n = \sigma[\{\{\omega \in \Omega \mid \omega_i = v_i,\ i = 1, 2, \cdots, n\} \mid (v_1, \cdots, v_n) \in \Sigma^n\}]\ \mathcal{F} = \sigma[\bigcup_{n\in\mathbf{N}}\mathcal{F}_n],\\ &n \in \mathbf{N},\ A \in \mathcal{F}_n\ \text{に対して}\ P(A) = r^{-n}\sharp A\,.\end{aligned}$$

さいころのときと同様に,文字列のどれでもいいから n 個分の文字を指定すると,それらの値がその通り生じる確率は 2^{-n} になる,という意味である.

ω_n は n 文字目に猿がタイプする文字とする. $\mathcal{F}_n$ は最初の n 文字 $(\omega_1, \cdots, \omega_n)$ で表せる集合($(\omega_1, \cdots, \omega_n)$ を決めると決まる集合)の全体であり,そういう意味で時刻 n (n 文字目)までの情報を表す. $\{\mathcal{F}_n\}$ は filtration である.

k 文字の単語 $w_1 w_2 \cdots w_k$ が初めて現れる時刻を T とする: $T = \inf\{n \ge k \mid \omega_{n-k+1} = w_1, \omega_{n-k+2} = w_2, \cdots, \omega_n = w_k\}$. T は確率変数になる($\{T \le n\}$ は $(\omega_1, \cdots, \omega_n)$ で表せるので $\mathcal{F}_n$ の要素である). $T < \infty$, a.s., も言える(省略).

$\mathrm{E}[\,T\,]$ を求めたいのであるが,これを解くために次のような方法を採る.

各時刻 $n \ge 1$ に対して,時刻 n (の直前)に,猿の打つ文字の予想を立てる賭けに参加するギャンブラーの時刻 m 財産を ${X^{(n)}}_m$ とする.このギャンブラーは最初 1 単位の財産を持っており,時刻 n に文字 w_1, 以下時刻 $n+k-1$ に文字 w_k までその都度全財産を賭ける.賭けは敗れれば没収され,従って,以後 0 のまま,勝てば財産は r 倍になる.式で書けば $X_0 = 1$ であって,

$${X^{(n)}}_m = \begin{cases} 1, & m < n,\\ r\,{X^{(n)}}_{m-1}\,\mathbf{1}_{\omega_m = w_{m-n+1}}, & n \le m < n+k,\\ {X^{(n)}}_{n+k-1}, & m \ge n+k\,. \end{cases}$$

Note that

$$X^{(n)}{}_{n+k-1} \neq 0 \Longleftrightarrow (\omega_n, \cdots, \omega_{n+k-1}) = (w_1, w_2, \cdots, w_k) \Longrightarrow T \leq n+k-1\,.$$

（定義から明らかに， $0 \leq X^{(n)}{}_m(\omega) \leq r^k$ が任意の n, m, ω に対して成り立つ（つまり，有界）．が，これは使わなくてもよい.）そして $n \leq m < n+k$ のとき

$$\mathrm{E}[\, X^{(n)}{}_m \mid X^{(n)}{}_{m-1} \,] = r\, X^{(n)}{}_{m-1}\, P(\omega_m = w_{m-n+1}) = X^{(n)}{}_{m-1}$$

となるので，martingale である（$m < n$, $m \geq n+k$ では明らか）．従って Doob の optional stopping theorem の最初の補題が使えて，$\mathrm{E}[\, X^{(n)}{}_{T\wedge(n+k-1)} \,] = \mathrm{E}[\, X^{(n)}{}_0 \,] = 1$ となる．

$$a_i = \begin{cases} 1, & w \text{ の最初と最後の } i \text{ 文字が等しいとき}, \\ 0, & \text{otherwise}, \end{cases}$$

とおく． T の定義から

$$X^{(n)}{}_{T\wedge(n+k-1)} = \begin{cases} X^{(n)}{}_{n+k-1} = 0, & T \geq n+k, \\ r^k, & T = n+k-1, \\ r^{T-n+1} a_{T-n+1}\,, & n \leq T < n+k-1, \\ X^{(n)}{}_T = 1, & T < n. \end{cases}$$

従って，

$$1 = P(T = n+k-1) r^k + \sum_{i=1}^{k-1} P(T = n+i-1) a_i r^i + P(T < n)$$

だから

$$P(T \geq n) = P(T = n+k-1) r^k + \sum_{i=1}^{k-1} P(T = n+i-1) a_i r^i.$$

これを用いると

$$\mathrm{E}[\, T \,] = \sum_{n\geq 1} n P(T = n) = \sum_{n \geq 1} P(T \geq n) = P(T \geq k) r^k + \sum_{i=1}^{k-1} P(T \geq i) a_i r^i.$$

最低 k 文字タイプしないと単語は出てこないので $n \leq k-1$ ならば $P(T = n) = 0$，従って $n \leq k$ ならば $P(T \geq n) = 1$ となる．よって

$$\mathrm{E}[\, T \,] = r^k + \sum_{i=1}^{k-1} a_i r^i.$$

ABRACADABRA については $r = 26$, $k = 11$, $a_i = 1$ となるのは $i = 1, 4$，だから

$$\mathrm{E}[\, T \,] = 26^{11} + 26^4 + 26\,.$$

Doob の optional stopping を強化したのが Doob の optional sampling で，確率過程における確率積分（マルチンゲール変換の連続極限）で重要な役割を果たす.

4.7.3 羊たちの沈黙．

結果のみ．D. Williams §15.3 参照．The Mabinogion （Wales の伝説 Peredur ap Efrawg 物語）の中に白と黒の羊の話があるそうだ．各時刻 $1, 2, 3, \cdots,$ に羊が一頭ランダムに鳴く．すると，鳴いた羊の色と反対の色の羊が一頭鳴いた羊と同色に変わる．人は各時刻 $0, 1, 2, \cdots,$ （の直後）に白い羊を何匹でも消すことができる．目的は，全部黒い羊になったとき（それで最終状態）黒い羊の数の期待値を最大にすることである．

このようにランダムな現象に途中で人が手を加えて何かの期待値を最善にしようとする問題を stochastic control と呼ぶ．あらゆる現実の制御は，必ず制御不能な撹乱があるという意味で stochastic control である．マルチンゲール理論を用いることで羊の話について最善の政策が決定できる．

結論（最善の政策）：各時刻毎に黒い羊の方が多い（か黒い羊が残っていない）ときは何もしない．さもなければ，白い羊を黒い羊の数から一頭引いた数に減らす．

この政策の下で黒白 k 頭ずつから始めたときの，最後の黒い羊の数の期待値 E_k は，$\lim_{k\to\infty} E_k - (2k + \frac{1}{4}\pi - \sqrt{\pi k} = 0$ を満たすことが知られている．例えば一万頭ずつから始めたときおおよそ 19824 頭残る．

5 少しだけ，確率過程．(金曜日)

5.1 ランダムウォーク．－確率連鎖－

人生は偶然の連続，という考え方がある.「人生万事塞翁が馬」という言い回しは，人生いつまでたってもその先が予測不能，という意味である．必然性のないようにみえる突然の出来事の連続で，不運（幸運）にも人生がころころ変わる．

ファミコンの RPG (role playing game) は各場面ごとに選択肢があって，一人の主人公についていろいろな冒険が可能である．[12] 各選択場面でさいころを振って選択肢を決めるとすると，ある時刻 t における RPG の主人公の状態（例えば存在位置）を X_t と書いたとき， X_t が各 t で確率変数になっている，といえる．

実数 $t \in \mathbf{R}$ で番号づけられた確率変数の族（集まり） $\{X_t\}_{t\in\mathbf{R}}$ を 確率過程 と呼ぶ．
（これは正確な定義である．もちろん，確率変数を考えるのだから，確率空間 $(\Omega, \mathcal{F}, P)$ が与えられている，とする．）t は実数全体 $\mathbf{R}$ を動かなくてもよい．例えば区間 $[0,1]$ や非負実数区間 $[0,\infty) = \{t \mid t \ge 0\}$ などがよく用いられる．

整数値で（離散的に）番号づけられている場合も stochastic process と呼ぶこともあるが， stochastic chain（確率連鎖）と分けて呼ぶことも多い．この場合は添字を n （には限らないが）にして，$X_n,\ n = 0,1,2,\cdots$, と記すことが多い．独立確率変数やマルチンゲールの話で確率変数列と呼んだものは確率連鎖の例である．添字（番号）t あるいは n を，time parameter （時刻を表す変数）と呼ぶのが普通だが，必ずしも時刻を表さなくてもよい．

例えば：(i) 実験データ $X_n,\ n = 1,2,\cdots$. 大勢の人が同じ実験をやったとき，制御不能な撹乱要因（測定誤差）のために値が異なる場合，これを確率変数とみてデータ処理を行うことが考えられる．n は誰が行った実験かを区別する．(ii) （１次元）画像データ $X_x,\ x \in [a,b)$. データに望まない信号（ノイズ）が加わって本来の画像がランダムに乱される場合，各点 x 毎の値（例えば白黒画像ならば輝度）X_x が確率変数と考えられる．x は空間的な位置を表す変数である．平面画像を扱う場合は $X_{(x,y)}$ のように座標２変数を parameter とする必要が生じうる．この講義ではそこまでは扱えない．

時間変数の値 t を一つ決める毎に X_t は定義によって確率変数 $X_t:\ \Omega \to \mathbf{R}$ である．今まで確率変数は実数値関数としてきたが，実数値でない場合に拡張できる．X_t がある空間 Σ を値域とするとき，

$$X_t:\ \Omega \to \Sigma,$$

Σ をこの確率過程の **状態空間** と呼ぶ．

硬貨を投げて表裏に応じて左右に一こまずつ進むゲームを簡単な例にとって，具体的に確率過程を説明する．さいころを永久に投げ続ける試行と同様に（公平な）硬貨を永久に投げ続ける試行を考える．

$$\Sigma = \{H, T\},\ r = \sharp\Sigma = 2,\ \Omega = \{\omega = (\omega_1, \omega_2, \cdots) \mid \omega_i \in \Sigma\ (i \in \mathbf{N})\},$$
$$\mathcal{F}_n = \sigma[\{\{\omega \in \Omega \mid \omega_i = v_i,\ i = 1,2,\cdots,n\} \mid (v_1,\cdots,v_n) \in \Sigma^n\}]\ \mathcal{F} = \sigma[\bigcup_{n\in\mathbf{N}} \mathcal{F}_n],$$
$n \in \mathbf{N},\ A \in \mathcal{F}_n$ に対して $P(A) = r^{-n}\sharp A$.

どの回でもいいから n 回分の硬貨の裏表を指定すると，結果がそうなる確率は 2^{-n} になる，という意味である．例えば $n = 1$ の例で言うと，

$$P[\{\omega \mid \omega_n = H\}] = P[\{\omega \mid \omega_k = T\}] = 1/2$$

[12] ファミコンや RPG は少なくとも 1990 年代前半までは流行していました．もう死語になっていたら，ごめんなさい．

が全ての k について成り立つ．これは,（いかさまでない，まともな）硬貨を（無限回）投げたとき，第 n 回目に表が出れば $w_n = T$, 裏が出れば $w_n = H$ とおいたことに相当する．このように定義すれば確率空間が定義できることは証明を要する（$\bigcup_n \mathcal{F}_n$ の上だけで P が定義されているが，これを $\mathcal{F}$ に拡張できるかどうかの確認が必要である.）

原理的にはこれで $(\Omega, \mathcal{F}, P)$ 上の任意の確率変数の期待値が計算できることになる．しかし，P が密度を持たないので，素朴な高校時代の計算は注意を要することがある．確率過程の具体的計算は，通常，有限個の確率変数ほど素朴には行かない．無限個の確率変数の問題は確率連鎖では極限の問題として現れる．

任意の自然数 n に対して

$$X_n(\omega) = \begin{cases} 1, & \omega_n = H, \\ -1, & \omega_n = T, \end{cases}$$

また $S_n = X_1 + X_2 + \cdots + X_n$ とおく．$\{X_n\}$ も chain だが，$\{S_n\}$ も chain である．ここではこの S_n たちに注目する．$\{X_n\}$ そのものは独立確率変数列なので，例えば積の期待値を計算するには，それぞれ別々に計算すればよい．これに対して $\{S_n\}$ は独立ではないので積の期待値は別々に計算することができない．しかし，独立確率変数の和であることを生かせば良い計算方法があるはずだと期待するのは自然なことである．実際，マルチンゲールはそのような手段の一つである．ここでは path という見方に注目しよう．

注：元になる確率空間 $(\Omega, \mathcal{F}, P)$ を明示しなくても，S_t の間の関係だけで確率過程 S_t は定まるが，分かりやすく，確率空間を指定するところから始めた．適切な確率空間の指定の仕方は一通りではない．簡単な一例をあげた．

Sample を一つ，例えば $\omega = (+ - + + \cdots)$ と固定しよう（簡単のため符号だけ書いた）．各 n 毎に $S_n(\omega)$ が定まる．これを $n \in \mathbf{Z}_+$ の関数と見ることができる．n を横軸に，$S_n(\omega)$ を縦軸にとればグラフに書ける．関数 $S.(\omega)$ は各時刻毎にゲームのコマがどこにいるかをたどっていく．つまり，コマの軌跡（道のり）を表す．（‘$\cdot$’ は，そこが関数の変数を代入する場所であることを表す便法）．$S.(\omega)$ を ω に対する**道 (sample path, sample function)** と呼ぶ．即ち，確率過程には二つの見方があることになる：

(1) $n \in \mathbf{Z}_+$ を固定したとき，$S_n(\cdot)$： $\Omega \to \mathbf{Z}$ （確率変数），
(2) $\omega \in \Omega$ を固定したとき，$S.(\omega)$： $\mathbf{Z}_+ \to \mathbf{Z}$ （時間の関数，即ち道）．

第 1 の見方が確率過程の定義であり，世の中は各時刻 n 毎にランダムで制御しきれない確率変数である，という「人の立場」と言えるかもしれない．すると，第 2 の見方は賽の目 ω を全て知っている「神の立場」ということになろうか．第 2 の見方で眺めると，Ω から関数の集まり（関数空間）への関数 $S.$ があることになる．Ω には確率が定義されているので，関数の集合が事象であるような確率空間を考えることになる，とも言える．

例では，異なる ω には異なる path $S.(\omega)$ が対応し，無数の path が実現しうる．これが「確率過程らしい」確率過程である．この例は **1 次元酔歩 (one-dimensional simple random walk)** と呼ばれる．一本道を一歩毎にでたらめに行ったり来たりする，という描像である．例は単に酔歩（酔っぱらいの行動）やすごろくを表すだけではない．ω_n は各時刻 n 毎にランダムに値をとり，しかも，時刻間で独立なので（P の定義と独立の定義から $\omega_1, \omega_2, \cdots$, が独立であることが示せる），本来 0 になるべき実験データ（あるいはゼロ信号）に加わったランダムな誤差（ノイズ）とみることができる.（比較的）簡単だが大変重要な，典型的な例である．

ω_n を実験データとみなすと，平均 $n^{-1} S_n$ に興味がある．実はこの例で $n^{-1} S_n(\omega)$ を計算すると，$n \to \infty$ の極限では確率 1 で 0 になる：$\lim_{n\to\infty} S_n = 0$, a.s.. この結果は **大数の法則** と呼ばれる定理の具体例になっている．

例では，時間がたつほどいろいろな道が可能になり，$n \to \infty$ を考えることが難しく見える．実際やさしくはない．しかし，$\{S_n\}$ は独立確率変数の和であることを生かすと，いろいろなことが分かる．一般に n が大きくなっていくときの確率過程の傾向を与える結果は極限定理と呼ばれる．大数の法則も極限定理の一つ．例に対して成り立つ，大数の法則より一段精密な極限定理に **中心極限定理** がある；$n^{-1/2} S_n$ の分布は $n \to \infty$ のとき正規分布 $N(0, 1)$ に収束する．大偏差値原理を用いればもっと細かいことが分かるはずである．

時刻 n の状態 S_n は S_{n-1} だけで分布が定まり，それ以前の状態（過去の履歴）には全くよらない（独立になる）．このような性質を持つ確率過程を **マルコフ過程** と呼ぶ.

マルコフ性：

$$\mathrm{E}^{\mathrm{a}}[\,\mathbf{1}_{\{X_n\le a\}}\mid X_1,X_2,\cdots,X_m\,]=\mathrm{E}^{\mathrm{X_m}}[\,\mathbf{1}_{\{X_{n-m}\le a\}}\,]\,,\ m<n\,.$$

（記号が悪い．右辺の $\mathrm{E}[\ \cdot\ ]$ の肩の X_m は右辺 $\mathrm{E}[\ \cdot\ ]$ の中で X_m と書くのとは意味が違う.）

非マルコフ過程はマルコフ過程に比べて，一般に極めて計算が困難である（過去の履歴を全て調べないと次の状態が決まらないので計算が大変）．このため，今日まで多くの原理的研究及び実用的応用がマルコフ過程であった.（研究者としては残念な状況である.）マルコフ過程でない典型例として self-avioding walk がある．そういうことに関連した研究もしている.

問１.

講義で取り上げた確率過程の例（硬貨を投げて表裏に応じて左右に一こまずつ進むゲーム）の例について，時刻 $n\in\mathbf{Z}_+$ における期待値 $\mathrm{E}[\,X_n\,]$ を計算せよ．特に $n=0,1,2,\cdots$ の値によって期待値が変わるか否かを指摘せよ.（期待値の線型性から n について帰納法で計算できる.）

問２.

１次元酔歩の例について，計算機で ±1 を確率 1/2 ずつで発生する乱数列を作り，定義を用いて見本 (sample path) $S_1,S_2,\cdots$ を発生させ，図示して見よ．横軸に $n\in\mathbf{Z}_+$ 縦軸に $S_n\in\mathbf{Z}$ をとれ．n が大きくなるとどのような傾向があるか．乱数の初期値を変えて，見本を何個か図示して見よ.（計算機による計算についてはプログラムもつけること．Mathematica などの高級ソフトの利用も歓迎するが，どの場合でも，プログラムまたはソースをつけること.）

5.1.1 Weierstrass 近似定理　（省略可能？）.

類題に Cehbyshev 不等式を用いて Weierstrass 近似定理を証明できる.

$f:[0,1]\to\mathbf{R}$ が連続関数とする．$\epsilon>0$ に対して $\displaystyle\sup_{0\le x\le1}|B(x)-f(x)|\le\epsilon$ を満たす多項式 B が存在する.

不公平な硬貨を永久に投げ続ける試行を考える．$0\le p\le1$ を固定する．$\{Y_n\}$, $n\in\mathbf{N}$, を i.i.d. で, $P(Y_n=1)=p$, $P(Y_n=0)=1-p$ を満たすとする．また $T_n=Y_1+Y_2+\cdots+Y_n$ とおく．期待値の線型性より $\mathrm{E}[\,S_n/n\,]=\mathrm{E}[\,Y_n\,]=p$. また $\mathrm{V}[\,Y_n\,]=p(1-p)\le1/4$ と Chebyshev の不等式より $P(|n^{-1}S_n-p|>\delta)\le(4n\delta^2)^{-1}$. ２項定理より $P(S_n=k)=\begin{pmatrix}\mathrm{n}\\\mathrm{k}\end{pmatrix}p^k(1-p)^{n-k}$, $0\le k\le n$ だから

$$B(p):=\mathrm{E}[\,f(n^{-1}S_n\,]=\sum_{k=0}^{n}f(n^{-1}k)\begin{pmatrix}\mathrm{n}\\\mathrm{k}\end{pmatrix}p^k(1-p)^{n-k}\,.$$

f は $[0,1]$ で一様連続なので $\epsilon>0$ に対して $\delta>0$ がとれて $|x-y|<\delta$ ならば $|f(x)-f(y)|<\epsilon/2$, また，有界なので $K\ge0$ がとれて $0\le y\le1$ に対して $f(y)\le K$. $Y=|f(n^{-1}S_n)-f(p)|$, $Z=|n^{-1}S_n-p|$ とおくと

$$|B(p)-f(p)|\le\mathrm{E}[\,Y\,]=\mathrm{E}[\,Y;\ Z\le\delta\,]+\mathrm{E}[\,Y;\ Z>\delta\,]\le\frac{1}{2}\epsilon+\frac{2K}{4n\delta^2}\,.$$

n を十分大きくとれば右辺を $0\le p\le1$ によらず ϵ 未満にできる.

Doob 分解はブラウン運動の２乗平均変位が時間に等しいこと $\mathrm{E}[\,{W_t}^2\,]=t$ のある意味で拡張概念ということができる.

5.2 ブラウン運動．－確率過程－

前節で，時刻変数 n が離散（整数）的な場合について，簡単な例で説明した．時刻変数が離散的な場合でも，極限定理を中心に，問題はたくさんある．応用上も，多くの社会的な活動は，RPG やすごろくのように，ランダムな選択が離散的に起こる．計算機や digital 通信は，下部構造そのものが離散時間的である．

他方，自然現象は時間が連続であり，いつでも「ノイズ」が生じうる．講義の冒頭で挙げた窓口の処理待ち問題は，連続時間確率過程の問題である．基礎事項の詳細には立ち入れないが，連続時間の場合，有限の時間間隔でも無限個の確率変数があり，sample path $X_t(\omega)$ の t に関する連続性のような，短い時間での変化を問う問題が生じる．

時間変数も状態空間も連続な確率過程の代表例，**Wiener（ウィーナー）過程**，を取り上げる．Wiener 過程は，確率過程論の出発点になった代表例中の代表例であり，多くの興味深い，特徴的な性質を持つ．応用上の重要性も格別である．自然現象は時間的にも空間的にも連続なものが多いから，当然かも知れない．

1827 年，花粉の中の微小な粉が水中で不規則な運動を行うことが顕微鏡観察で発見された．Wiener 過程の研究の出発点となったこの運動は，発見した植物学者にちなみ，Brown（ブラウン）運動と呼ばれる．Einstein（アインシュタイン）等の研究により，不規則な運動の原因は水分子の熱運動による撹乱であることが確立した．１９世紀末に「観測できないものは存在しない」という哲学から，見ることのできない分子や原子を否定した物理学者・科学哲学者 Mach（マッハ）が，Brown 運動を顕微鏡で見せられて分子の存在を認めた，という科学史上のエピソードもある．Wiener 過程は，Brown 運動の数学モデルであると同時に，通信回線等の（アナログ）電気回路の熱雑音のモデルでもある．名前は，研究の端緒を与えた電気工学者 Wiener に由来する．

時間が連続な確率過程 $X_t,\ t \geq 0,$ が Wiener 過程であるとは，次の条件を満たすことである．

(1) 状態空間（X_t の値域 Σ）が実数 $\mathbf{R}$ で，sample path （時間 t の関数としての $X_t(\omega)$）は連続関数である．
(2) 加法性（$t_1 < t_2 < \cdots < t_n$ とするとき，$X_{t_n} - X_{t_{n-1}}$ が，$X_{t_{n-1}} - X_{t_{n-2}},\ \cdots,\ X_{t_1} - X_{t_0},$ と独立なこと）．
(3) 時間的一様性（$X_t - X_s$ の分布が $t - s$ だけで決まること）．
(4) $\mathrm{E}[\, X_1 - X_0\,] = 0$，かつ，$\mathrm{V}[\, X_1 - X_0\,] = 1.$
(5) $X_0 = 0.$

以上の性質を持つ確率過程が存在することは証明されている．Wiener 過程のように種々の強力な公式が成り立つ確率過程が存在すること自体が，深い内容である.（計算しやすいように安易にいろんな性質を仮定すると，そんな話のうまい確率過程は存在しない，ということになりかねない.）

条件 (5) は原点を決めるだけで，重大な内容はない．すぐ述べるように，条件 (4) も本質的ではないので，Wiener 過程の定義のうちで本質的な条件は (1)(2)(3) （連続性，加法性，時間的一様性）である．条件 (4) が本質的ではない根拠を述べる．条件 (1)(2)(3)(5) を満たす確率過程 X_t が与えられたとし，$m = \mathrm{E}[\, X_1\,]$, 及び $v = \mathrm{V}[\, X_1\,]$, とおく．このとき，$Y_t = X_{t/v} - m\,t/v$ とおくと，Y_t は条件を全て満たし，Wiener 過程になる（(4) 以外を満たすことは明らか）．逆に言えば，X_t が条件 (4) 意外を満たせば，Wiener 過程 Y_t を用いて $X_t = Y_{v\,t} + m\,t$ と書ける．これは次の定理による（証明はしない）．

定理．

確率過程 X_t が条件 (1)(2)(3)(5) を満たせば，ある実数 m 及び $v \geq 0$ がとれて，任意の $t > s$ に対して $X_t - X_s$ は，平均 $\mathrm{E}[\, X_t - X_s\,] = m\,(t-s)$，分散 $\mathrm{V}[\, X_t - X_s\,] = v\,(t-s)$ の正規分布に従う．特に X_t が Wiener 過程ならば（(4) も満たせば），$m = 0,\ v = 1$ である $(t = 1,\ s = 0$ と (4) から$)$．

平均 m，分散 v の正規分布とは，実数上の確率であって，密度 $\rho_{m,v}(x)$ が

$$\rho_{m,v}(x) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x-m)^2}{2\,v} \right)$$

で与えられるものを言う．即ち，定理の言うことは

$$P(X_t - X_s \le x) = \int_{-\infty}^{x} \rho_{m(t-s),v(t-s)}(y)\,dy$$

ということ．

Wiener 過程の種々の確率や期待値などは $\rho_{m,v}(x)$ を用いて計算できる．一見抽象的に見えるかも知れない定義から $\rho_{m,v}(x)$ の具体的な公式が出る．例えば，$t > s$，かつ，$A \subset \mathbf{R}$（正しくは $A \in \mathcal{B}$）のとき，$X_s = y$ という条件下で $X_t \in A$ となる確率は，

$$P[X_t \in A \mid X_s = y] = \int_{x\in A} \rho_{0,t-s}(x-y)\,dx$$

となる．ここで，事象 $A, B \in \mathcal{F}$ に対して，記号 $P[A \mid B]$ は条件付き確率 $P[A \mid B] = P[A \cap B] \,/\, P[B]$ を表す．B が起こったという前提の下で A の起こる確率，である．(導出：$X_s = X_s - X_0$ なので加法性から，左辺は $P[X_t - X_s \in A - y \mid X_s - X_0 = y] = P[X_t - X_s \in A - y]$．ここで，$A - y = \{z \mid z + y \in A\}$．これに定理を用いればよい.)

特に $s = 0, y = 0$ とおけば，定義より $X_0 = 0$ なので，$P[X_0 = 0] = 1$．これから，$P[X_t \in A] = P[X_t \in A \mid X_0 = 0]$．よって，

$$P[X_t \in A] = \int_{x\in A} \rho_{0,t}(x)\,dx\,, \quad \rho_{0,t}(x) = \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{x^2}{2\,t}\right). \tag{1}$$

この式は，時刻 0 に原点 0 から出発 ($X_0 = 0$) した Wiener 過程の t 時間後の位置 X_t が平均 0 分散 t の正規分布に従うことを意味する．少し平たく言えば，時刻 t に点 x にいる相対的な確率（確率密度）が，正規分布で表される．

水の中に一滴のインキを落とすと，次第に薄まりながら広がっていく．金属棒や板の一箇所を一瞬暖めると，次第に熱は広がって全体が一様な温度になる．これらの現象は，分子の熱運動という不規則な運動の，巨視的な効果である．微視的には（インキ粒子一つは）Wiener 過程で記述される不規則な運動だが，多数の粒子が集まってインキ一滴となると，微視的な運動の生じる確率に応じた割合で，密度分布が実現する（これは粒子間の運動が独立であるという仮定の下で正しい）．例えば，長細い容器に入った水にインキを落とし，落とした点を原点 0，時刻を 0 とする．t 時間後点 x でのインキの密度（薄さ） $u(t, x)$ は (1) の確率密度に比例する；$u(x, t) \propto \rho_{0,t}(x)$．時間がたつにつれて山が低くなる正規分布となる．微視的 (sample path) には不規則な運動でも，巨視的（密度）にはなめらかな関数になる．確率過程のおもしろさも難しさもここにある．

5.2.1 ランダムウォークの連続極限：長いすごろくを遠くから眺める．

Wiener 過程は存在する，と書いたが,「作り方」もいろいろある．理論的にはもっとも単純とは言えないが，計算機上で作りやすく，またある観点からは極めて興味深い,「作り方」を説明する．

前の節で 1 次元酔歩 (simple random walk) という離散時間確率過程を定義した．定義を要約すると，整数時刻の整数値確率過程 Z_n で，sample path が

$$Z_n(\omega) = \sum_{m=1}^{n} \omega_m\,, \quad n = 0, 1, 2, \cdots,$$

と書けるものであった（X ばかり使うと紛らわしいので，simple random walk は Z で表した）．ここで m 回目の硬貨投げで，表ならば $\omega_m = 1$，裏ならば $\omega_m = -1$，とおく（確率 1/2 で ± 1 の値をとり，$\{\omega_m\}$ は独立，ということ）．

定理 1． 離散時空確率過程 $c^{-1/2}\, Z_{[c\,t]}$ は $c \to \infty$ のとき *Wiener* 過程 X_t に法則収束する．(Z の添字の $[c\,t]$ は，実数 $c\,t$ の整数部分をとる *Gauss* の記号.)

ここで，確率変数列 Y_c が $c\to\infty$ のとき確率変数 W に **法則収束する** とは，直感的には Y_c の従う分布 Q_c が W の従う分布 R に近づくという意味である．

（詳しく言うと，分布 Q_c が R に **弱収束** する，という意味である．これは，$R[\partial A]=0$ を満たす任意の集合（正しくは，$A\in\mathcal{F}$）に対して，$\lim_{c\to\infty} Q_c[A]=R[A]$ が成り立つ，という意味である．$R[\partial A]=0$ とは A の境界集合の，R で測ったときの確率が 0 という意味である．この一見ややこしい留保条件は,「分布が近づく」という言葉の直感的意味に近い形に弱収束を定義するための条件だが，詳しい説明は，定理 1 の証明とともに，割愛する．

定理 1 では，X_t を確率変数の集まり（確率過程）としてではなく，$\Omega\ni\omega$ から sample path （時間の関数） $X.(\omega)$ への関数（一つの関数値確率変数）と見て法則収束を論じている．次の 定理 2 でもその見方をしたときの確率（分布）を論じている.）

定理 1 は，simple random walk という，計算機でもシミュレーションしやすく，数学的にも直感的にも比較的簡単な確率過程の極限として，Wiener 過程を「作る」ことができることを表している．時空ともに離散的な確率過程が時空ともに連続的な Wiener 過程に近づく，というのも興味深いが，上記の定理は，次の定理の意味で，Wiener 過程の「自己相似性」を表すので，非常に興味深い．

定理 2. *Wiener* 過程 X_t の分布は，任意の正の実数 a に対して *Wiener* 過程を時空についてスケール変換（拡大・縮小）した確率過程 $a^{-1/2}\,X_{[a\,t]}$ の分布に等しい．

証明. 定理 1 において，c を $c\,a$ と置き換えて $c\to\infty$ としても（極限をとる変数の変数変換だから）構わない．即ち，$(c\,a)^{-1/2}\,Z_{[c\,a\,t]}$ は $c\to\infty$ のとき Wiener 過程 X_t に法則収束する．$(c\,a)^{-1/2}\,Z_{[c\,a\,t]}=a^{-1/2}\,c^{-1/2}\,Z_{[c\,(a\,t)]}$ だから，定理 1 において $t\to a\,t,\ Z\to a^{-1/2}\,Z$ と置き換えたものになっている．よって，$c\to\infty$ で $a^{-1/2}\,X_{a\,t}$ にも法則収束する．故に，$a^{-1/2}\,X_{a\,t}$ と X_t の分布は等しい．

定理 1 は X_t の時間と空間をうまく呼応させてスケール変換（拡大・縮小）した後，極限をとることになっている．このような極限の取り方を **scaling limit** と呼ぶ．他方，定理 2 は Wiener 過程が，時間と空間をうまく呼応させてスケール変換すると，再び Wiener 過程に戻る（分布が不変である）こと，を表す．これを分布の **スケール不変性 (scale invariance)** という．物理用語では**統計的自己相似性 (statistical self-similarity)** ともいう.「統計的」という修飾語は，一つ一つの見本関数 sample path の関数としての自己相似性ではなく，分布（確率）の自己相似性に言及していることからつく．

大雑把に言うと，Brown 運動を顕微鏡で 3 倍拡大してビデオに録画し，$1/9=1/3^2$ 倍のスロー再生で再生画像を見ると,「元の Brown 運動らしく」見える，ということである．拡大と早回し（時空のスケール変換）の関係は古典力学でもおなじみである．怪獣映画やアクション映画で，ビルが爆破されて破片が飛び散る（放物運動）シーンを，ミニチュア撮影ですませるには，スローモーションで再生する必要がある．違いは，放物運動の場合は拡大が 3 倍ならばスピードは $1/\sqrt{3}$ 倍のスロー再生，ということである．2 乗か 1/2 乗かの違いがある．

「統計的」という修飾語にもう一つの特徴がある．放物運動は時空の適切なスケール変換で正確に元の運動に一致する（自己相似である）．Wiener 過程の sample path は自己相似ではない．不規則な運動が拡大・スロー再生でぴたっと一致したら，かえっておかしい．あくまで「典型的な振る舞いが元の path と同じ」に過ぎない．しかし，そもそも，Wiener 過程の典型的な sample path の振る舞いとはなにか？これは直感的にすら容易ではない．諸君が各自計算機で sample path を発生させてそれをいくつか図示したときに，共通するものを直感したならば，それが Wiener 過程に対する諸君の直感ということになろうか．

Wiener 過程は自然現象（例えば，分子の大きさの熱運動に由来する現象を扱う統計力学の分野）を表すには単純過ぎて，十分当てはまらない現象が多いことが分かっている．しかし，Wiener 過程とは一致しなくても，統計的自己相似性には関連すると見られる現象が普遍的に知られている．この意味で，定理 2 は Wiener 過程を越えて興味ある結果であり，この性質に直接結びついた Wiener 過程の「作り方」 定理 1 は，数学的にも重要になってきている．

問１．

正規分布の密度 $\rho_{m,v}(x)$ について，次のような公式が成り立つことが知られている．この式が成り立つように，右辺の M と V を左辺の $m_1,\, m_2,\, v_1,\, v_2,$ を用いて表せ．

$$\int_{-\infty}^{\infty} \rho_{m_1,v_1}(x-z)\,\rho_{m_2,v_2}(z-y)\,dz = \rho_{M,V}(x-y)\,.$$

特に，$m_1 = m_2 = 0$ かつ $v_1 = t-u,\, v_2 = u-s$ のときの結果を，Wiener 過程の言葉で解釈せよ．

公式 $\displaystyle\int_{-\infty}^{\infty} \exp(-a\,(z-b)^2)\,dz = \sqrt{\frac{\pi}{a}}$ を用いてもよい．

問２．

定理 1 を用いれば simple random walk Z_n を用いて Wiener 過程の性質を数値計算で調べることができる．適当に大きな定数 c を用いて $c^{-1/2}\,Z_{[c\,t]}$ で近似するのである．計算機で sample path を多数発生（乱数の初期値をいろいろ変えて計算）させて，固定した t （例えば $t=1$）における $c^{-1/2}\,Z_{[c\,t]}$ の値の統計（ヒストグラム）をとることにより，Wiener 過程の遷移確率密度 $\rho_{0,1}(x)$ の概形を求めてみよ．理論的結果（(1)）と比べてどうか？c は計算機のスピードと相談して様子を見ながら決めればよい．

解答にはプログラムもつけること．Mathematica などの高級ソフトの利用も歓迎するが，どの場合でも，プログラムまたはソースをつけること．

参考：プログラム例の主要部．

```
      implicit real*8(a-h,o-z)
（中略）
      parameter (ir30=1073741824, irm=48828125, r31=2147483648d0)
      data irr/1000001/
      dx=dsqrt(dt)
（中略）
1     continue
      it=it+1
      irr=irr*irm
      if (irr .lt. 0) irr=(irr+ir30)+ir30
      r=dfloat(irr)/r31
      if (r .le. 5d-1) then
      ix=ix+1
      else
      ix=ix-1
      endif
      t=dfloat(it)*dt
      x=dfloat(ix)*dx
（中略）
      if (it .eq. ilast)     goto 9999
      goto 1
（後略）
```

問3.

Wiener 過程の「作り方」には 定理 1 以外にもある．例えば，$\{Y_k\}$, $k = 0, 1, 2, \cdots$, を独立同分布な確率変数列で，各々正規分布 $N(0,1)$ （平均 0 分散 1）に従うとする．このとき

$$X_{N,t} = \frac{t}{\sqrt{\pi}}\, Y_0 + \sqrt{\frac{2}{\pi}} \sum_{n=1}^{N} \frac{\sin(n\,t)}{n}\, Y_n\,, \quad t \in [0, \pi]\,,$$

とおくと，見本関数 $X_{N,t}(\omega)$ は（確率 1 で） t について一様に収束し，極限 $X_t = \lim_{N\to\infty} X_{N,t}$ は Wiener 過程になる．

正規分布に従う乱数列を計算機で発生させて，$X_{N,t}$ の sample path （$0 \le t \le \pi$）を生成してみよ．N を増やしていくときの変化を，定理 1 に基づいて生成した図と比較して見よ．

解答はプログラムもつけること．Mathematica などの高級ソフトの利用も歓迎するが，どの場合でも，プログラムまたはソースをつけること．

5.3 Poisson 過程.

5.3.1 バスはなぜ来ないか.

たいていのバスは時刻通り来ない．いくつかの大都市では，バス停に電気仕掛けで次のバスがどれくらい近くにいるかを表示している．時刻表通り来るならばこのような表示はいらないはずである．それほどバスは時刻通り来ない.（頑張っているところもある．アメリカワシントン州シアトルのバスはちゃんと来た.）遅れや進みが極端になってバスが 2 台以上つながって来ることすら多い.（先に来たバスに乗客が多く乗るので，一度バスの間隔が等間隔でなくなるとその傾向はどんどんひどくなる，という面もあるが，この問題には立ち入らない.）

場所によっては停留所に，時刻でなく，例えば「10 分に 1 本」という書き方をしている．ちょっと頭をひねって「10 分に 1 本ならば，平均すると 5 分待てば次のバスが来る」と考えて待つと，ちっとも来ない.（この計算の根拠はすぐ次に示す.）遅れたり早くなったりするのは，交通事情もあってやむを得ないとも思う．しかし，単に時刻通り来ないだけでなく，時刻表よりもバスの来る量が少ない気がすることも多いのはどういうわけだろう．時刻表より実際の運行回数が少ない「間引き」が行われていると即断していいのだろうか？

100 分の間に合計 10 本のバスを走らせる路線があったとする．平均的には 10 分間隔で運行していることになる．平均運行間隔 10 分である．乗客はこの停留所に勝手な時刻にやって来てバスを待つとする．このとき，この停留所で待つ乗客は平均何分待てば乗れるだろうか？

まず，バスが等間隔に来る場合を考える．10 分ごとに 1 本バスが来る．乗客が停留所に到着してから次のバスが来るまでに待つ時間 Z は 0 分（次のバスの来る直前＝幸運な場合）から 10 分（前のバスが行った直後＝悪夢の場合）までの可能性がある．Z のばらつきは乗客には予測・制御のできない，確率変数であると考える．Z の値は $0 < Z < 10$ の範囲に分布することになる．この分布（乗客のバス停到着時刻の分布）が一様（等確率）とすると，平均待ち時間 $\mathrm{E}[\,Z\,]$ （確率変数 Z の期待値を $\mathrm{E}[\,Z\,]$ と書く）は

$$\mathrm{E}[\,Z\,] = \int_0^{10} t\,dt/10 = 5$$

即ち，

平均運行間隔 10 分のバス停において，乗客の到着時刻の分布が一様分布のとき，この乗客の平均待ち時間 $\mathrm{E}[\,Z\,]$ は 5 分である.

次にバスが等間隔に来ない場合を考える．先ほどと同様に，平均運行間隔 10 分の路線を考える．交通事情や乗客の具合で，ある停留所に来たときには等間隔でないとする．それでも，平均運行間隔 10 分，つまり 100 分の時間幅の間に最初から最後まで 10 台のバスが通るとする．このときこの停留所で待つ乗客は平均 5 分待てば乗れるだろうか？

答えは，一般には平均 5 分より長く待たないといけない．これを理解するために，極端な場合を考える．10 本のバスが全部一斉に 100 分目に到着したとしよう．平均運行間隔は $100/10 = 10$ 分である．乗客は t 分目に到着したとすると，待ち時間は $100 - t$ である．t は 0 から 100 の間のどの時刻も同じ確率でとるとすると，平均は

$$\int_0^{100} (100 - t)\, dt/100 = 50$$

つまり，

バスが 100 分ごとに 10 台数珠繋ぎになって来るバス停においては，平均運行間隔は 10 分だが，到着時刻の分布が一様分布の乗客の平均待ち時間は $\mathrm{E}[\,Z\,] = 50$ 分である．

バスは平均 10 分に 1 本走っているのに，乗客は平均 50 分も待たないといけない！

なぜ，平均運行間隔が等しくても平均待ち時間が 5 分から 50 分まで違うのか？バスの運転間隔が狭いと一つ前のバスを逃してもすぐ次が来る．しかし，間隔が狭いということは偶然その時間帯に乗客が停留所に到着する可能性も低いということである．より多くの場合，乗客は間隔のあいた時間帯に到着するから，平均より長く待つ可能性の方が高いことになる．

平均運行間隔が同じ二つのバス路線を比べた場合，（大雑把に言うと）等間隔に近い走り方をしているほど平均待ち時間は短く，等間隔からずれているほど平均待ち時間は長い．

さていよいよ，バスが全くでたらめに運行している場合を考える．でたらめ，というのは，運行間隔（あるバスが通ったあと次のバスが来るまで）T が（道路事情などで）ばらつくので（正の値を取る）確率変数であるということである．どんなふうにでたらめか，ということを定義する必要があるが，数学的に最も「単純」とされるのは独立性を用いて次のように定義される．（書くと「単純」に見えないかも知れないが，いろいろな計算が可能なことが知られているので実用上も非常に利用される．）

「運行間隔を表す確率変数 T の分布の密度が指数分布

$$\lambda \exp(-\lambda\, t), \quad t > 0$$

で与えられる．λ（ラムダ）は平均運行間隔の逆数．しかも，T は注目したバスより前のバス達がどんな間隔で通ったかということと独立である．」（このとき，時刻 t におけるバスの累積到着台数（時刻 t までに何台通ったか）$X(t)$ が平均運行間隔 $1/\lambda$ の Poisson（ポワッソン）確率過程になっているという．）

T の分布の密度が上述の指数分布で与えられるとは，t_0 と t_1 を $0 < t_0 < t_1$ を満たす定数とするとき $t_0 < T < t_1$ となる確率 $\mathrm{Prob}[\,t_0 < T < t_1\,]$ が

$$\mathrm{Prob}[\,t_0 < T < t_1\,] = \int_{t_0}^{t_1} \lambda \exp(-\lambda\, t)\, dt\,,$$

で計算されるということである．

今の例では平均運行間隔が 10 分であったから $\lambda = 1/10$．このとき，乗客の平均待ち時間は $1/\lambda = 10$ となる（計算は後述）．即ち平均運行間隔に等しい．

平均運行間隔 10 分で，累積到着台数が Poisson 確率過程になっているバスに対して，乗客の平均待ち時間は $\mathrm{E}[\,Z\,] = 10$ 分である．

このことを逆に見ると，

もし，乗客の平均待ち時間と平均運行間隔が一致しないならば，バスの到着は Poisson 確率過程になっていない．それは，バスの運行間隔が「全くのでたらめではない」ことを意味する．

先ほどの例と合わせて考えると，平均待ち時間が短い場合は，バスが「でたらめに来る」場合（到着が Poisson 確率過程）に比べて等間隔に近い（つまりバスの定時運行努力が実っている）ことになり，平均待ち時間が長い場合は，バスが数珠繋ぎになったり間隔が空いたりが極端になっていると予想される（利用が多く，乗降ののべ時間が長い場合にはバスが団子状態になりやすい，と考えられる）．

乗客の平均待ち時間 $= 1/\lambda$ になることの証明．

T の分布の密度の式 ($\mathrm{Prob}[\,t_0 < T < t_1\,]$) から，乗客の平均待ち時間が $1/\lambda$ になることを導いておく．

ある特定のバスに乗れるケースに注目する．それは直前のバスが通り過ぎてから問題のバスが到着する直前までに乗客が停留所に来た場合である．仮に乗客は直前のバスが通り過ぎてから時間 s たって停留所に着いたとしよう．ということは，問題のバスと直前のバスは間隔 T が s 以上だということである．この条件の下で待ち時間 $T-s$ の期待値 $\mathrm{E}[\,T-s\mid T>s\,]$ を計算すればよい．($\mid T>s$ は $T>s$ という条件付きで期待値を取ることを表す.）T の分布の密度の式から

$$\mathrm{E}[\,T-s\mid T>s\,]=\frac{\int_s^{\infty}(t-s)\lambda\exp(-\lambda\,t)\,dt}{\int_s^{\infty}\lambda\exp(-\lambda\,t)\,dt}\,.$$

（条件 $T>s$ がついているので積分範囲が $t>s$ となる.）これを計算するには分母分子とも $t=u+s$ という積分変数変換をするのがよい．すると定数 $\lambda\exp(-s\,\lambda)$ が分母分子で打ち消して

$$\mathrm{E}[\,T-s\mid T>s\,]=\frac{\int_0^{\infty}u\exp(-\lambda\,u)\,du}{\int_0^{\infty}\exp(-\lambda\,u)\,du}$$

と変形され，これを計算すると

$$\mathrm{E}[\,T-s\mid T>s\,]=1/\lambda$$

を得る．この値はどのバスに乗れたかにも s がいくらかにもよらないので，結局この式を導いた条件に関係なく常に平均待ち時間は $1/\lambda$ になる．

大文字小文字の使い分けは，確率変数は T のように大文字で書くことが多いという習慣による．分布の密度を用いて期待値の積分を実行するときは確率変数 T が t という値を取る，即ち，$T=t$ での密度という意味で，小文字の t になっている．

今回の話では平均運行間隔という平均値が一定でも平均待ち時間が変わりうることを示した．直感的には「１時間に４本なら１５分くらい待てば１本来るだろう」と考えたいところだが（そしてバスが「でたらめ」に来ればその通りなのだが)，一般には平均の取り方によって答えが違う．もちろん，これは取り上げた「バスの到着」という例が見かけに反して（確率過程という）高度な問題だったからであるが，確率論の教えるところには安易な直感の誤りを指摘するものが他にもある．誤った直感を排して，正しい推論に導くために現代確率論は，一見直感的でない用語や定義を用いる．それらは取っつきにくく，この講義でもどれだけ触れられるか分からないが，確率論の問題は根本に立ち返って考えないといけない，という注意は学んでほしい．

問１．

平均運行間隔が定数 a で与えられるバスのバス停において，乗客の到着時刻の分布が一様分布のとき，この乗客の平均待ち時間を次の各場合に求めよ：

(1) バスが a ごとに等間隔に到着する場合.
(2) バスが到着間隔 $3a/2$ と $a/2$ を交互に繰り返して到着する場合（つまり，到着時刻が， $3a/2, 2a, 7a/2, 4a, \cdots,$ のとき.）
(3) $0<r<1$ を定数とするとき到着間隔 $r\,a$ と $(1-r)\,a$ を交互に繰り返して到着する場合
(4) バスの累積到着台数が平均運行間隔 a の Poisson 確率過程になっている場合.

問２．

あなたはバス会社（その他どんな顧客サービス会社でも当てはまりうるが）の顧客相談窓口に配属されているとする．乗客から次のような苦情が来た.「停留所の時刻表では１０分に１本となっているのに２０分待たされることもざらだ．平均を取ったら１５分待たされている．間引き運転しているのではないか？」あなたの会社が間引き運転をしていない良心的な会社の場合，この苦情に正しく説明するにはどう返事をすればよいかを考えよ.（回答の鍵は講義（資料）に全て書かれているが，客は確率論の専門家でもないし，この講義も聞いていない，という前提で何とか分かるように説明せよ.）

5.3.2 Poisson 過程．

今回は，**時間変数が連続で，状態空間（値域，X_t の取りうる値）が離散的** な確率過程の代表例として，Poisson process （ポワッソン過程）を紹介する．

Poisson 過程は，１個単位で時間的にでたらめに発生して累積していく量の数学的モデルである．前節のバスの例では，特定の停留所で見張っていて，時刻 t までに到着（発車）したバスの累積台数を X_t とおいた．いつでもバスが到着しうる，という意味で時間変数は連続であり，台数が整数値であるという意味で状態空間は離散である．さらに，もし X_t の性質が Poisson 過程でよく表されることが分かれば,「でたらめに」到着していると判断できる．

Poisson 確率過程 $X(t)$ は，バスの到着だけでなく，時間的にでたらめに１個単位で発生する確率現象のモデルとして実用上も重要である．例えば，電話交換台が処理する電話の累積回数，レジや案内カウンターやチケット売場などのサービスカウンターへの客の累積到着数，また，通信回線や画像などの（ビットやピクセル単位の）ノイズの累積個数，等のモデルに利用される.（実用上相互に無関係に見える問題に適用できるということが Posson 確率過程の重要性を表している.）

時間が連続な確率過程 $X_t,\ t \geq 0,$ が Poisson 確率過程であるとは，次の条件を満たすことである．

(1) 状態空間（X_t の値域）が非負整数で，sample path （t の関数としてみた $X_t(\omega)$）の時間的変化は 1 ずつ増えるだけでそれ以外の時は時間的に一定（階段関数）である．
(2) **加法的**，即ち， $t_1 < t_2 < \cdots < t_n$ とするとき，増分 $X_{t_n} - X_{t_{n-1}}$ （時間 $(t_{n-1}, t_n]$ に何件到着したか）が，それ以前の時刻の増分 $X_{t_{n-1}} - X_{t_{n-2}}, \cdots, X_{t_1} - X_{t_0}$, と（確率変数として）独立である．
(3) **時間的一様**，即ち，確率変数 $X_t - X_s$ の分布は $t - s$ だけで決まり，s そのものにはよらない．
(4) $X_0 = 0$. （測定開始時 $t = 0$ は累積ゼロ件という意味で，本質的でなく，省いてもよい.）

条件 (2) 及び (3) は Wiener 過程と共通である．$X_t - X_s$ の分布は明示していないことに注意．定義上は分布の具体形によらず，上記４条件を満たせば（即ち，幅 1 で増大し，加法的かつ時間的一様な連続時間の確率過程を）Poisson 過程と呼ぶ．ところが，以上の事実を用いると，いろいろな量を式変形で計算することができる．例えば，バス運行の例で T の分布密度の式に出てくる λ が実際に平均運行間隔の逆数（即ち単位時間当たり通るバスの台数の平均）$\mathrm{E}[\, X(a+1) - X(a)\,]$ に等しいことなども示すことができる．また，平均待ち時間が $1/\lambda$ に等しいことの証明も，加法性と時間的一様性が本質的な役割を果たしている．

これらの具体的計算は次の事実が証明できることに基づく．

定理．

上記４条件を満たせば，$X_t - X_s$ の分布は，平均値が $t - s$ に比例する Poisson 分布になる．

Poisson 分布 とは，非負整数 $n = 0, 1, 2, \cdots$ 上の分布であって値 n をとる確率が

$$Q[\{n\}] = \exp(-\mu)\,\frac{1}{n!}\,\mu^n$$

で表されるものを言う．正定数 μ は Poisson 分布の平均値である；

$$\sum_{n=0}^{\infty} nQ[\{n\}] = \mu\,. \tag{2}$$

確率変数 Z が平均 μ の Poisson 分布に従うとは，$\mathrm{Prob}[\, Z = n\,] = Q[\{n\}]$ ということであった．定理の言うことは，X_t が Poisson 過程ならば，任意の $s < t$ に対して $X_t - X_s$ （時間 $(s, t]$ 内の到着数）が

$$P[X_t - X_s = n] = \exp(-\lambda\,(t-s))\,\frac{1}{n!}\,\{\lambda\,(t-s)\}^n\,, \quad n = 0, 1, 2, \cdots, \tag{3}$$

に従う，ということである．ここで正定数 λ は単位時間幅 1 内の到着数の期待値であることは Poisson 分布の説明から明らかであろう．Poisson 過程という名前はこの事実に由来する．一見計算の役に立たないように見える Poisson 過程の定義（４つの条件）から，具体的な公式 (3) が得られることに注意.

定理の証明

きちんと行う余裕はないが，一般原理（加法性，一様性）から (3) という具体形が出る「気持ち」を示す．

$P[X_t-X_s=n]$，即ち時間幅 $(s,t]$ に n 回 event が発生する（X が n だけ増える）確率，を計算するために，区間 $(s,t]$ を N 等分して，ひと区画当たり時間間隔 $(t-s)/N$ にする．N を n に比べて十分大きくとれば，短い時間間隔 $(t-s)/N$ に event 発生が 2 回以上起こる可能性は極めて小さくなる.（正確には，N が大きいときこの可能性が小さいことを，加法性と一様性を用いて言うのだが，省略する.）Event が N 個の小区画のうちの n 箇所で起こる．単位時間当たり平均 λ 回 event が発生するとすると，時間間隔 $(t-s)/N$ では $p=\lambda\,(t-s)/N$ の確率で 1 回 event が発生する．小区画あたり確率 p で起こることが N 個の小区画のうちの n 箇所で起こる確率は（高校で習ったように），${}_NC_n\,p^n\,(1-p)^{N-n}$ である（この式を導くところで加法性と一様性を使っている）．但し，これは 1 つの小区画の中で 2 回以上 event が発生するケースを除外して導いたので，この式は $N\to\infty$ で初めて $P[X_t-X_s=n]$ に等しくなる；

$$P[X_t-X_s=n]=\lim_{N\to\infty}{}_NC_n\left(\frac{\lambda\,(t-s)}{N}\right)^n\left(1-\frac{\lambda\,(t-s)}{N}\right)^{N-n}.$$

後は標準的な計算で証明できるが，参考までに右辺の計算方法の方針を示しておく．${}_NC_n=N!/(n!\,(N-n)!)$ と分数にして，さらに分母と分子をそれぞれ $\sqrt{2\pi N}\,N^N\,\exp(-N)$ で割っておく．公式

$$\lim_{N\to\infty}\frac{N!}{\sqrt{2\pi N}\,N^N\exp(-N)}=1$$

を $N!$ と $(N-n)!$ に適用する．$(\frac{\lambda\,(t-s)}{N})^n$ からも分母に N^n が出ることに注意すると，右辺は

$$\lim_{N\to\infty}\frac{1}{\sqrt{1-n/N}\,(1-n/N)^{N-n}\,\exp(n)\,n!}\,(\lambda\,(t-s))^n\left(1-\frac{\lambda\,(t-s)}{N}\right)^{N-n}$$

と変形される．a,b が N によらないとき成り立つ公式 $\lim_{N\to\infty}(1-a/N)^N=\exp(-a)$ 及び $\lim_{N\to\infty}(1-a/N)^b=1$ を用いれば (3) を得る．

定理．

X_t が変化する（即ち 1 だけ増える）ことを，event が発生した，ということがある．この表現の気持ちは明らかであろう.（時刻ゼロから測定開始して）初めて event が発生した時刻を T_0, 以下 n 件目と $n+1$ 件目の event 発生の **時間間隔** を T_n $(n\geq 1)$ とおく．T_n たちは $\{X_t\}$ で定まる確率変数である．これについて次の性質が証明できる．

X_t が Poisson 過程ならば，$T_0,T_1,T_2,\cdots$, は独立で，全て同じ分布を持つ（この性質を **独立同分布** であると言い，i.i.d. と書く）．T_0 の分布は平均 $\mathrm{E}[\,T_0\,]=1/\lambda$ の指数分布になる（どの T_n でも同じ）．ここで λ は (3) の λ，即ち，単位時間内の到着数の期待値 $\mathrm{E}[\,X_1-X_0\,]$ である．

（平均 $1/\lambda$ の）指数分布とは $t\geq 0$ 上の分布で，密度が

$$\rho_\lambda(t)=\lambda\exp(-\lambda\,t) \tag{4}$$

で与えられるものを言う．例えば，時間間隔 T_n が a 以上である確率は

$$P[T_n\geq a]=\int_a^\infty\rho_\lambda(t)\,dt$$

で計算される．

定理の証明．

$\{T_n\}$ が独立同分布であることは，Poisson 過程の加法性と一様性から導かれるのだが，省略する（以下の証明にかなり含まれている）．T_n が平均 $1/\lambda$ の指数分布に従うことだけ証明しよう．最初に述べたことによ

り，Poisson 過程の定義から (3) が導かれることが（証明は完全にはしなかったが）分かっているので，(3) から (4) を得ることができればよい．$a \geq 0$ を定数とする．T_0, T_1, $\cdots$, T_{n-1} をそれぞれある値に固定するという条件の下で $T_n > a$ となる条件付き確率（これを $P[T_n > a \mid T_0, T_1, \cdots, T_{n-1}]$ と書く）を計算する．$S = \sum_{i=0}^{n-1} T_i$ とおいておく．T_n の定義から，求める確率は，時間間隔 $(S\,,\ S+a]$ の間に event が発生しない確率，即ち，この時間に X_t が変化しない確率に等しい．従って，

$$P[T_n > a \mid T_0, T_1, \cdots, T_{n-1}] = P[X_{S+a} = X_S \mid S] = \exp(-a\,\lambda)\,.$$

最後の等号は (3) において $n = 0$, $s = S$, $t = S + a$ とおいて得られる．右辺は S によらないから，T_0, T_1, $\cdots$, T_{n-1} の値によらない（即ち，これらの変数と T_n は独立である）．だから，条件をはずしても等号が成り立つ；$P[T_n > a] = \exp(-a\,\lambda)$. この式と，$T_n$ の分布の密度 $\rho_\lambda(t)$ の定義 $P[T_n \geq a] = \int_a^\infty \rho_\lambda(t)\,dt$ を比べれば $\rho_\lambda(t)$ は　$\exp(-a\,\lambda)$ を a で微分して符号を変えたものになることが分かる．これは（a を t と書き換えれば） (4) の右辺に他ならない．

定理．

上の定理の逆も成り立つ．

$\{T_n\}$, $n = 0, 1, 2, \cdots$, が独立同分布で，平均 $1/\lambda$ の指数分布ならば，

$$X_t = \min\{n \mid \sum_{i=0}^{n} T_i > t\}$$

で定義される確率過程 X_t, $t \geq 0$, は Poisson 過程になる．

この定義の右辺は時刻 t までに発生した event の総数を表している（各自確認せよ）．この定理の証明は省略する．X_t の加法性と独立性を証明すればいいのだが，$\{T_n\}$ が独立同分布であることだけでなく，指数分布に従うことも用いる必要がある．

以上の事実を用いると，いろいろな量を計算できる．既に定理の中で，λ が単位時間当たり通る車の台数の平均 $\mathrm{E}[\ X_{t+1} - X_t\ ]$ に等しく，さらに，平均通過時間間隔の逆数 $1/\mathrm{E}[\ T_n\ ]$ に等しいことを示した．これが，平均待ち時間の逆数に等しいことは，前節で（以上の結果を先取りして）証明した．平均待ち時間と言ったのは，ある時刻 s に停留所に到着した乗客がバスがくるまで待つ時間の平均値 $\mathrm{E}[\ \sum_{i=0}^{X_s} T_i - s\ ]$ のことである（この式がそういう意味を持つことは各自確認せよ）．前節で述べたようにこの値は，$T_n > s'$ なる条件下での $T_n - s'$ の期待値（条件付き期待値） $\mathrm{E}[\ T_n - s' \mid T_n > s'\ ]$ に等しいことが証明できる．そしてそれが $1/\lambda$ に等しいことは，前節で計算したとおりである．

最後に，極めて重要，かつ驚くべきことは，このように，計算しやすい性質をいくつも持つ確率過程が本当に存在するという事実である．講義では触れなかったが，Poisson 過程の定義を満たす確率過程が，適当な確率空間の上の確率過程（確率変数の集まり）として，存在することが証明されている．

参考．

上の定理は event 発生の累積観測データを得たとき，それが Poisson 過程になっている（つまりでたらめな event）かどうかを確かめるのに便利である．(i) まず，累積到着数 sample path $X_t(\omega)$ を多数集める．1 時間毎に（毎日，など）データを異なる sample path とみなす.（t は 0 から 1 時間（1 日）しか動かないことになる.）(ii) それぞれ，最初の event 発生時刻を $T_0(\omega)$ とし，以下順に発生間隔を $T_n(\omega)$ に割り振る．(iii) 各 n ごとに T_n の sample として割り振られたデータのヒストグラム（度数分布）または，分布関数，を作って，それが指数分布になることを確認する．(iv) 異なる n の間の T_n が確率変数として独立なことを確認する．

以上を満たせば，定理より，この観測データは Poisson 過程になっていると結論される．Poisson 過程の定義（事実上，加法性と一様性）をチェックする方法も良い．

問１．

Poisson 分布の定義（$Q[\{n\}]=\exp(-\mu)\,\frac{1}{n!}\,\mu^n$）を用いて (2) を証明せよ．さらに，定義に出てくる（n によらない）定数 $\exp(-\mu)$ は全事象の確率が（確率の定義通り）１になるように決まっているはずである；

$$\sum_{n=0}^{\infty} Q[\{n\}]=1\,.$$

この式を証明せよ．

また，指数分布の密度 (4) についても，全事象の確率 $\int_0^\infty \rho_\lambda(t)=1$ であること，及び，平均が $1/\lambda$ になること，を証明せよ．最後に，平均 $1/\lambda$ の指数分布に従う確率変数 T が，定数 a 以上になる確率 $P[T\ge a]$ を計算せよ．

問２．

計算機の乱数を利用して Poisson 過程 X_t の sample path を発生して見よ．乱数の初期値を変えていろいろ発生させ，いくつかを，X_t を縦軸 t を横軸にして図示せよ．Sample を多数作って，適当な s,t をいくつかとって，時間 $(s,t]$ の間の到着数 X_t-X_s の統計分布（0 の sample がいくつ，といったふうに）をとってみよ．また，Event 間隔 T の統計分布も図示して見よ．

Sample の発生には，event 発生間隔 T が指数分布に (4) に従うことを使うのが便利かも知れない．(計算機で，区間 $[0,1)$ の一様分布に従う乱数 w から，指数分布に従う乱数 u を作るには，$u=-\log(w)$ と置けばよい.）

解答はプログラムもつけること．Mathematica などの高級ソフトの利用も歓迎するが，どの場合でも，プログラムまたはソースをつけること．

問３．

ランダムウォークの問で発生させた１次元酔歩について，いろいろな乱数の初期値をとって，１次元酔歩の長い（N 歩目までとるとする）sample をたくさん（T 本とるとする）用意せよ．第 i $(1\le i\le T)$ sample の n 歩目の位置を $X_n(i)$ とする．このとき，X_n の分散の sample 平均

$$V_n=\frac{1}{T}\sum_{i=1}^{T} X_n(i)^2$$

が n のどのような関数になているか，図示してみよ.（ヒント．N よりもむしろ T を大きくとる必要がある．乱数は直前の sample を作るのに使った最後の値を続けて使えばよい．十分 sample 数 T が大きければ $V_n=n$ に近づくはず.）

解答はプログラムもつけること．Mathematica などの高級ソフトの利用も歓迎するが，どの場合でも，プログラムまたはソースをつけること．

6　マルコフ性．

6.1　マルコフ過程．

前の節で，確率過程の実例をみてきた．時間変数と状態空間が離散的な確率過程の代表例として random walk，時間変数も状態空間を連続な代表例として Wiener 過程，時間変数が連続で状態空間が離散的な代表

例として Poisson 過程を取り上げた．ところで，Random walk，Wiener 過程，Poisson 過程，はいずれも **マルコフ性** を持つ．マルコフ性は理論上も応用上も大変本質的な役割を果たす性質である．

確率過程の分類学上，Wiener 過程と Poisson 過程はともに，実数 $\mathbf{R}$ 上の，時間的にも空間的に一様な，マルコフ過程である．全く異なる性質を持つが，ある観点からは「親類」である．

マルコフ性を持つ確率過程（マルコフ過程）は，非マルコフ過程に比べて，一般に格段に計算が容易である．確率論以外の数学の分野との深い関連がはっきりしていて，それを利用していっそう詳しい計算ができる．そのため，応用上も広く用いられる．

いつもの通り，確率空間 $(\Omega, \mathcal{F}, P)$ と，その上の実数値確率過程 X_t が与えられているとする．即ち，X_t の状態空間は $\Sigma = \mathbf{R}$ である．

X_t が **マルコフ過程** であるとは，任意の自然数 n と，任意の $n+1$ 個の時刻 $s_1 < s_2 < \cdots < s_n < t$，及び，任意の（$\Sigma$ の）事象 $A \in \mathcal{B}$，任意の n 個の Σ の点（つまり，値） $y_1, y_2, \cdots, y_n\ (\in \Sigma)$，に対して

$$P[X_t \in A \mid X_{s_1} = y_1,\ X_{s_2} = y_2,\ \cdots,\ X_{s_n} = y_n] = P[X_t \in A \mid X_{s_n} = y_n] \tag{5}$$

が成り立つことをいう．ここで，$P[A \mid B] = P[A \cap B] \,/\, P[B]$ は，B が起こったという前提の下で A の起こる確率，即ち，条件付き確率である．

大雑把にいうと，マルコフ過程とは，時刻 t における位置（値）の分布 X_t は直前の時刻 s_n での位置 X_{s_n} だけで決まり，それ以前の時刻 $s_1, \cdots, s_{n-1}$ にどこにいたかによらない，ということである．すごろくの例は，X_n の分布が直前の位置 X_{n-1} だけで定まり，それ以前の状態（過去の履歴）によらない（独立になる）．離散時間のマルコフ過程の例である．既に Poisson 過程と Wiener 過程の例を挙げた．両者とも定義によって加法性を満たすが，（$X_0 = 0$ を用いれば）加法性からマルコフ性が導けるので，Poisson 過程と Wiener 過程は連続時間マルコフ過程の例である．

非マルコフ的な，即ち，直前の時刻の位置だけで決まらない，というのは，小説や劇，ドラマの伏線という概念に例えられる．主人公が，直前の状況からは，突拍子もなくみえる行動に走るシーンがある．その行動の理由を説明する背景が最初の方にさりげなく描かれていた，というストーリーの構成法である．突拍子もない行動とは，そのような行動が起こる確率が小さい，ということであろう．直前の状態の下では確率の小さい行動が，過去のある条件の下で高い確率になる，即ち過去に依存する非マルコフ性がある．確率過程が過去の「記憶」を持っている，とも言えるだろう．

§5.2 の式 (5) の右辺の形の条件付き確率を遷移確率（推移確率）と言い，

$$P(s, y, t, A) = P[X_t \in A \mid X_s = y]\,, \quad (s < t,\ \ y \in \Sigma,\ \ A \in \mathcal{F})\,,$$

と書く．マルコフ過程は，遷移確率 $P(s, y, t, A)$ と初期分布（X_0 の分布）だけで決まることが知られている．

$$P(s, y, t, A) = P[X_t \in A \mid X_s = y] = \int_{x \in A} \rho(s, y, t, x)\, dx$$

と書けるとき，ρ を**遷移確率密度** という．Wiener 過程について既に見たように，

$$P[X_t \in A \mid X_s = y] = \int_{x \in A} \rho_{0, t-s}(x - y)\, dx$$

が成り立つので，その遷移確率密度は，正規分布の密度

$$\rho_{m,v}(x) = \frac{1}{\sqrt{2\pi v}} \exp\left(-\frac{(x-m)^2}{2\, v} \right)$$

を用いて

$$\rho(s, y, t, x) = \rho_{0, t-s}(x - y) \left(= \frac{1}{\sqrt{2\pi(t-s)}} \exp\left(-\frac{(x-y)^2}{2\,(t-s)} \right) \right), \tag{6}$$

と書けることが分かる．

Wiener 過程は sample path が（確率 1 で）連続関数である．一般に sample path が（確率 1 で）連続関数になる連続時間マルコフ過程を**拡散過程** と呼ぶ．遷移確率密度が (6) と書ける拡散過程を **Brown 運動** という．Brown 運動で初期状態を $X_0 = 0$ としたのが Wiener 過程である．

Poisson 過程は状態空間 $\Sigma = \mathbf{Z}$ が離散的なので，遷移確率密度はないが，遷移確率が簡単に書ける．X_n が Poisson 過程の場合，§5.3.2 の (3) から直ちに次を得る（(3) では増分を n とおいた）．

$$P_{Poisson}(s,m,t,n) = P[X_t = n \mid X_s = m] = e^{-\lambda\,(t-s)}\,\frac{\{\lambda\,(t-s)\}^{(n-m)}}{(n-m)!}\,, \quad s<t, \quad m \le n\,.$$

参考：現代のマルコフ過程の定義では，状態空間 Σ は局所コンパクト Haudorff 空間で第 2 可算公理を満たすものならばよい．また，消滅時刻 $0<\zeta\le\infty$ と消滅点 ∂ を考えて，$X_t=\partial,\ t\ge\zeta$, をマルコフ過程の定義に加える．$\zeta=\infty$ が非消滅に対応.

マルコフ性については話がつきないのでここまでとし，Wiener 過程に戻る．あとの準備もかねて，Wiener 過程と random walk の状態空間 Σ を d 次元空間 $\mathbf{R}^d$ に拡張する．d を空間の次元を表す自然数とする．(いままでの確率過程の例は $d=1$ に対応する．) $\mathbf{R}^d$ 上の連続時間確率過程を考える．d 個の実数値確率過程を成分に持つベクトル値確率過程 $X_t = (X_{1,t}, X_{2,t}, \cdots, X_{d,t})$ ということである（数学の本ではベクトルでも $\vec{X}_t$ とせず，X_t などと書くことが多い）．

X_t が d 次元 Wiener 過程であるとは，各成分 $X_{i,t}$ が Wiener 過程であって，成分 $X_{1,\cdot}, X_{2,\cdot}, \cdots, X_{d,\cdot}$ が独立なものを言う．各成分毎に独立に計算すればよいので，§5.2 の (1 次元)Wiener 過程の公式を用いることができる．例えば d 次元 Wiener 過程の遷移確率密度は

$$\rho(s,y,t,x) = \frac{1}{\sqrt{2\pi\,(t-s)}^{\,d}}\exp\left(-\frac{1}{2\,(t-s)}\sum_{i=1}^{d}(x_i-y_i)^2\right), \quad x=(x_1,x_2,\cdots,x_d), \quad y=(y_1,y_2,\cdots,y_d), \tag{7}$$

と書けることは，成分間の独立性と §5.2 の諸公式から証明できる．d 次元 Wiener 過程もマルコフ過程である（拡散過程でもある）．

d 次元 Wiener 過程に対応して，1 次元酔歩 (random walk) を d 次元 random walk に拡張する．1 次元 random walk は，原点から出発する整数時刻の整数値確率過程 Z_n で，sample path が $Z_n(\omega)=\sum_{m=1}^{n}\omega_m$, $n=0,1,2,\cdots$, と書けるものであった．ここで，$\{\omega_m\}$, $m=1,2,\cdots$, は独立で，それぞれ確率 $1/2$ で $+1$ または -1 となる．d 次元 random walk は，状態空間が d 次元格子空間 $\mathbf{Z^d}$ の場合である．即ち，Z_n が d 成分ベクトルで，各成分が整数値をとる；$Z_n=(Z_{1,n},Z_{2,n},\cdots,Z_{d,n})$. そして，各時刻毎に次に，現在の位置 Z_n の（$2d$ 個の）隣の点のどれかに各確率 $1/(2d)$ で移る．言い換えれば，$Z_n(\omega)=\sum_{m=1}^{n}\omega_m$ であって，$\{\omega_m\}$, $m=1,2,\cdots$, は独立，かつ，各 ω_m は，それぞれ確率 $1/(2d)$ で $2d$ 個のベクトル $(+1,0,0,\cdots,0)$, $(-1,0,0,\cdots,0)$, $(0,+1,0,\cdots,0)$, $\cdots$, $(0,0,0,\cdots,+1)$, $(0,0,0,\cdots,-1)$, になる．種々の性質は 1 次元と同様である（計算は煩雑になる場合がある）．例えば，d 次元 random walk と d 次元 Wiener 過程の間にも §5.2 で紹介した scaling limit の関係がある．

定理．

d 次元離散時空確率過程 $c^{-1/2}\,Z_{[c\,t]}$ は $c\to\infty$ のとき d 次元 Wiener 過程 X_t に法則収束する．

§5.2 の定理と，c のべきの値 $-1/2$ まで含めて，全く同じ定理が全ての d で成り立つ．

6.2 Self-avoiding walk.

確率過程の典型例が全てマルコフ過程であったように，これまでの確率過程論の研究は（応用も）マルコフ過程中心であった．マルコフ性の成り立たない例として self-avoiding walk （自己回避道）を取り上げる．Self-avoiding walk とは空間上の道であって同じ点に二度と戻らないものである．

d 次元 random walk で N 歩目までの道のりを考えると，原点から出発して隣の点へ移りながら N 歩でたどれる可能な全ての歩き方が等しい確率（確率 $(2d)^{-N}$）で出現する．w が d 次元格子空間 $\mathbf{Z^d}$ 上の長さ N

の walk であるとは $\mathbf{Z^d}$ の列 $w=(w(0),w(1),\cdots,w(N))$ であって，

$$|w(i)-w(i+1)|=1,\ i=0,1,\cdots,N-1,$$

を満たすものである，と定義する．$|\cdot|$ は通常のユークリッド距離である．上式の条件は，一歩毎に隣の格子点に移る，ということを表す．すると，random walk という確率過程の定義は，原点から出発する $(w(0)=0)$ 長さ N の walk がどれも，等しい確率 $(2d)^{-N}$ を持つ確率空間（が全ての N で定義されていること），と見ることができる．（これは §5.1 で説明した２つの見方のうち,「神の立場」と言えよう.）これに対して，w が d 次元格子空間 $\mathbf{Z^d}$ 上の長さ N の **self-avoiding walk** とは長さ N の walk $w=(w(0),w(1),\cdots,w(N))$ であって，$w(j)\neq w(i),\ j\neq i,$ を満たすものと定義する．W_N を原点から出発する $(w(0)=0)$ 長さ N の self-avoiding walk の集合とする．この集合の各要素に（random walk のときのように）等しい確率を与えたのが d 次元 self-avoiding walk の確率空間である．条件 $w(j)\neq w(i),\ j\neq i,$ は一度通った点は２度と通らないという意味である．過去の状態（通った点）を全て覚えていないと次の動きが決まらないから，非マルコフ性を持つ典型的な例である．これが self-avoiding （自己回避的）という名前の由来である．

$$\mathrm{E}[\ |w(N)|^2\ ]=\frac{1}{C_N}\sum_{w\in W_N}|w(N)|^2$$

は self-avoiding walk が N 歩で出発点からどれくらい遠くまで到達できるかを測る期待値で，mean-square displacement と呼ばれる．C_N は 原点から出発する N 歩の self-avoiding walk の本数.

$$\nu=\lim_{N\to\infty}\frac{\log\mathrm{E}[\ |w(N)|^2\ ]}{2\log N} \tag{8}$$

とおく．Random walk について同様の期待値をとって計算すると，d の値によらず $\nu=1/2$ となる（ §6.1 終わりの定理で c のべきが d によらず $-1/2$ になることに対応する）．$d=1$ の self-avoiding walk とは直線上を（右か左に）まっすぐ進むことだから，$|w(N)|=N$ となるので，(8) より $\nu=1$ である．$d\geq 2$ では self-avoiding walk の性質は難しい．$d\geq 5$ の self-avoiding walk では random walk と同じ値 $\nu=1/2$ になることが Hara–Slade によって最近証明された．高い次元の self-avoiding walk の，N が大きいときの性質は random walk のそれに近い，という意味である．

$d<4$ のときは self-avoiding walk の ν の値は 1/2 からずれる，と予想されている．予想値 (Flory の値) $\nu=\frac{3}{d+2}$ は数値的にはいい近似値になっている．低次元（小さい d）で値が random walk の値 1/2 からずれるということは，低次元空間上の self-avoiding walk では非マルコフ性（自己回避効果）が顕著に現れることを意味する．しかし，$d\leq 4$ の self-avoiding walk については証明はまだない.（参考書： N. Madras, G. Slade, The Self-Avoiding Walk, Birkhäuser, 1993 年.）

「少しだけ，確率過程」という題は，分野の広がりに比べて少ししか講義できない，という意味である．確率過程は奥行きの深い分野で，Wiener 過程についてすら，紹介できなかった多くの研究成果がある．それでも，これからの数学研究の発展が期待される応用上の問題も多い．種々の有益な計算ができるほど詳しい性質が明らかになっている確率過程はマルコフ過程だけ，と言ってよく，今回紹介したマルコフ性を持たない self-avoiding walk は今後の重要な課題の一つである．非マルコフ過程は物理や経済への応用の機運が高い．例えば，株価変動は不規則に見えるが（新聞などの株価のグラフを見よ），これを適当な確率過程（の sample path）と見よう，というのである．

問１.

原点から出発する N 歩の２次元 $(d=2)$ self-avoiding walk （$w(0)=0$ で，１歩毎に平面内の４つの隣の格子点のどれかに移り，一度通った点は２度と通らない N 歩の歩き方）の総本数 C_N をいくつかの N について求めよ．例えば $C_1=4,\ C_2=12$.（計算機で大きな N まで数え上げればなお望ましい．ν の定義 (8) の右辺の，lim の中身（極限をとる前の量）が各 N でいくらになるかも求められればなおよい.）

計算機による解答にはプログラムもつけること．Mathematica などの高級ソフトの利用も歓迎するが，どの場合でも，プログラムまたはソースをつけること．

問２．

(d=1) Wiener 過程の遷移確率密度 $\rho(s,y,t,x)$ は次の形の偏微分方程式を満たすことが，具体形 (6) を微分することで確かめられる．下式が成り立つように定数 A の値を求めよ．

$$\frac{\partial \rho(s,y,t,x)}{\partial t} = A\,\frac{\partial^2 \rho(s,y,t,x)}{\partial x^2}\,.$$

一般の d について，d 次元 Wiener 過程の遷移確率密度 (7) はどんな偏微分方程式を満たすか？（ヒント：2 回偏微分が d 次元ラプラシアンに一般化される．なお，これらの偏微分方程式は拡散方程式と呼ばれる.）

問３．

これまでの課題のうちで興味を持ったものを，引き続き発展させよ．例えば，(i) 証明し残した性質を証明する，(ii) より詳しい計算を行う，(iii) 数値シミュレーションのデータをはるかに増やす，など．

計算機による解答にはプログラムもつけること．Mathematica などの高級ソフトの利用も歓迎するが，どの場合でも，プログラムまたはソースをつけること．