# 確率・統計入門

## この講義録を拡充して図版や例および練習問題を多数加えた「統計と確率の基礎」が学術図書から2006年に刊行しました．

教科書に特化した会社なので書店店頭にはありません．店頭注文または web 注文をお願いすることになりご迷惑をおかけします．詳しくは「**服部哲弥**」（弥の字に注意）で検索していただくか
`http:/www.math.tohoku.ac.jp/~hattori/gakjutu.htm`
をご覧下さい．

この講義録は不満が残っていますが，改訂は打ち切りました．よろしくお願い申し上げます．

toukei.tex

服部哲弥

19960107;27;30 etc.;0410;13–16;21;27-30;0501;03-16;18;19-20(rnd);21-23;
24-26(stat);27-29;0602;08;14-18;20-28(ver 0);
0707(realsinfty);12(egg data);15(appendix);17(egg data);26(irrelev.); 27(egg data);1229(corrections);
19991024;
20040606(corrections);

# 確率・統計入門

章目次

# 第 0 部
# 序

## -1　イントロ（講義の動機付け）

01

> 「確率とは何か」ということに関しては古くから論争があり，現在も決着はついていない．これについてはいろいろなものの考え方があり，おそらく永久に決着がつくことはないと思われる [楠岡].

**数学とは何か？**　数学科で特に学べることであって，社会では習う機会の少ないものの見方考え方，普通の人が学ばなかった特技，としての数学的理解，とは何か？いろいろあるはず．本来は各自自分が何を得たかを考えるべきこと．例えば，

**理想化**　人間にとって分かりやすい（数学は分からないというのは壮大な幻想である）．知らないうちに思考の節約をしていることになる．

**厳密性**　再現性＝誰がやっても何回やっても同じ結論＝安心できる．期待通りのことが起きる．(cf. 占い＝当たるも八卦当たらぬも八卦．エステのCM.）

**精密性**　わずかな有利不利を議論できる可能性．強力な手段．(cf. くだらない宗教の予言「明日は晴れか晴れでないかのどちらか」.）→抽象性素朴には理解できない部分にまで立ち入る．

**普遍性**　種々の問題をまとめて解くことができる．公式．

**講義の目的**　数学的理解の現実的価値に関する一般論は全て確率論について当てはまる (§1)．この講義は何をやるか？以下の意味での確率論入門.

- 現代の確率論は測度論に立脚した定式化によって厳密かつ明快な優れた数学になっているが，測度論を講義する解析学 I は３年の科目なので，詳しい確率論の展開はこの講義では行わない．
- 詳しい数学的理論展開を諦める代わりに，数学の「外側」のお話を織り込むことで他の講義に対する特色としたい．例えば，数学や確率論への動機付けや現実への応用などである．

**教科書**　講義の前半は [楠岡 2]，後半は [小針]．特に，これらの教科書の問題は是非自習してほしい．

**参考書**　この講義で触れない確率論の進んだ概説 [楠岡].

**確率とは何か？**　現象（または描像）：決定論と非決定論.

- 決定論：ものごとの動きは一通りに定まっている．例：古典力学．法則の形が，現在の状況（初期条件）を与えると未来が一通りに決まるようになっている．
- 非決定論：いろいろな可能性がある．例：保険，賭事，釣り．現在の状態を完全に与えても，未来を一通りに定められない．

どちらの立場をとるかで，行動原理が全く異なる．例えば，電車に乗って待ち合わせの場所に行く場合，電車が決定論に従って時刻表通り動くと考えるか電車のスケジュールにはばらつきがあると考える（非決定論）か.

**決定論**　約束の時刻より前のできるだけぎりぎりにつくように時刻表から電車を選ぶ.

**非決定論１**　５分くらいの遅れがある場合を想定して，一本前の電車にする．

**非決定論 2** 大きな事故で他の路線による振り替え輸送になる場合も考えて，駅には予定より３０分ほど早く着くようにして，あまった時間は近くの本屋や図書館でつぶす.

決定論的理解の場合それに応じた行動はたいてい確定するが，非決定論の場合，他の条件（約束に遅れることがどれくらい重大か）との関係でしか結論が出ない点も難しさの一つ.

非決定論的状況でも殆ど一つの選択肢だけが生じる場合（例：通常電車は時刻表通り来る）であって，外れた場合の損害が小さい場合（例：待ち合わせの相手が時刻にうるさくない）は決定論的に考えて行動しても問題はない．また，非決定論的な法則の理解や応用は難しいので，多くの場合，人は世の中が決定論に従っていると考えて行動する．だからこそ，期待と異なる事態が生じると驚く．例：競馬で大穴が出たとき，大事故の新聞報道.「ねずみがねこをかむと新聞記事になる.」

ルーレットやさいころのように複数の場合が小さくない可能性で起きるときや，飛行機事故に対する保険金の決定や原子炉設計や耐震構造の基準の決定のように，めったに起きなくても結果が重大なときは，非決定論的に考える必要があるはず.

非決定論的な状況の原因はいろいろある．例えば正しい値を一つ求めたい測定実験の場合，正しい値は一つのはずでも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．天気予報のように観測能力や予測能力の限界から予測値（雨の量など）が不確実な場合もある.

確率は非決定論的な現象を理解する一つの数学的概念である.

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.

不規則性の分析は現実のシステムのあり方を決める上で積極的な意味を持つ．有名な分布の例に偏差値がある．能力に分布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.

さらに積極的に不規則性が自然法則として観測される場合もある．物理学の統計力学では非決定論を前提としている．ブラウン運動のように確率論で明快かつ精密に説明できる自然現象もある．保険は確率論が産業に応用されている例である.

不規則性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.

## 0 確率論と統計学

確率論を現実に適用するときの特徴的な問題．例：天気予報で「明日の降水確率が４０％」とあって，実際は雨が降った場合，天気予報は当たったか当たらなかったか？１日限りの観測では判定のしようがない.

確率モデルを現実に当てはめるときに当てはまったかどうかの検証自体が高度な考察を必要とする．分布やパラメータをどう選ぶかもこの考察に基づく.

**確率論** 数学的対象として確率モデル自体を論じる

**（数理）統計学** 現実への確率モデルの適用，モデルの選択やパラメータの決定や信頼性などを論じる

実験や観測を繰り返すと平均値や分散などの（さらには分布も）統計量が安定した値や形に近づくことが観測されれば，客観的な確率が現実に存在すると言えるだろう.（例えばさいころの各目が出る確率が 1/6 に近づくことが観測された場合.）このとき確率論で記述できる現象（法則）が存在すると結論できる．観測から結論に至る前提として統計学が必要．逆に，ある自然現象が確率論で記述できると仮定すれば，そして，現象を記述する確率モデルを決めれば，そこから予測や説明をするために確率論が必要.

講義目標.

- 非決定論的現象を数学的に理解し利用すること.

- 確率論の視点（数理的な確率モデルの適用）.
- 確率論と（数理）統計学の役割分担.

講義目次.

(i) ０．序．確率論と統計学に関連するいろいろな「お話」，動機付け
(ii) １．確率論からの準備．確率論の基礎事項（統計学の基礎となる部分使う部分だけ，直感的に説明し証明はあまりしない）
(iii) ２．データ．確率論的視点から見たデータ，模擬的なデータの発生（実験）としての乱数
(iv) ３．統計学入門．統計的推測（推定と検定，特に正規母集団に関する推測）
(v) ４．その他の話題．発展的話題（余裕があれば，確率過程論などの紹介）

４年生の講義は確率論．今回の講義は前半が確率論のお話，後半が統計学入門.

# 1 数学としての確率論 02

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.

こだわるならば,「確率とは何か」という問に答える定義ではなく，以下に示すいくつかの数学的性質を満たす対象を確率空間と呼んで考察の対象とする，という宣言（公理化）である.

現代の確率論の決定版は N. Kolmogorov の公理 (1933) から出発する．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.

$\Omega$ （オメガ，ギリシャ文字） を集合とする（全体集合）.

$\Omega$ の部分集合の集合族（部分集合からいくつか選んだ集まり，集合の集合）であって空でないもの $\mathcal{F}$ が $\sigma$-加法族（完全加法族，可算加法族）であるとは以下を満たすことと定義する.

**補集合** $A \in \mathcal{F}$ ならば $A^c \in \mathcal{F}$ （$A^c$ は補集合）,

**可算和** $A_n \in \mathcal{F}$, $n \in \mathbf{N}$, ならば $\displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$,

**例** $\Omega = \{0,1\}$ のとき 部分集合は $\phi$, $\{0\}$, $\{1\}$, $\Omega$．部分集合族としては $\{\phi\}$, $\{\phi, 0\}$, $\cdots$ がある.（問．$\sigma$-加法族になっているのはどれか？）

$A_n \subset \Omega$, $n = 1,2,3,\cdots$, が $n \neq m$ ならば $A_n \cap A_m = \phi$ を満たす時に限り，$\displaystyle\bigcup_{n=1}^{\infty} A_n$ のことを $\displaystyle\sum_{n=1}^{\infty} A_n$ と書く.「$n \neq m$ ならば $A_n \cap A_m = \phi$ を満たすとき」という文を省略するための記号法として用いる.

$P$ が $(\Omega, \mathcal{F})$ 上の確率測度であるとは $\mathcal{F}$ で定義された集合関数（$\mathcal{F}$ に属する集合を与えると実数値が決まる関数）$P:\ \mathcal{F} \to \mathbf{R}$ であって，以下を満たすことと定義する.

**非負** $P(A) \geq 0$, $A \in \mathcal{F}$,

$\sigma$ **加法性** $P(\displaystyle\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, $A_i \in \mathcal{F}$, $i \in \mathbf{N}$,

**規格化** $P(\Omega) = 1$

$\mathcal{F}$ の要素 $A \in \mathcal{F}$ を**事象**，$P(A)$ を事象 $A$ の**確率**という.

組 $(\Omega, \mathcal{F}, P)$ を確率空間と呼ぶ.

**注**. (i) $\sigma$-加法族の定義から，$A_n \in \mathcal{F}$, $n \in \mathbf{N}$, ならば $\displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ なので $P(\displaystyle\sum_{n=1}^{\infty} A_n)$ が定義されている.

(ii) $\sigma$-加法族の定義から，有限和に関して閉じていること，積 $A \cap B \in \mathcal{F}$ （de Morgan を使う），差，空集合 $\phi \in \mathcal{F}$, 全体集合 $\Omega \in \mathcal{F}$, 高々可算積に関して閉じていること，確率の連続性 $A_n \supset A_{n+1}$ ならば $P(\bigcap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} P(A_n)$，などが従う．特に，$P(\Omega)$ が定義されている．よって，確率測度の定義は well-defined.

(iii) ある集合族が$\sigma$-加法族であることを証明するときは 2 つの性質のみ証明すればよく，$\sigma$-加法族であると分かっている（仮定している）族に対しては直ちに上記全ての性質を持っている前提で話を進めてよい，というように使い分ける．以下でも同様.

## 2 気持ち

§1 の性質を満たす組 $(\Omega, \mathcal{F}, P)$ はなんでも確率空間と呼び,「確率とは何か」という問に答える定義ではない．実際，例えば $\Omega$ を何と思うかという「気持ち」には任意性がある．しかし，しばしば背後に想定される「気持ち」があるのも事実である．公理の背後にある確率空間の定義の気持ちを知ることは理解の助けになる.

全体集合 $\Omega$ とは，考察対象の空間，即ち，想定する事態の集合．制御できない撹乱の結果の全体である．皮肉な言い方をすれば，確率的現象においては $\Omega$ の元のたった一つが実現し，残りは架空の「起きたかも知れないこと」である.（従って，確率論の現代的定義に即しても，天気の確率予報は実際に晴れたら当たったと言えるかどうか，という問題はそのままでは解決しない.）例．さいころの目 $\Omega = \{1,2,3,4,5,6\}$．$\Omega$ の要素 $\omega \in \Omega$ 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.

公理で要求する性質を見てみよう．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．そのときの直感に基づく素朴な計算はそのまま正しいと考えたい.

$P(\Omega) = 1$ 即ち全体集合は確率 1 というのは必ず何かが起こるという意味と理解すれば素朴な確率のイメージからは当然成り立つべきことである．$P(A) \geq 0$ も確率が負にならないという当然のこと.

$\sigma$-加法族は「確率が定義できる $\Omega$ の部分集合の全体」のこと．余事象の確率は $P(A^c) = 1 - P(A)$ であるべきだから，$A$ の確率が定義されていれば，余事象の確率も定義されているはず．即ち $A \in \mathcal{F}$ ならば $A^c \in \mathcal{F}$.

有限個の事象の場合は素朴な確率論のイメージから和事象の確率が定義できるべきだし，排反な事象の和の確率は個々の事象の確率の和である．即ち，$A_n \in \mathcal{F}$, $n = 1, 2, \cdots, N$, ならば $\bigcup_{n=1}^{N} A_n \in \mathcal{F}$, および $P(\sum_{i=1}^{N} A_i) = \sum_{i=1}^{N} P(A_i)$ （有限加法性）．しかし，数学としての精密性のためには極限が取れないと不十分である（微分のように）．そこで，可算個の事象についても以上の類推が成り立つことを要求すれば，$A_n \in \mathcal{F}$, $n \in \mathbf{N}$, ならば $\bigcup_{n=1}^{\infty} A_i \in \mathcal{F}$, および $P(\sum_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

従って，Kolmogorov の公理が要求する性質は順当なものばかりである．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（可算加法性，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.

残された問題.

(i) 可算加法性よりもっと強く $P(\sum_{i\in\Lambda} A_i) = \sum_{i\in\Lambda} P(A_i)$ を任意の濃度の集合 $\Lambda$ に対して要求してはいけないのか？

(ii) なぜ確率 $P$ は集合関数として $\Omega$ の部分集合（事象）に対して定義するのか？$\Omega$ 上の普通の関数として $\Omega$ の元に対して定義しないのか？

(iii) なぜ$\sigma$-加法族を考えるのか？なぜ，$\Omega$ の部分集合全てに確率を定義しないのか？$\Omega$ の部分集合全てからなる集合（集合族）を $2^{\Omega}$ と書く．例．$\Omega = \{0,1\}$ ならば $2^{\Omega} = \{\phi, \{0\}, \{1\}, \Omega\}$．素朴に考えれば $\mathcal{F} = 2^{\Omega}$ でもよいはず．そういう場合もある．

これらの疑問に答えるために次の例を見よう．区間 $(0,1]$ 上の**一様分布**を考える．この分布は直感的には $A$ を区間 $(0,1] = \{x \mid 0 < x \leq 1\}$ 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．なお，ここでは区間を $0 < x \leq 1$ と 0 を含まず 1 を含む定義にしているが，これは趣味の問題で両端を含むか含まないかは本質的な意味はない．

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 $\mathrm{P}(\ 0 < X \leq 0.1\ ) = 0.1$ などとなる．$0 \leq x < 1$ 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．

$(0,1]$ の上の一様分布 $P$ では１点 $x$ $(0 < x \leq 1)$ 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 $y$ 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 $0 < x \leq 1$ 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：

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

このようにできれば，集合関数としてではなく，普通の関数のように $x \in \Omega$ に対して定義すれば十分になる．ところが，$\displaystyle\bigcup_{x\in(0,1]} \{x\} = (0,1]$ 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 $\displaystyle\bigcup_{x\in(0,1]} \{x\}$ 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** (countable) 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．

この例から，一般には $\Omega$ 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 $(\mathcal{F} = 2^{\Omega})$ と，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．一様分布の場合は，もし $P(A + x \bmod 1) = P(A)$ を全ての $A$ に対して要求し，かつ，$P((a,b]) = b - a$ が全ての $0 \leq a \leq b \leq 1$ に対して成り立ち，かつ，選択公理が成り立つならば，$\mathcal{F} = 2^{\Omega}$ とはできない（Lebesgue 非可測集合の存在）ことが証明できる．

$G = (0,1]$ の中の有理数の全体 $H$ は「加法」に関して $G$ の部分群をなす．$G/H$ の各剰余類から代表元を一つずつ取り出して固定して，その全体を $A$ とする．$G = \displaystyle\sum_{r\in H}(A + r \bmod 1)$．このとき $P(A) = 0$ としても $P(A) > 0$ としてもそれぞれ矛盾する．即ち $P$ の定義域を $A$ まで拡張できない．

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 $\sigma$-加法族と呼び，確率を$\sigma$-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．$\sigma$-加法族の元を可測集合とも呼ぶ．幸い，実数上の Lebesgue 測度の場合は考察の対象

になる「たいていの」集合が Lebesgue 可測集合 ($\sigma$-加法族の元) になる．例えば，開集合，閉集合，これらの可算個の和や積．ここでは

たいていの集合は可測集合である．

と信じて深入りしないことにする．厳密にやるには避けられない．進んだ話題では必須．抽象的なのが難．

以上の公理はに測度論（Lebesgue 積分論）立脚した定義．$\sigma$-加法族や測度という言葉はそこから来た．特に，確率が「Lebesgue 積分」で定義されていれば，自動的に確率論の公理を満たす．さいころなどの場合は場合の数を数える素朴な方法で正しい．

一般に $\rho$（ロー）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，

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

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．$\rho$ をこの確率の **密度** と呼ぶ．全体集合 $\Omega$ は実数全体や，区間 $(0,1]$ など実数の部分集合が考えうる．

例えば，区間 $(0,1]$ 上の一様分布は，$\mathrm{P}(\,A\,)$ が集合 $A$ の長さに等しいと定義したが，これは積分

$$\mathrm{P}(\,A\,) = \int_A 1\,dx\,, \quad A \subset (0,1],$$

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．

公理主義＝上記を満たす対象ならば何でも確率空間と呼ぶ．確率の公理化＝確率論を数学の一分野とするということ（数学的定式化）.「確率とは何か」という哲学的問題の回避.

理想化, 普遍性　非負，$\sigma$ 加法性，規格化，という「当然のこと」しか要求しない．いろんな対象が確率論の研究対象になりうる．

厳密化　無意味な混乱（素朴に考えたとき陥りやすい間違い §3）を避けられる．

精密化　一旦公理にあてはめれば（定義を満たすことを確認すれば），教科書の種々の定理（公式）を利用して詳しい結論を得ることができる．

哲学的問題を解決しないですむ問題を扱う→確率論の急速な発展（確率過程，確率微分方程式，無限次元解析，等）.

確率論的な結果が精密に成り立つ自然現象（自然法則）の発見→数学としての確率論の根拠．例．Brown 運動 (R. Brown 1827, A. Einstein 1905) →きちんとやるには準備が必要．社会活動は複雑すぎて確率論の当てはめの成功失敗が議論しにくい．

例．物理では統計力学と呼ばれる分野．粒子がたくさんあるときの法則が，確率論で記述できる.（平衡系）統計力学で導入する確率の基本．$r$ 個の粒子を $N$ 箇所の「指定設置場所」に置く場合の数．

**古典粒子**　1 個当たり $N$ 箇所のどこにでも置ける．$r$ 個あって，1 箇所に何個でも重ねておけるならば，$N^r$ 通り．

**Bose 粒子**　1 箇所に何個でも重ねておけるが，どの粒子をどこに置いたかは意味がなくて，各場所に何個あるかという配置の場合の数を数えるべきならば，重複組み合わせ ${}_N H_r = {}_{N+r-1}C_r$ 通り．

**Fermi 粒子**　どの粒子をどこに置いたかは意味がなくて，各場所に何個あるかという配置の場合の数を数えるのだが，1 箇所に 1 個しか重ねておけないならば，${}_N C_r$ 通り ($N \geq r$).

古典的なボールのイメージで考えると 1 番目の数え方が正しそうだが，いろいろな実験によって，原子より細かい粒子については 2 番目の数え方が正しいか 3 番目の数え方が正しいかどちらかであることが分かっている．

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．実験によって具体的な確率が簡単な形に予測できて，それで全ての現象が説明できる場合は問題ない．統計力学などはそうやって成功した．

社会活動などに応用するときは個々の問題毎に $(\Omega, \mathcal{F}, P)$ それぞれが何であるかを自分で判断しないといけない．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．

統計学は判断（現実との関わり）があるので数学になりきれない部分がある．例えば，現実に起きることはただ一つなのに確率的に予想するとはどういうことか？保険会社にとっては事故の起きる確率かける契約者数が事故の起きる平均値．大数の法則と呼ばれる定理が成り立つ状況ならば，契約者の数が十分大きければ実際に起きる事故の数と平均値の比は必ず１に近づく．つまり，保険会社の経営にとっては確率は現実的な意味がある．しかし，自分の乗る自動車の事故で死ぬ確率が 0.01% と言われたとして，それがどんな意味があるのか？１回死んだらおしまいなのに平均値に意味があるだろうか？→確率論と統計学を分けて考えるもう一つの理由．

## 3　確率論と現実の関わり

03

確率論を知っていると知っていないで現実の生活の知恵が変わるか？確率論の知識をいくらか先取りして，ある実体験を話したい．

銀行のキャッシュコーナー，電話ボックス，飛行機のチェックインカウンター，など，窓口（またはボックス）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」が数年前に新聞などでも話題になった．昔は，たいていの場合，スーパーのレジのように全てボックスや窓口の前に別々に並んでいた.「一列待ち」が話題になってからしばらくは「一列待ち」がいろいろなところで試みられたと思う．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ていないように思う.（話題になったとき「欧米ではそうなっている」ということだったが，たしかに，アメリカの飛行場のチェックインカウンターはロープをひっぱて「一列待ち」にしている．日本のはどうなっているのだろう.）

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，４年ほど前のことだが,「一列待ち」が話題になってしばらく後,「一列待ち」に変えた．ところがＹ駅は一旦開始した「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

(i) 平均時間は変わらない，
(ii) 列が長く見えてＹ駅は混んでいると思われてしまう，
(iii) 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意味がない」という主張にみえるが，これは適切な理由とは言えない．今回の講義では

**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．

(i) 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．

(ii) 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．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．確率変数とは $\Omega$ 上の関数であるということ（正確には後述）．$\Omega$ は制御不能な撹乱要因あるいはその結果として起こりうるかも知れない sample の全体だから，その上の関数というのは，値が制御不能な要因によってばらつきうる量，という気持ちである．客によって処理時間がばらつくことをこのようにとらえよう，ということである．確率変数 $X_1, X_2, \cdots$, が独立とは，$\mathrm{P}(\ X_1 \in A_1,\ X_2 \in A_2,\ \cdots\ ) = \prod_i \mathrm{P}(\ X_i \in A_i\ )$ ということ（後述）．

(iii) 「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．列のでこぼこは考えない．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．この問題についてはここでは扱わない．

さて，一人の客の処理をするのに平均 $\tau$（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 $\tau$ もよく使われる）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，

**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 $\mathrm{P}(\ T > t_0\ )$ 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．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」$\mathrm{P}(\ T > t_0\ )$ 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．常識的には，分布が広がっているほど $\mathrm{P}(\ T > t_0\ )$ 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．

**注**. 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 $\mathrm{P}(\ T > t_0\ )$ 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 $\mathrm{P}(\ T > t_0\ )$ 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 $\mathrm{P}(\ T > t_0\ )$ も計算可能になる．

「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．$\sigma$ 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.

(i) $\{S_i\}$ が独立ならば，$\mathrm{V}[\,S_1 + S_2 + \cdots + S_n\,] = \mathrm{V}[\,S_1\,] + \mathrm{V}[\,S_2\,] + \cdots + \mathrm{V}[\,S_n\,]$.
(ii) $a$ 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

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

$\mathrm{V}[\,T^{(2)}\,]$ について証明しておこう.

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

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

$M > 1$ だと $\mathrm{V}[\,T^{(1)}\,] > \mathrm{V}[\,T^{(2)}\,]$ となるから，並列待ちが分散の大きい場合，一列待ちが小さい場合に対応する．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．**

ＪＲの職員は期待値が変わらないという数学的結論を知っていたという点で優れている．確率論を勉強しなくても，分かる部分がある．しかし，問題が期待値ではなく分散，あるいは，より正確には，平均値からの大きなずれの確率であることの認識がなかった．確率論を知るということは問題をそこまで煮詰められるということである.

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，を指摘したい．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．今回は確率論の結果を先取りして話をしたが，これからしばらく根拠となるべき確率論の基礎を講義していく.

## 第 1 部
# 確率論からの準備

05–08

数学に深入りしないといっても，せっかくあるのだから使わない手はない.

**定義 1 (Kolmogorov)** 確率空間とは全体集合 $\Omega$, $\Omega$ の部分集合よりなる$\sigma$-加法族 $\mathcal{F} \subset 2^{\Omega}$, $(\Omega, \mathcal{F})$ 上の確率測度 $P$ の組 $(\Omega, \mathcal{F}, P)$ を言う．

ここから確率論を展開するのだが，この先に関して，（確率過程の前までについて）重要な概念として以下のものがある．

(i) 確率変数，確率変数の分布，期待値（分散，モーメント）
(ii) 独立性（確率変数の独立性，事象の独立性，Borel-Cantelli の定理）
(iii) 収束：分布の収束（特性関数，弱収束），確率変数の収束（概収束，確率収束，平均収束，法則収束）
(iv) 大数の法則，中心極限定理（期待値，分散，正規分布の重要性），小数の法則（Poisson 分布の意味）

以下，順にこれらの内容を概観していく．

# 4 確率変数と期待値

## 4.1 確率変数と分布

**確率変数 (random variable)** とは関数のことである．確率空間 $(\Omega, \mathcal{F}, P)$ が与えられたとき，$\Omega$ の上の関数を確率変数と呼ぶ．正確には「確率の計算できる」関数だが，通常考えるような関数は全て許される．具体的には，（区分的に）連続な関数などは確率変数である．この講義では実数値関数のみを考える：$X:\ \Omega \to \mathbf{R}$.

確率変数の定義の「気持ち」は，値に散らばり（不規則性）を生じる量である．制御不能な要因 $\Omega$ によって値が変わりうる量，という意味で $\Omega$ 上の関数は気持ちの上では全て確率変数の資格がある．関数をわざわざ確率変数と呼ぶのは，集合を事象，積分を確率，と呼ぶのと同様に，これらの基本的な数学がそのまま確率論の「気持ち」を表しているからである．

確率変数の各々の値が実現する可能性を確率変数の**分布**と呼ぶ．確率変数 $X$ は関数だから，実数の部分集合 $A \subset \mathbf{R}$ に対して，$X$ の値が $A$ に含まれる確率が考えられる．つまり，事象 $\{\omega \in \Omega \mid X(\omega) \in A\}$ の確率 $\mathrm{P}(\ \{\omega \in \Omega \mid X(\omega) \in A\}\ )$ である．これを $Q(A)$ とおくと，$Q$ は実数の集合 $A$ を与えるごとに値が決まる．$Q(A) = \mathrm{P}(\ \{\omega \in \Omega \mid X(\omega) \in A\}\ )$ を，**$X$ が $A$ に入る確率** といい，（実数空間 $\mathbf{R}$ 上の）確率 $Q = P \circ X^{-1}$ 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，上記の $Q$ 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 (formulate)（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．

確率変数に関する以上の説明は（気持ちの上では正しいが）数学的には不正確である．前回注意したように，必ずしも常に勝手な集合の確率が全て存在するわけではない．確率を定義できる集合，つまり事象，の集まりとして $\sigma$-加法族を定義した．同様に，勝手な関数 $X$ を全て考慮の対象にできない．$X$ が確率変数である，というときの「気持ち」には，確率が計算できる（$X$ の分布 $Q$ が存在する）ことも期待されている．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．確率変数の正しい定義を見ておこう．

実数値関数が取る値は実数なので，確率変数の分布は実数上の確率（$\Omega = \mathbf{R}$ のときの確率）になる．実数の場合，区間の確率，即ち，例えば $0.1 \le X < 0.3$ となる確率，が当然ほしいから，どんな $\sigma$-加法族でもいいのではなく，区間 $(a, b]$ は全て $(a < b)$ 含まれるような $\sigma$-加法族 $\mathcal{F}$ が必要である．区間を全て含む最小の $\sigma$-加法族を 1 次元 Borel 集合族と呼び，記号で（$\mathcal{F}$ の代わりに）$\mathcal{B}$ と書く．

$\sigma$-加法族の定義から，$\mathcal{B}$ は開区間 $(a, b)$ や閉区間 $[a, b]$ も含むことが証明できるので，片側開区間 $(a, b]$ を用いて $\mathcal{B}$ を定義したのは単なる趣味．$\sigma$(開集合族) $=$ $\sigma$(開区間族) $=$ $\sigma$(閉区間族) $=$ $\sigma$(片側開区間族 $) = \sigma(\{(-\infty, r) \mid r \in \mathbf{Q}\}) = \mathcal{B}$ $\sigma$(n次元直方体族) $\stackrel{\mathrm{def}}{=} \mathcal{B}_n$ を $n$ 次元 Borel 集合族．1 次元と同様．

$(\Omega, \mathcal{F}, P)$ を確率空間とする．$X:\ \Omega \to \mathbf{R}$ が**確率変数**であるとは

$$A \in \mathcal{B} \Rightarrow X^{-1}(A) \in \mathcal{F},$$

を満たすことと定義する．ここで，$X^{-1}$ は $X$ の逆関数を表す．即ち，$X^{-1}(A)$ とは $X(\omega) \in A$ となるような点（見本）$\omega \in \Omega$ の集合である．

$A \in \mathcal{B}$ に対して $Q(A) = \mathrm{P}(\ X^{-1}(A)\ ) = \mathrm{P}(\ \{\omega \in \Omega \mid X(\omega) \in A\}\ )$ で $\mathcal{B}$ 上の関数 $Q = P \circ X^{-1}$ を定義すれば，$Q$ は $(\mathbf{R}, \mathcal{B})$ 上の確率測度になる．$Q$ を $X$ の分布（法則）という．また，$X$ は分布（法則）$Q$ に従う，ともいう．

なお，以後上のような量を

(1)　$\mathrm{P}(\ X \in A\ ) \overset{\mathrm{def}}{=} \mathrm{P}(\ \{\omega \in \Omega \mid X(\omega) \in A\}\ )$

と略記する．（この略記は $X$ が $A$ に入る確率，あるいは $A$ が実現する確率，という素朴な語法に近いので私は好ましいと思う．）

通常考えるような実数の集合は $\mathcal{B}$ に含まれるし，通常考えるような関数は，確率変数の定義を満たす．

**注**. $P$ は $X$ の定義域 $(\Omega, \mathcal{F})$ の上の確率測度，$Q = P \circ X^{-1}$ は $X$ の値域 $(\mathbf{R}, \mathcal{B})$ の上の確率測度である．$X$ が可測関数であるという定義から，$P$ が $X$ を通して $Q$ を誘導している．

$\Omega$ 上の実数値関数 $g$ が階段関数であるとは以下を満たすこと．$\exists N \in \mathbf{Z}_+,\ \exists a_i \in \mathbf{R},\ i = 1, 2, \cdots, N,\ \exists A_i \in \mathcal{F},\ i = 1, 2, \cdots, N,\ A_k \cap A_n = \phi,\ k \neq n;\ g = \displaystyle\sum_{i=1}^{N} a_i \chi_{A_i}$.

**命題 1** すぐ導かれる性質．$f,\ g,\ f_n,\ n = 1, 2, 3, \cdots : \Omega \to \mathbf{R};\ \mathcal{F}$ 可測.

*(i)* 階段関数は確率変数．例 $\chi_A,\ A \in \mathcal{B}$.

*(ii)* $\mathbf{R^n}$ 上の実数値連続関数は $\mathcal{B}_n$ 確率変数である．

*(iii)* $h :\ \mathbf{R^n} \to \mathbf{R},\ \mathcal{B}_n$ 確率変数とすると，$g(f_1(\omega), \cdots, f_n(\omega))$ も $\mathcal{F}$ 可測．例：$f \pm g,\ f \cdot g,\ f \cdot \chi_A,\ f_\pm \overset{\mathrm{def}}{=} \max\{\pm f, 0\},\ |f|$.

*(iv)* $\inf_n f_n,\ \sup_n f_n,\ \liminf_n f_n,\ \limsup_n f_n$, は $\Omega$ の各点で有限値ならば $\mathcal{F}$ 確率変数．$\{\omega \in \Omega \mid \exists \lim_n f_n(\omega)\} \in \mathcal{F}$.

*(v)* さらに $f :\ \Omega \to \mathbf{R}_+$ ならば非負階段関数で下から各点近似できる．即ち

$$\exists g_n\ n = 1, 2, 3, \cdots,\ \text{階段関数};\ \lim_{n\to\infty} g_n(\omega) = f(\omega),\ 0 \leq g_1(\omega) \leq g_2(\omega) \leq \cdots,\ \omega \in \Omega\,.$$

*(vi)* 一般に $g :\ \mathbf{R^n} \to \mathbf{R},\ \mathcal{B}_n$ 確率変数，$\mathbf{X} :\ \Omega \to \mathbf{R^n},\ n$ 次元確率変数とすると $g \circ \mathbf{X}(\omega) = g(\mathbf{X}(\omega))$ は $\Omega$ 上の確率変数となる．$Q$ を $\mathbf{X}$ の分布とすると $\mathrm{E}[\ g \circ \mathbf{X}\ ] = \int_{\mathbf{R^n}} g(x)\, Q(d^n x)$.

$X_i,\ i = 1, \cdots, n$：確率変数のとき $\mathbf{X} \overset{\mathrm{def}}{=} (X_1, \cdots, X_n) :\ \Omega \to \mathbf{R^n}$ を $n$ 次元確率変数という．

**命題 2** $(\Omega, \mathcal{F}, P)$：確率空間，$\mathbf{X}$：$n$ 次元確率変数．$A \in \mathcal{B}_n$ に対して

$$P(\mathbf{X} \in A) = Prob(\mathbf{X} \in A) = P(\mathbf{X}^{-1}(A)) \overset{\mathrm{def}}{=} Q(A)$$

で $\mathcal{B}_n$ 上の関数 $Q = P \circ \mathbf{X}^{-1}$ を定義すれば $Q$ は確率になる．

$(\mathbf{R^n}, \mathcal{B}_n)$ 上の確率 $Q$ を $\mathbf{X}$ の分布あるいは $X_1, \cdots, X_n$ の結合分布という．

## 4.2　期待値，分散，モーメント

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

確率空間及び確率変数の定義の一つの重要な数学的意味は，この定義の下で**確率 $P$ が測度（長さの拡張概念）となり**，期待値 $\mathrm{E}[\ X\ ]$ が積分として定義できることである．

**数学的には確率論は測度論（積分論）である．**

確率 $P$ が密度 $\rho$ を持っているとき，即ち，任意の $A \in \mathcal{F}$ に対して $\mathrm{P}(\,A\,) = \int_A \rho(x)dx$ と書けるときには,「普通の」積分で期待値が定義できる：

$$
\begin{aligned}
(2) \qquad \mathrm{E}[\,X\,] &= \int_\Omega X(\omega)\,\rho(\omega)\,d\omega\,, \\
(3) \qquad \mathrm{E}[\,X; A\,] &= \int_A X(\omega)\,\rho(\omega)\,d\omega = \mathrm{E}[\,X\,\chi_A\,]\,.
\end{aligned}
$$

式 (3) は，集合 $A$ の上での期待値，の定義である．$\chi_A$ は集合 $A \in \mathcal{F}$ の定義関数（$\omega \in A$ のとき $\chi_A(\omega) = 1$, $\omega \notin A$ のとき $\chi_A(\omega) = 0$ となる関数）．例えば，$\chi_\Omega$ は恒等的に 1 になる定数関数であり，$\mathrm{E}[\,\chi_\Omega\,] = P(\Omega) = 1$.

積分の線型性から，**期待値の線型性**，即ち，定数 $a$, $b$ と確率変数 $X, Y$ に対して

$$
(4) \qquad \mathrm{E}[\,aX + bY\,] = a\mathrm{E}[\,X\,] + b\mathrm{E}[\,Y\,]
$$

が成り立つ.

以上は積分論を既知として期待値を説明したのであるが，確率 $P$ が与えられたときに，積分即ち期待値 $\mathrm{E}[\,X\,]$ を（密度に頼らずに）定義する現代的な積分論の粗筋を述べておく.（証明はしない）.

記号として，(2), (3) を一般化してそれぞれ

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

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

と書く．複数の確率測度を考えるときなど $\mathrm{E}[\,X\,]$ を $\mathrm{E}^{\mathrm{P}}[\mathrm{X}]$ と書く.

先ず，$X$ が集合 $A \in \mathcal{F}$ の定義関数 $\chi_A$ （$\omega \in A$ のとき $\chi_A(\omega) = 1$，$\omega \notin A$ のとき $\chi_A(\omega) = 0$ となる関数)，のとき（ちょうど $A$ の上でだけ 1 になるから)，$\mathrm{E}[\,\chi_A\,] = \displaystyle\int_A dP(\omega) = \mathrm{P}(\,A\,)$ で定義する.（$P$ は確率だから $\mathrm{P}(\,A\,)$ は存在する.）

次に，$X$ が階段関数，即ち，複数の可測集合の定義関数の和 $X = \displaystyle\sum_{i=1}^N a_i \chi_{A_i}$ （$a_i$ は定数，$A_i \in \mathcal{F}$）のとき（定義関数の重ね合わせだから)，$\mathrm{E}[\,X\,] = \displaystyle\sum_{i=1}^N a_i \int_{A_i} dP(\omega) = \sum_{i=1}^N a_i \mathrm{P}(\,A_i\,)$ で定義する.

最後に一般の確率変数 $X$ の場合，階段関数の列 $g_n$, $n \in \mathbf{N}$, であって，$X$ に収束する ($\displaystyle\lim_{n\to\infty} g_n(\omega) = X(\omega)$) 増大する ($g_1(\omega) \le g_2(\omega) \le \cdots$) 列が取れるので,

$$
\mathrm{E}[\,X\,] = \lim_{n\to\infty} \mathrm{E}[\,g_n\,]
$$

で定義する．確率変数の定義から，このような階段関数（$A_i \in \mathcal{F}$）の近似列がとれ，極限が近似列の取り方によらないことが保証される.

確率 $P$ が密度 $\rho$ を持っているときは,「普通の」積分で期待値が書ける：

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

$A \in \mathcal{B}$ に対して $Q(A) = \mathrm{P}(\,X \in A\,) = \mathrm{P}(\,X^{-1}(A)\,)$ 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 $\rho_Q$ のとき,

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

が定義と一般論から導かれる．一般には（$Q$ が密度を持たない場合も） $\mathrm{E}[\,X\,] = \int_{\mathbf{R}} x\,Q(dx)$ が導かれる．証明しないが，積分変数変換の公式から導かれる.

気持ちとしては,

$$
\mathrm{E}[\,X\,] = \int X\,dP = \lim \sum_i a_i \mathrm{P}(\,X \in [a_i, a_i + \delta)\,) = \lim \sum_i a_i Q([a_i, a_i + \delta)) = \int xQ(dx) = \mathrm{E}^{\mathrm{Q}}[\omega]\,.
$$

$Q$ 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 $\mathrm{P}(\ X = n\ ) = q_n$ $(\sum_n q_n = 1)$ とすると，同様に次式を得る：

$$\text{(8)} \quad \mathrm{E}[\ X\ ] = \int_{\mathbf{Z}} n dQ(n) = \sum_{n=-\infty}^{\infty} n\, q_n\,,$$

測度論ではこれらをまとめて $\mathrm{E}[\ X\ ] = \int_\Omega \omega\, Q(d\omega)$ と書く．

また，多次元積分を考えることで，多次元空間上の分布も考えることができる．例えば，それぞれ $(0,1]$ の上の一様分布に従う独立な 2 つの確率変数 $X$ と $Y$ があったとき，$X+Y \le 1$ となる確率は

$$\mathrm{P}(\ X+Y \le 1\ ) = \int\!\!\int_{x+y\le 1,\ 0<x,y\le 1} 1\, dx\, dy = \int_0^1 (\int_0^{1-y} dx) dy = \int_0^1 (1-y)\, dy = \frac{1}{2}$$

などと積分で計算できる．それぞれ密度 $\rho_X$ および $\rho_Y$ で定義される分布に従う，互いに独立な確率変数 $X$ と $Y$ の結合分布は密度 $\rho(x,y)$ が，積 $\rho(x,y) = \rho_X(x)\,\rho_Y(y)$ で書ける分布に従う．一様分布の場合は $\rho(x) = 1$ なので結合分布の密度も 1 である．

これらのことは，証明するのに準備がいるが，結論は全て正しいことが知られている．

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 (4) から直ちに次の性質が分かる．

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

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．最後の性質は，分散には加法性が一般にはないことを表す．分散が加法性を持つのは，$X$ と $Y$ が独立な場合である．

期待値と分散は $\mathrm{E}[\ X^n\ ]$ の $n = 1, 2$ によって決まることが分かる．一般に $n$ を自然数とするとき $\mathrm{E}[\ X^n\ ]$ を $X$ の $n$ 次のモーメントと呼ぶ．実用上の問題で，データを増やして観測を続けると，多くの場合 $n$ の小さい方から順に究極の値に早く近づいていく（値が安定する）．$n$ が小さいモーメントほど実用上は重要な場合が多い．このことを数学的な視点でとらえたときの根拠として，後述の大数の法則や中心極限定理がある．モーメントは全て $X$ の分布 $Q$ を用いて $\mathrm{E}[\ X^n\ ] = \int x^n\, Q(dx)$ と書ける．

# 5 独立性

## 5.1 2 つの事象の独立性

[1]

[1] 19960614 講義後追加項目．条件付き確率も先ず 2 つでやるべき．他の問題は全て，$n$ そして $\infty$ 個への拡張として Appendix に回すべき．

## 5.2 事象の独立性

有限個の事象 $A_k \in \mathcal{F},\ k=1,2,\cdots,n,$ が独立とは

$$\mathrm{P}(\ A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_m}\ ) = \mathrm{P}(\ A_{i_1}\ )\,\mathrm{P}(\ A_{i_2}\ ) \cdots \mathrm{P}(\ A_{i_m}\ )$$

が任意の部分列 $A_{i_k},\ k=1,2,\cdots,m,$ に対して成り立つことと定義する．

$n=2$ の場合を考えれば，これは素朴な意味の独立ということである．

事象の無限列 $A_k,\ k=1,2,\cdots,$ が独立とは任意の $n$ に対して有限列 $A_k,\ k=1,2,\cdots,n,$ が独立になること．

事象の無限列に関して次の定理が基本である．

**定理 3 (Borel-Cantelli の定理)** 事象の無限列 $A_k \in \mathcal{F},\ k=1,2,\cdots,$ が与えられたとき．

*(i)* $\displaystyle\sum_{n=1}^{\infty} \mathrm{P}(\ A_n\ ) < \infty$ ならば $\mathrm{P}(\ \displaystyle\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\ ) = 0,\ \mathrm{P}(\ \displaystyle\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} {A_k}^c\ ) = 1,$ 即ち無限個の事象のうち有限個しか起こらない確率が 1 である．

*(ii)* もし $A_k \in \mathcal{F},\ k=\mathbf{N},$ が**独立**であって $\displaystyle\sum_{n=1}^{\infty} \mathrm{P}(\ A_n\ ) = \infty$ ならば $\mathrm{P}(\ \displaystyle\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\ ) = 1,\ \mathrm{P}(\ \displaystyle\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} {A_k}^c\ ) = 0.$ 即ち無限個の事象が起こる確率が 1 である．

**証明**. [楠岡, p.9] 確率の連続性に持ち込む．前後半とも 0 になるほうを証明し，もう一つの式は deMorgan を使う．前半は $\mathrm{P}(\ \displaystyle\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k\ ) \le \mathrm{P}(\ \displaystyle\bigcup_{k=n}^{\infty} A_k\ ) \le \displaystyle\sum_{k\ge n} \mathrm{P}(\ A_k\ ),$ 後半は

$$\mathrm{P}(\ \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} {A_k}^c\ ) = \lim_{n\to\infty}\lim_{m\to\infty} \mathrm{P}(\ \bigcap_{k=n}^{m} {A_k}^c\ ),$$

で独立性を使って積集合の確率を確率の積にしたのち，仮定からそれが 0 になることをいう．

あることが起こる確率が 1 であるとき,「殆ど確実に起こる」と言う.（現実には「必ず起こる」という意味である！）定理 3 から，$\displaystyle\sum_{n=1}^{\infty} \mathrm{P}(\ A_n\ ) < \infty$ ならば殆ど確実に無限個の事象のうち有限個しか起こらないし，$\displaystyle\sum_{n=1}^{\infty} \mathrm{P}(\ A_n\ ) = \infty$ で独立な事象列ならば殆ど確実に無限個の事象が起こる．

## 5.3 確率変数の独立性

$n$ 個の確率変数 $X_i,\ i=1,\cdots,n,$ が**独立**とは,

(9) $$\mathrm{P}(\ X_{i_1} \in A_{i_1}, \cdots, X_{i_m} \in A_{i_m}\ ) = \prod_{i=1}^{m} \mathrm{P}(\ X_{i_k} \in A_{i_k}\ ), \qquad A_i \in \mathcal{B},\ i=1,2,\cdots,n,$$

が，$1,2,\cdots,n$ の任意の部分列 $i_1, i_2, \cdots, i_m$ $(m=2,3,\cdots,n)$ に対して成り立つことと定義する．

積分変数変換の公式を用いると，分布の言葉でいうことができて，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．

証明は再びしないが，このとき $A \in \mathcal{B}_n$ に対して

(10) $$\mathrm{P}(\ (X_1,\cdots,X_n) \in A\ ) = Q(A) = \int_{(x_1,x_2,\cdots,x_n)\in A} Q_1(dx_1)\cdots Q_n(dx_n)$$

と書ける．$Q_i$ が積分で書けているとき，即ち $Q_k(A)=\int_{x\in A}\rho_k(x)\,dx$ などと書けているときは

(11) $\mathrm{P}(\,(X_1,\cdots,X_n)\in A\,)=Q(A)=\int_{(x_1,x_2,\cdots,x_n)\in A}\rho_1(x_1)\cdots\rho_n(x_n)\,dx_1\cdots dx_n\,.$

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**．

**命題 4** 確率変数の簡単な性質その 2．

*(i)* $X_i,\ i=1,\cdots$ が独立ならば定数たち $c_i$ に対して確率変数 $X_i-c_i,\ i=1,\cdots$ も独立．
*(ii)* $\phi_{i,j}:\ \mathbf{R}^{\mathbf{k_i}}\to\mathbf{R},\ j=1,\cdots,l_i$ を $\mathcal{B}_{k_i}$ 可測関数，$\mathbf{X_i}:\ \Omega\to\mathbf{R}^{\mathbf{k_i}},\ i=1,\cdots,n$ を独立な $k_i$ 次元確率変数とすると，$l_i$ 次元確率変数 $\mathbf{Y_i}:\ \Omega\to\mathbf{R}^{\mathbf{k_i}},\ i=1,\cdots,n$ も独立．
*(iii)* $X_i,\ i=1,\cdots$ が独立ならば，それらを $k_1,k_2,\cdots$ ずつ一まとめにして作った $k_i$ 次元確率変数達も独立．
*(iv)* $X_k,\ k=1,\cdots,n$ 独立．各 $X_k$ が可積分ならば $\prod_{k=1}^{n}X_k$ も可積分で $\mathrm{E}[\,\prod_{k=1}^{n}X_k\,]=\prod_{k=1}^{n}\mathrm{E}[\,X_k\,]$．（$X,Y$ が独立だと結合分布 $Q=Q_X\times Q_y$．*Fubini*（と積分の線型性）から.）

**命題 5** *(i)* 任意の（可算または有限）独立事象列 $A_k,\ k=1,2,\cdots,$ に対して $\phi,\ A_k,\ k=1,2,\cdots,\Omega$ は独立である．
*(ii)* 事象 $A,\ B$ が独立ならば $A,\ B^c$ も独立である．

次の意味で確率変数の独立性は事象の独立性の拡張概念になっていることが分かっている．

**定理 6** $X_k=\chi_{A_k},\ k=1,2,\cdots,$ の場合には $\{X_k\}$ の独立性と $\{A_k\}$ の独立性は同値である．

**証明**. $X=\chi_A, Y=\chi_B,$ について証明してみる．値域は $\{0,1\}$ だから (9) で $A_k$ が $0,1$ を含むか含まないかだけが問題．$A_i$ が $\phi,\mathbf{R},Z=\{0\}\in\mathcal{B},Z^c=\{1\},$ の場合についてのみ証明すればよい．$\{\omega\in\Omega\mid X(\omega)\in Z\}=A,$ $\{\omega\in\Omega\mid X(\omega)\in Z^c\}=A^c,$ などより，$X$ と $Y$ の独立性は $\sigma(A)\stackrel{\mathrm{def}}{=}\{\phi,A,A^c,\Omega\}$ と $\sigma(B)\stackrel{\mathrm{def}}{=}\{\phi,B,B^c,\Omega\}$ の中から各一つずつ選んだときの事象の独立性が全て成り立つことと同値.（このことを$\sigma$-加法族 $\sigma(A)$ と $\sigma(B)$ の独立性と呼ぶ．後述.）命題 5 よりこれは $A,\ B$ の独立性と同値.

期待値との関連では，次のことが言える．

**命題 7** $X_i,\ i=1,\cdots,n,$ が独立で，$f_i,\ i=1,\cdots,n,$ が（$\mathcal{B}$ 可測実数値）関数のとき，

(12) $\mathrm{E}[\,f_1(X_1)\,f_2(X_2)\,\cdots\,f_n(X_n)\,]=\mathrm{E}[\,f_1(X_1)\,]\,\mathrm{E}[\,f_2(X_2)\,]\,\cdots\,\mathrm{E}[\,f_n(X_n)\,]$

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．**

（$f\circ X$ を改めて $X$ と置けばよいのだから $f$ を明記する必要はない.）
特に，このことから，$X$ と $Y$ が独立ならば，

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

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．**

$\sigma_x=\sqrt{\mathrm{V}[\,X\,]}$ を標準偏差と呼ぶ.

統計学では相関という概念もよく用いる．どれくらい独立か（独立でないか）を表す目安．分散の親戚.

**定義 2** • 共分散 $C(X,Y) = \mathrm{E}[\,(X - \mathrm{E}[\,X\,])(Y - \mathrm{E}[\,Y\,])\,]$.

• $\mathrm{V}[\,X\,]\mathrm{V}[\,Y\,] \neq 0$ のとき，相関係数 $r(X,Y) = \mathrm{E}[\,(X - \mathrm{E}[\,X\,])(Y - \mathrm{E}[\,Y\,])\,]/\sqrt{\mathrm{V}[\,X\,]\,\mathrm{V}[\,Y\,]}$.

**命題 8** *(i)* $X$, $Y$ が独立ならば $C(X,Y) = 0$.
*(ii)* $\mathrm{V}[\,X\,]\mathrm{V}[\,Y\,] \neq 0$ のとき $-1 \leq r(X,Y) \leq 1$.

**証明**. 前半は明らか．後半は $f(t) = \mathrm{E}[\,(t(X - m_x) - (Y - m_y))^2\,] \geq 0$ の判別式.

### 5.4 情報 ($\sigma$-加法族) の独立性

確率変数の独立性のもう一つの表現．$\sigma$-加法族 $\mathcal{F}$ はたいていの場合殆ど全ての集合と思って良いと言ったが，実際は情報の精密さという現実的な意味を持つ概念であり，重要である．特に確率過程をやるためには決定的に重要な意味がある．

可測空間 $(\Omega, \mathcal{F})$ があるとき，$\sigma$-加法族の（可算または有限）列 $\mathcal{F}_k \subset \mathcal{F}$, $k = 1, 2, \cdots$, が独立とは，任意の事象列 $A_k \in \mathcal{F}_k$, $k = 1, 2, \cdots$, が独立なことと定義する．

**命題 9** $A \neq \emptyset$, $A \neq \Omega$ ならば，事象 $A (\in \mathcal{F})$ を含む最小の $\sigma$-加法族は $\sigma(A) \stackrel{\mathrm{def}}{=} \{\phi, A, A^c, \Omega\}$ である．

**定理 10** 事象 $A$, $B$ が独立であることと，$\sigma(A)$, $\sigma(B)$ が独立であることは同値である． $\sigma(A)$ は 命題 *9* で定義した.

この定理は，容易に想像できるように任意可算または有限の事象列に拡張できる．定理 10 は 命題 5 を用いて証明できる．この定理の意味で，$\sigma$-加法族の独立性は事象の独立性の定義の拡張になっている．

さらに，次の意味で確率変数の独立性の定義の拡張になっていることも分かっている [数学辞典, 47C].

**命題 11** 確率変数 $X$ に対して，集合族 $\sigma(X) \stackrel{\mathrm{def}}{=} \{X^{-1}(A) \mid A \in \mathcal{F}\}$ は $\sigma$-加法族になる．これを $X$ が生成する $\sigma$-加法族と呼ぶ.

**定理 12** （有限または可算）確率変数列 $X_k$, $k = 1, 2, \cdots$, が独立であることと$\sigma$-加法族列 $\sigma(X_k)$, $k = 1, 2, \cdots$, が独立なことは同値である.

この定理は，$X_k = \chi_{A_k}$, $k = 1, 2, \cdots$, の場合には定理 6 と 定理 10 に帰着される．

## 6 分布と特性関数

### 6.1 分布

$(\mathbf{R}, \mathcal{B})$ 上の確率測度を確率分布という．(実数値) 確率変数の分布は $\mathbf{R}$ 上の確率測度，即ち，確率分布である．$\Omega = \mathbf{R}$ のときは元 (sample) を $\omega$ と書かずに $x$ と書くのが普通．$Q$ を確率分布とするとき $m = \int_{\mathbf{R}} xQ(dx)$ を $Q$ の平均，$v = \int_{\mathbf{R}} (x - m)^2 Q(dx)$ を分散という．$Q$ を分布に持つ確率変数の期待値と分散になっていることはすぐわかる．

確率変数とは数学的には（可測）関数のことであったが，現実への応用ではデータ（標本・サンプル）の値を表す．従って，現実への応用では確率変数の法則，即ち確率分布，が問題になる．

$X$ 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．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．

$\mathcal{B}$ は区間 $(a,b]$ を含む最小の $\sigma$-加法族であった．任意の $x \in \mathbf{R}$ に対して $x \in \mathcal{B}$ である．(なぜなら，$(x-1/n, x] \in \mathcal{B}$ と可算積に関して $\sigma$-加法族が閉じているから.）分布 $(\mathbf{R}, \mathcal{B}, P)$ に対して，$P(\{x\}) > 0$ となる $x$ の集合を $D \subset \mathbf{R}$ とおく．$D$ は高々可算集合である.（なぜなら，$P(\{x\}) > 1/n$ なる点の集合は $n$ 個以下で，それらの $n$ についての和集合に $D$ は含まれるから.）よって $P(D) = \sum_{x \in D} P(\{x\})$．$P(D) = 1$ となる分布を離散分布と呼び，$P(D) = 0$ となる分布を連続分布と呼ぶ．一般の分布は両方とも確率正で入りうるが，例示するときは分けて行う．

## 6.2 離散分布

離散分布では，各 $x \in D$ 毎に $P(\{x\})$ を与えれば可算加法性から任意の $\mathcal{B}$ の元に対して確率が決まる．従って，$P$ を $\Omega = D, \mathcal{F} = 2^D$, の上の確率測度とみなして調べれば十分である．ここでは $D = \mathbf{Z}$ またはその部分集合として確率分布の典型的な例を挙げる．個々の分布の詳しい性質は [小針] 参照.

### 6.2.1 2 項分布

$p \in (0,1]$ に対して，$\Omega = \{0, 1, 2, \cdots, n\}$, $\mathcal{F} = 2^\Omega$, $P(\{k\}) = {}_nC_k p^k (1-p)^{n-k}$,（有限加法性から任意の $A \in \mathcal{F}$ (i.e. $A \subset \Omega$) に対して $P$ が定義できたことになる.）で定義される確率測度.（$\Omega$ を $\mathbf{R}$ に埋め込められるので確率分布と呼べる）．これを $B_{N,p}$ と書くことにする．

**応用** 硬貨投げ $\Omega = \{0, 1\}$, $\mathcal{F} = 2^\Omega$, $P = P_p$; $P_p(\{1\}) = p$, $P_p(\{0\}) = 1 - p$, を $n$ 回行ったとき $k$ 回表 1 が出る確率を表す確率分布.

**注** 確率変数の言葉で言えば，$(\Omega, \mathcal{F}, P)$ 上の確率変数 $X$ であって $\mathrm{P}[\,X = 1\,] \stackrel{\mathrm{def}}{=} \mathrm{P}[\,\{\omega \in \Omega \mid X(\omega) = 1\}\,] = p$, $\mathrm{P}[\,X = 0\,] \stackrel{\mathrm{def}}{=} \mathrm{P}[\,\{\omega \in \Omega \mid X(\omega) = 0\}\,] = 1 - p$, $\mathrm{P}[\,$その他$\,] = 0$, を満たすものの分布が $P_p$.

(9) より明らかに，

**命題 13** $P_p$ を分布に持つ $n$ 個の独立な確率変数 $X_1, \ldots, X_n$ が与えられたとき，$Y_n = \sum_{k=1}^{n} X_k$ の分布が 2 項分布 $B_{n,p}$.

そういう意味で極めて基本的な分布．同じ分布に従う独立な確率変数列を i.i.d. と略記することが多い．

$A_k = \{\omega \in \Omega \mid X_k(\omega) = 1\}$ とおけば $A_k \in \mathcal{F}$ であって，$\{A_k\}$ は独立な事象である．従って独立な事象の出現数に関する分布とも言える．

**公式** [小針, pp.56–57] $B_{N,p}$ について $m = \mathrm{E}[\,X\,] = Np$, $v = \mathrm{V}[\,X\,] = Npq$.

### 6.2.2 Poisson 分布

$\lambda > 0$ に対して，$\Omega = \mathbf{Z}_+$, $\mathcal{F} = 2^\Omega$, $\mathrm{P}[\,\{k\}\,] = \exp(-\lambda)\lambda^k/k!$, で定義される確率分布.

**公式** [小針, pp.63–64] $m = \mathrm{E}[\,X\,] = \lambda$, $v = \mathrm{V}[\,X\,] = \lambda$.

#### Poisson 小数の法則

**命題 14** 2 項分布 $B_{n,p}$ に対して，$B_{n,\lambda/n}$ は $n \to \infty$ で平均 $\lambda$ の *Poisson* 分布に弱収束する．弱収束の一般的な定義は後述．$\mathcal{F} = 2^\Omega$ の場合は，各 $k \in \mathbf{Z}_+$ 毎に $\lim_{n\to\infty} P_{n,\lambda/n}(\{k\}) = P(\{k\})$ が成り立つことを意味する．

**証明**.

$$
\begin{aligned}
&P_{n,\lambda/n}(\{k\}) \\
&= \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^k \prod_{i=1}^{k-1}\left(1-\frac{i}{n}\right).
\end{aligned}
$$

$n\to\infty$ として $\displaystyle\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = \exp(-\lambda)$ を用いる.

この定理は次の Poisson 小数の法則の特別な場合である.

**定理 15** $A_{n,k},\ k=1,2,\cdots,N_n,\ n=1,2,\cdots,$ は事象の列, $\lambda>0,\ Z$ は確率変数で, 次の条件を満たすとする：

*(i)* 各 $n$ 毎に $A_{n,k},\ k=1,2,\cdots,N_n,$ は独立.
*(ii)* $\max\{\mathrm{P}(\ A_{n,k}\ )\mid k=1,2,\cdots,N_n\}\to 0,\ n\to\infty.$
*(iii)* $\displaystyle\lim_{n\to\infty}\sum_{k=1}^{N_n}\mathrm{P}(\ A_{n,k}\ )=\lambda.$
*(iv)* $Z$ の分布は平均 $\lambda$ の *Poisson* 分布.

このとき $\displaystyle\sum_{k=1}^{N_n}\chi_{A_{n,k}}\to Z$ が成り立つ.

$N_n=n,\ A_k=\{X_k=1\},$ とおけば 定理 15 から命題 14 を得ることはすぐに分かる. 証明には特性関数を用いるが詳細は略す [楠岡, p.23].

**気持ち** 有限個の独立事象が, 一つ一つの起こる確率は小さいが,(十分たくさんあるので）実現する事象の個数の期待値は無視できないほど大きいならば, 何個起こるかの分布は個数の期待値だけで決まる Poisson 分布になる.(逆に言えば, そのような状況では, 観測データから関与する事象の総数を決めることはできない.)

### 6.2.3 単位分布

$D\subset\mathbf{R}$ を可算集合または有限集合 とする（例えば自然数）. 確率空間 $(D,2^D,P)$ が与えられているとき, $P$ を $(\Omega,\mathcal{F})=(\mathbf{R},\mathcal{B})$ 上の分布とみなす時の形式的な点を整理しておく.

$\mathcal{B}\supset\mathcal{I}$ なので, 可算加法性から $2^D\subset\mathcal{B}$. よって $2^D$ は $\mathcal{B}$ の部分 $\sigma$-加法族であることが分かる. そこで $\tilde P:\ \mathcal{B}\to\mathbf{R}$ を $\displaystyle\tilde P(A)=\sum_{x\in D\cap A}P(\{x\}),\ A\in\mathcal{B},$ で定義する. このとき $\tilde P$ が well-defined であり, $(\mathbf{R},\mathcal{B},\tilde P)$ が確率空間になることが定義から確かめられる.

特に, $A\subset D$ ならば $\tilde P(A)=P(A)$ になる. 即ち $\tilde P$ は $P$ の拡張になっている. 以後断らなければこのようにして得られる $\tilde P$ を $P$ と同一視し, 有理数（またはその部分集合）上の離散分布は全てこの意味で $(\mathbf{R},\mathcal{B})$ 上の分布とみなす.

特に, $D$ が 1 点 $(x_0\in\mathbf{R})$ よりなるとき, $P(\{x_0\})=P(D)=1$ となり, $\tilde P(A)=1,\ x_0\in A,\ \tilde P(A)=0,\ x_0\notin A,$ となる. これを単位分布と呼び, $\delta_{x_0}$ と書くこともある.

一般の場合は, $\displaystyle\tilde P=\sum_{y\in D}P(\{y\})\delta_y$ と書けることは明か.

**大数の法則** 事象数 $n$ を増やすと同時に一つの事象当たりの生起確率 $p$ を反比例で減らせば, 生起事象数 $\displaystyle Y_n=\sum_{k=1}^{n}X_k$ の分布は Poisson 分布に近づいた. これが 2 項分布 $B_{n,p}$ に対して $B_{n,\lambda/n}$ を考えることに相当した.

$p$ を固定して, 生起事象数ではなく生起率 $\displaystyle Y_n=\frac{1}{n}\sum_{k=1}^{n}X_k$ を考えると大数の法則を得る.

**命題 16** 確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数の無限列 $\{X_k\}$ が $\{0,1\}$ に値を取る独立同分布の確率変数で，その分布は $P_p$ であるとき，$Y_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ は確率 1 で $p$ に収束する．即ち $\mathrm{P}(\ \{\omega \mid \exists \lim_{n\to\infty} Y_n(\omega) = p\}\ ) = 1$.（このことを，$Y_n$ は $p$ に概収束するという.）

**注**. $A_k = \{\omega \in \Omega \mid X_k(\omega) = 1\}$ とおけば $A_k \in \mathcal{F}$ であって，$\{A_k\}$ は独立な事象である．従って独立な事象の平均出現数に関する法則とも言える．

より一般的に，独立同分布確率変数列の平均は（適当な条件の下で）一変数の期待値に概収束する．

**定理 17** 確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数の無限列 $\{X_k\}$ が独立同分布の確率変数で，その分布について 4 次のモーメントが有限 $\mathrm{E}[\ X_1^4\ ] < \infty$ ならば，$Y_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ は $p = \mathrm{E}[\ X_1\ ]$ に概収束する．即ち $\mathrm{P}(\ \{\omega \mid \exists \lim_{n\to\infty} Y_n(\omega) = p\}\ ) = 1$.

命題 16 について証明するが，この証明は 定理 17 に対しても通用する．

**命題** *16* **の証明**. $\epsilon > 0$ を任意に取り，$B_n = \{\omega \in \Omega \mid |Y_n - p| \geq \epsilon\}$ とおくと，独立性と期待値の線型性と $\mathrm{E}[\ X_k\ ] = p$ から

$$\begin{aligned}\mathrm{P}(\ B_n\ ) &\leq \epsilon^{-4}\mathrm{E}[\ (Y_n - p)^4\ ] \\ &\leq (n\epsilon)^{-4}\left\{\sum_{k=1}^{n}(\mathrm{E}[\ (X_k - p)^4\ ] - \mathrm{E}[\ (X_k - p)^2\ ]^2) + (\sum_{k=1}^{n}\mathrm{E}[\ (X_k - p)^2\ ])^2\right\} = O(n^{-2})\,.\end{aligned}$$

よって $\displaystyle\sum_{n=1}^{\infty}\mathrm{P}(\ B_n\ ) < \infty$ となるから Borel-Cantelli の定理から

$$\mathrm{P}(\ \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} {B_k}^c\ ) = 1\,,$$

即ち無限個の事象のうち有限個しか起こらない確率が 1 である．即ち確率 1 である $n$ 以降は $B_n$ が起きない．$Y_n \to p$ とならない確率が正だと矛盾することは容易に分かる．

## 6.3 連続分布

以下 $\Omega = \mathbf{R}$, $\mathcal{F} = \mathcal{B}$ 上の連続分布を考える．特に

$$\text{(13)}\quad \mathrm{P}(\ A\ ) = \int_A \rho(x)\,dx,$$

のように積分で書ける分布をルベーグ測度に対して絶対連続な確率測度と呼ぶ．$\rho$ を分布の密度と呼ぶ．このように書ける分布は連続分布であるが，大事な連続分布である．統計学では連続分布と言えば通常この場合のみ考える．

$\displaystyle\int_{\mathbf{R}} \rho(x)\,dx = 1$ でなければならない．

### 6.3.1 一様分布

$0 \leq a < b \leq 1$ に対して $P((a,b]) = b - a$, $A \cap [0,1] = \phi$ ならば $P(A) = 0$ を満たす $(\Omega, \mathcal{F})$ 上の確率測度がただ一つ存在する．これを $[0,1]$ 上の一様分布と呼ぶ．

**注** 自明に，任意の有限区間 $[a,b]$ 上の一様分布を考えることができる．無限区間では一様分布に対応する確率測度は存在しない．(規格化ができない.)

### 6.3.2 指数分布

$\lambda > 0$ に対して

(14) $\mathrm{P}(\ A\ ) = \lambda \int_{A\cap\{x>0\}} \exp(-x\lambda)\,dx, \quad A \in \mathcal{B},$

で定義される分布．つまり密度が $\lambda\exp(-x\lambda)\chi_{[0,\infty)}$ であるような分布．

**公式** $m = \lambda^{-1}$, $v = \lambda^{-2}$.

**Poisson process**

**命題 18** 確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数列 $\{T_k\}$, $k = 0, 1, 2, 3, \cdots$, が次を満たすとする．

*(i)* $0 \stackrel{\mathrm{def}}{=} T_0 \le T_1 \le T_2 \le \cdots$, *(*$\forall\omega \in \Omega$*)*,
*(ii)* $S_k \stackrel{\mathrm{def}}{=} T_k - T_{k-1}$, $k = 1, 2, 3, \cdots$, は独立同分布確率変数列,
*(iii)* $S_1 = T_1$ の分布は平均 $1/\lambda$ の指数分布.

$t \ge 0$ に対して $X_t(\omega) = \max\{k \in \mathbf{Z} \mid T_k(\omega) \le t\}$ とおく．このとき $X_t$ は非負整数に値を取る確率変数であって，平均 $t\lambda$ の *Poisson* 分布に従う.

**証明**. $X_t$ が非負整数に値を取る確率変数であることは定義から分かる．(10) から

$$\mathrm{P}(\ X_t = n\ ) = \mathrm{P}(\ T_n \le t < T_{n+1}\ ) = \mathrm{P}(\ \sum_{k=1}^{n} S_k \le t < \sum_{k=1}^{n+1} S_k\ )$$
$$= \int_{\sum_{k=1}^{n} x_k \le t < \sum_{k=1}^{n+1} x_k} \lambda^{n+1} \exp(-\sum_{k=1}^{n+1} x_k\lambda)\,dx_1\cdots dx_{n+1}\,.$$

$x_{n+1}$ で積分すると

$$\exp(-t\lambda)\int_{0\le\sum_{k=1}^{n} x_k \le t\lambda} dx_1\cdots dx_n$$

となるが，積分部分は一辺 $t\lambda$ の $n$ 次元三角錐の体積だから，Poisson 分布になっている.

**注**. この命題の逆も成り立つ．通常は $X_t$ にいくつかの性質を仮定すると，固定区間内の発生件数分布が Poisson 分布になることが導かれて，その後で発生間隔が指数分布になることを言う．大雑把に言うと，固定区間の発生件数が Poisson 分布であることと，発生間隔分布が指数分布であることが同値ということになる.

$X_t$, $t \ge 0$, は連続な変数をパラメータに持つ確率変数（の非可算個の集まり）である．このような確率変数の集まりを確率過程 (stochastic) process と呼ぶ．特に $\omega$ を固定する毎に $X_t(\omega)$ は $t$ の関数として 0 から始まって単調増加増分は 1 左連続な関数である．このような解析可能な穏やかな振る舞いをしている時に process の概念は有効である．解析可能な穏やかな振る舞いとは，連続関数（連続確率過程）や，左連続右有界 (cadlag) 関数を指す．$t$ を固定したときの $X_t$ の分布が正規分布になるような連続確率過程の有名な例が Wiener process である.

分布の話をしていたが，大事な話なので advanced topic ながら Poisson process に深入りしてみる.

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

(i) 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 ずつ増えるだけである（図 6.1）.
(ii) **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.

(iii) **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．
(iv) $X_0 = 0$．（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．）

$X_t - X_s$ 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，次の事実が証明できる．

**定理 19** 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．

定理 19 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

$$(15)\quad P[X_t - X_s = n] = \exp(-\lambda\,(t-s))\,\frac{1}{n!}\,\{\lambda\,(t-s)\}^n\,,\quad n = 0, 1, 2, \cdots,$$

に従う，ということである．ここで正定数 $\lambda$ は単位時間幅 1 内の到着数の期待値であることは Poisson 分布の定義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 (15) が得られることに注意．

**定理** *19* **の証明**．定理の証明をきちんと行う余裕はないが，一般原理（加法性，一様性）から (15) 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 $n$ 回事件が発生する（$X$ 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（図 6.2）．$N$ 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．）事件が $N$ 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 $\lambda$ 回事件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回事件が発生する．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），${}_N C_n\, p^n\,(1-p)^{N-n}$ 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 回以上事件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；

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

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．${}_N C_n = N!/(n!\,(N-n)!)$ 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 $\displaystyle\lim_{N\to\infty}\frac{N!}{\sqrt{2\pi N}\,N^N\exp(-N)} = 1$ 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 $\displaystyle\lim_{N\to\infty}(1-a/N)^N = \exp(-a)$ 及び $\displaystyle\lim_{N\to\infty}(1-a/N)^b = 1$ を用いれば (15) を得る．

$X_t$ が変化する（即ち 1 だけ増える）ことを，事件が発生した，とここだけの用語で呼ぶ.（時刻ゼロから測定開始して）初めて事件が発生した時刻を $T_1$, 以下 $n$ 件目事件発生の時刻を $T_n$, **時間間隔** を $S_n = T_n - T_{n-1}$ $(n \geq 1)$ とおく．但し，$T_0 \stackrel{\mathrm{def}}{=} 0$．$S_n$ 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．

**定理 20** $X_t$ が *Poisson* 過程ならば，$S_1$, $S_2$, $S_3$, $\cdots$, 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）．$S_1$ の分布は平均 $\mathrm{E}[\,S_1\,] = 1/\lambda$ の指数分布になる（どの $S_n$ でも同じ）．ここで $\lambda$ は *(15)* の $\lambda$，即ち，単位時間内の到着数の期待値 $\mathrm{E}[\,X_1 - X_0\,]$ である．

**定理 *20* の証明**. $\{S_n\}$ 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）．$S_n$ が平均 $1/\lambda$ の指数分布に従うことだけ証明しよう．定理 19 により，Poisson 過程の定義から (15) 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，(15) から (14) を得ることができればよい．$a \geq 0$ を定数とする．$S_1$, $S_2$, $\cdots$, $S_n$ をそれぞれある値に固定するという条件の下で $S_{n+1} > a$ となる条件付き確率（これを $P[S_{n+1} > a \mid S_0, S_1, \cdots, S_n]$ と書く）を計算する．$T_n = \sum_{i=1}^{n} S_i$ とおいておく．$S_{n+1}$ の定義から，求める確率は，時間間隔 $(T_n\,,\ T_n + a]$ の間に事件が発生しない確率，即ち，この時間に $X_t$ が変化しない確率に等しい．従って，

$$P[S_{n+1} > a \mid S_0, S_1, \cdots, S_n] = P[X_{T_n+a} = X_{T_n} \mid T_n] = \exp(-a\,\lambda)\,.$$

最後の等号は (15) において $n = 0$, $s = T_n$, $t = T_n + a$ とおいて得られる．右辺は $T_n$ によらないから，$S_1$, $S_2$, $\cdots$, $S_n$ の値によらない（即ち，これらの変数と $S_{n+1}$ は独立である）．だから，条件をはずしても等号が成り立つ；$P[S_{n+1} > a] = \exp(-a\,\lambda)$. この式と，(13) を見比べれば，$S_{n+1}$ の分布の密度は $\exp(-a\,\lambda)$ 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）(14) の右辺に他ならない．

#### 6.3.3 正規分布

$m \in \mathbf{R}$, $v > 0$, に対して $\mathrm{P}(\ A\ ) = \frac{1}{\sqrt{2\pi v}} \int_A \exp(-(x-m)^2/(2v))\,dx$, $A \in \mathcal{B}$, で定義される分布 $N_{m,v}$. Gauss 分布とも．

$X$ が $N_{m,v}$ に従えば

$$\mathrm{P}(\ X \in A\ ) = \frac{1}{\sqrt{2\pi v}} \int_A \exp(-(x-m)^2/(2v))\,dx\,,$$

$$\mathrm{E}[\ f(X)\ ] = \frac{1}{\sqrt{2\pi v}} \int f(x) \exp(-(x-m)^2/(2v))\,dx\,.$$

**諸性質の要約**

- $m = m$, $v = v$,
- 線型変換：$Y = (X-t)/u$ の従う分布は $N_{(m-t)/u, V/u^2}$,
- 和の分布：$X_i$, $i = 1, 2$, がそれぞれ $N_{m_i, v_i}$ に従えば，$X_1 + X_2$ は $N_{m_1+m_2, v_1+v_2}$ に従う．
- 独立同分布確率変数列 $\{X_k\}$ について（適当な条件下で），和 $S_n \stackrel{\mathrm{def}}{=} \sum_{k=1}^{n} X_k$ は $n \to \infty$ で
  - · 大数の法則 定理 17：$S_n = m\,n(1 + o(1))$, a.s.,
  - · 中心極限定理 定理 37：$S_n \sim N_{nm,nv}$,
- 特性関数 §6.4：$\phi(\xi) \stackrel{\mathrm{def}}{=} \mathrm{E}[\ \exp(\sqrt{-1}\xi X)\ ] = \frac{1}{\sqrt{2\pi v}} \int \exp(\sqrt{-1}x\xi) \exp(-(x-m)^2/(2v))\,dx$. 計算すると $\phi(\xi) = \exp(\sqrt{-1}m\xi - v\xi^2/2)$.

詳しい話は後述 §9.

### 6.4 特性関数

この節では証明はしない．

$\phi(\xi) \stackrel{\text{def}}{=} \mathrm{E}[\ \exp(\sqrt{-1}\xi X)\ ]$ を確率変数 $X$ の特性関数と呼ぶ．実部と虚部に分ければ，期待値を取るべき関数は三角関数と $X$ の合成関数だが，三角関数は連続関数なのでこれは可測であり，絶対値が 1 で押さえられるので,（有界収束定理より）期待値は存在して有限である．よって，どんな確率測度，どんな（実数値）確率変数でも特性関数は全ての $\xi$ で存在する．

また，確率測度 $P$ に対して $\phi(\xi) = \phi(\xi; P) \stackrel{\text{def}}{=} \int_{\mathbf{R}} \exp(\sqrt{-1}\xi x)\,P(dx)$ を $P$ の特性関数と呼ぶ．この定義は，確率分布が密度を持つ場合（通常の関数の積分で書ける場合）はフーリエ変換になっている．

確率測度は集合関数だが，特性関数は点関数なので通常の解析学的取り扱いが容易．$X$ は $\Omega$ 上の関数だが，特性関数は実数上の関数なので（$\Omega$ が解析的に良い性質を持っているとは限らないので），$X$ の分布を考える場合はやはり取り扱いが容易．さらに，分布と特性関数は次の意味で one to one に対応する．

**定理 21** $P_i,\ i = 1, 2,$ を分布とする．もし，$\phi(\xi; P_1) = \phi(\xi; P_2),\ \xi \in \mathbf{R},$ ならば $P_1 = P_2$.

（反転公式

$$\int \tilde{\chi}_{[a,b]}(x)P(dx) = \lim_{c\to\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{-1}{\sqrt{-1}\xi}(\exp(-\sqrt{-1}b\xi) - \exp(-\sqrt{-1}a\xi))\phi(\xi; P)d\xi$$

があるのでこれの正しいことが証明できる．$\tilde{\chi}$ は定義関数において，区間の端点で 1/2 としたもの.）逆は当然成り立つ．

測度と特性関数が one to one 対応なので，特性関数で一致を言えば分布も一致する．例えば「ある性質を持つ分布が既知のものに等しい」という命題を証明するのに便利.

確率変数列または分布の独立性を判定する（必要十分条件）こともできる．

**定理 22 (Kac)** 確率変数列 $X_1,\ X_2,\ X_3,\ \cdots,\ X_n$ が独立であることと全ての $\xi_i \in \mathbf{R},\ i = 1, 2, \cdots, n,$ に対して，$\mathrm{E}[\ \exp(\sqrt{-1}\sum_{k=1}^{n} \xi_k X_k\ ] = \prod_{k=1}^{n} \mathrm{E}[\ \exp(\sqrt{-1}\xi_k X_k)\ ]$ が成り立つことは同値である．

さらに，法則収束を特性関数の収束の言葉で表せるので，法則収束の議論に便利 §7.1.

モーメント $\mathrm{E}[\ X^n\ ]$ も特性関数の微分で計算できる．

**定理 23** 確率分布 $P$ の $n$ 次モーメントが存在すれば（即ち $\int |x|^n P(dx) < \infty$ ならば），$\phi(\xi, P)$ は $n$ 階連続微分可能であって，$\int x^n P(dx) = (-\sqrt{-1})^k \frac{\partial^k}{\partial \xi^k}\phi(\xi; P)\Big|_{\xi=0}$.

複素数値関数 $\phi$ が与えられたとき，それがある確率分布の特性関数になっているための必要十分条件も知られている．

**定理 24 (Bochner)** 実数上の複素数値関数 $\phi$ がある確率分布の特性関数になっていることと次の 3 条件の同時成立は同値である．

*(i)* $\phi(0) = 1,$

*(ii)* $\phi$ は原点で連続（結果として $\phi$ は連続関数になる），

*(iii)* 任意の $n \ge 1,\ \xi_k \in \mathbf{R},\ z_k \in \mathbf{C},\ k = 1, 2, \cdots, n,$ に対して $\sum_{k,l=1}^{n} \phi(\xi_k - \xi_l) z_k \bar{z}_l \ge 0.$

## 7 収束

解析学の最も本質的な概念は収束と極限であると言って過言では無かろう．確率空間においても収束の概念は数学の威力を見せつける．しかし，確率変数の収束はその直感的な意味からいろいろな同値でない定義が考えられていて，問題毎に使い分けられている．

## 7.1 分布の収束

**定義 3** 確率分布の列 $P_n$, $n \in \mathbf{N}$, が確率分布 $P$ に弱収束するとは，実数上で定義された任意の有界な実数値連続関数 $f$ に対して $\lim_{n\to\infty} \int f(x)\,P_n(dx) = \int f(x)\,P(dx)$ となることをいう．これをしばしば $P_n \to P$, *weakly*, $n \to \infty$, 等と書く．

弱収束と呼ぶが，特に確率論の問題で「強収束」という概念を用いるケースは知らない．弱収束には同値な種々の言い替えがある．確率分布，即ち $(\mathbf{R}, \mathcal{B})$ 上の確率測度の場合は上記の定義がやさしい．一般的な議論は [Billingsley] 参照．ここでは確率分布に即して同値条件を述べる．

**定理 25** $P$, $P_n$, $n \in \mathbf{N}$, を確率分布とする．このとき次の 3 つの条件は同値．

*(i)* $P_n \to P$, *weakly*, $n \to \infty$,
*(ii)* $\lim_{n\to\infty} \phi(\xi; P_n) = \phi(\xi; P)$, $\xi \in \mathbf{R}$,
*(iii)* $(\forall x \in \mathbf{R};\ P(\{x\}) = 0)\ \lim_{n\to\infty} P_n((-\infty, x]) = P((-\infty, x])$.

定理 25 は極限 $P$ を前もって知っていないと使えないが，特性関数の収束があれば弱収束が言える．これも特性関数の価値．

**定理 26** $P_n$, $n \in \mathbf{N}$, を確率分布とする．もし，各 $\xi \in \mathbf{R}$ に対して $\lim_{n\to\infty} \phi(\xi; P_n) = \psi(\xi; P)$ が存在して，かつ $\psi$ が $\xi = 0$ で連続な関数ならば，ある確率分布 $P$ が存在して，$P_n$ は $P$ に弱収束し，かつ，$\phi(\xi, P) = \psi(\xi)$, $\xi \in \mathbf{R}$, となる．

**証明**. 定理 24 から明らか.

## 7.2 確率変数の収束

確率変数の収束にはよく用いられるものだけで同値でない 4 種類がある．以下，この節 §7.2 では，確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数の無限列 $X_n$, $n \in \mathbf{N}$, と確率変数 $Y$ が与えられたとする．

### 7.2.1 概収束

定理 17 で先取りしたが，より一般的には概収束は次のように定義する．$X_n$ が $Y$ に概収束する (converge almost surely) とは

$$\mathrm{P}(\ \limsup_{n\to\infty} |X_n - Y| = 0\ ) = 1$$

が成り立つことと定義する.（$\limsup_{n\to\infty} |a_n - b| = 0$ と $\lim_{n\to\infty} a_n = b$ は同値であることに注意.）即ち，$\mathrm{P}(\ A\ ) = 1$ なる $A \in \mathcal{F}$ があって，$\lim_{n\to\infty} X_n(\omega) = Y(\omega)$, $\omega \in A$, となること．殆ど至るところ収束するという用語はこの意味で使われる．解析学の言葉では $A$ で各点収束ということ．

**記号** $X \to Y$, a.s.

### 7.2.2 確率収束

$X_n$ が $Y$ に確率収束する (converge in probability) とは

$$(\forall \epsilon > 0)\ \lim_{n\to\infty} \mathrm{P}(\ |X_n - Y| > \epsilon\ ) = 0$$

が成り立つことと定義する．

**記号**　$X \to Y$, in prob.

確率収束の定義より，次の定理が使いやすいことがある．

**命題 27** *$X \to Y$, in prob.,* と次の 2 つの条件は同値.

*(i)* $\lim_{n\to\infty} \mathrm{E}[\,|Y - X_n| \wedge 1\,] = 0.$
*(ii)* $\lim_{n\to\infty} \mathrm{E}[\,|Y - X_n|/(1 + |Y - X_n|)\,] = 0.$

**注**. $|\cdot|$ は $|\cdot|^p$ に替えても成立.

確率収束は平均収束 §7.2.4 に似ているが，平均収束が大きなずれを嫌うのに，確率収束はずれのある程度以上の大きさには鈍感（従って条件は弱い）ことが分かる．

### 7.2.3　法則収束

$X_n$ が $Y$ に法則収束する (converge in law) とは $X_n$ の分布が $Y$ の分布に（分布として）弱収束することと定義する．

**同値条件**

**定理 28** $\{X_n\}$ が $Y$ に法則収束することと次の各項はそれぞれ同値.

*(i)* $\lim_{n\to\infty} \mathrm{E}[\ \exp(\sqrt{-1}\xi X_n)\,] = \mathrm{E}[\ \exp(\sqrt{-1}\xi Y\ ],$
*(ii)* 任意の有界連続関数 $g(x)$ に対して $\lim_{n\to\infty} \mathrm{E}[\,g(X_n)\,] = \mathrm{E}[\,g(Y)\,].$

前半は 定理 25 より．

**記号**　$X \to Y$, in law.

### 7.2.4　平均収束

関数解析の視点から導入された収束の定義なので，今までの純確率論的視点からの定義とは少し毛色が違う，が，やはり重要.

$p \geq 1$ とする．$X_n$ が $Y$ に $p$ 次平均収束（$L^p$ 収束）する (converge in $L^p$) とは

$$\mathrm{E}[\,|Y|^p\,] < \infty,\ \mathrm{E}[\,|X_n|^p\,] < \infty,\quad n \in \mathbf{N},$$

であって，

$$\lim_{n\to\infty} \mathrm{E}[\,|X_n - Y|^p\,] = 0$$

が成り立つことと定義する．関数解析の言葉では $L^p$ norm に関する収束ということ．

### 7.2.5　収束の定義の関係

**定理 29** $X_n$ が $X$ に

*(i)* 概収束すれば確率収束する．
*(ii)* 確率収束すれば法則収束する．
*(iii)* 平均収束すれば確率収束する．

概収束 ↘
　　　　確率収束→法則収束
$L^p$ 収束 ↗

各矢印の「逆」には反例がある．

$(\Omega, \mathcal{F}, P)$ を区間 $(0, 1]$ 上の一様分布とする：$\Omega = [0, 1)$, $\mathcal{F} = \mathcal{B}$, $P([a, b)) - b - a$. $\Omega$ 上の確率変数列 $\{X_n\}$ と確率変数 $X$ について，

**確率収束し $L^p$ 収束するが概収束しない例**

$$X(\omega)=0,\quad X_n(\omega)=\begin{cases}1, & \omega\in[n\,2^{-[\log_2 n]}-1,(n+1)\,2^{-[\log_2 n]}-1),\\ 0, & \omega\notin[n\,2^{-[\log_2 n]}-1,(n+1)\,2^{-[\log_2 n]}-1),\end{cases}\quad \omega\in\Omega.$$

**法則収束するが確率収束しない例**

$$X(\omega)=X_1(\omega),\quad X_n(\omega)=[2^{n+1}\omega]-2[2^n\omega],\quad \omega\in\Omega.$$

**確率収束し概収束するするが $L^p$ 収束しない例**

$$X(\omega)=0,\quad X_n(\omega)=n\chi_{(0,1/n]},\quad \omega\in\Omega.$$

条件を追加すれば 定理 29 の「逆」も出る場合がある．

**命題 30** $X_n$ が $X$ に

*(i)* 確率収束してかつ $p+\epsilon$ ノルムに関して $\{X_n\}$ が有界になる（$\sup_n \mathrm{E}[\,|X_n|^{p+\epsilon}\,]<\infty$）$\epsilon>0$ があれば，$p$ 次平均収束する．

*(ii)* 法則収束して，かつ，$X$ の分布が単位分布 $\delta_a$ ならば，確率収束する．

**後半 の証明**. $X\to X-a$ によって $a=0$ でやればよい．このとき $X=0$, a.e.. 法則収束すれば有界連続関数 $g(x)$ に対して $\mathrm{E}[\,g(X_n)\,]\to\mathrm{E}[\,g(X)\,]$ (定理 28)．$g(x)=\mathrm{E}[\,x\cap 1\,]$ と 命題 27 より．

# 8 極限定理

独立確率変数列の和（平均）については古くから特徴的な量が詳しく分かっている．最も重要な量が平均（期待値），その次に分散，最後に平均から非常に遠く離れた値を取る確率，がとらえられる．それぞれ大数の法則，中心極限定理，大偏差値原理と呼ばれる定理がある．

独立でない変数列の場合はこれが良い描像とは限らない．統計力学では中心極限定理がそのままでは拡張できず，正規分布に近づかない，極めて難しい問題があることが分かっている．

この講義の後半では相関のある場合の難しい問題は避けて，正規分布に基づく統計学的推定という古典的な話題を紹介する．

## 8.1 大数の法則

独立同分布確率変数列の平均は（適当な条件の下で）一変数の期待値に概収束する．

**定理 31 (定理 17)** 確率空間 $(\Omega,\mathcal{F},P)$ 上の確率変数の無限列 $\{X_k\}$ が独立同分布の確率変数で，その分布について 4 次のモーメントが有限 $\mathrm{E}[\,X_1^4\,]<\infty$ ならば，$Y_n=\dfrac{1}{n}\displaystyle\sum_{k=1}^n X_k$ は $p=\mathrm{E}[\,X_1\,]$ に概収束する．即ち $\mathrm{P}(\,\{\omega\mid\exists\lim_{n\to\infty}Y_n(\omega)=p\}\,)=1$.

元の分布の詳細に関係なく極限が元の分布の平均だけで決まるという点が重大．元の分布を知らなくても極限が分かると言える一方，多数の変数の和（平均）しか見なければ元の分布についての情報はどんどん消えて最後まで残る情報が期待値（分布の平均）であるとも言える．分布あるいは確率変数の持つ情報の中での期待値の重要性を示す．

定理 17 よりも仮定はもっと弱められる．

**定理 32 (大数の法則２)** 確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数の無限列 $\{X_k\}$ が独立確率変数で，期待値が等しく $\mathrm{E}[\ X_k\ ] = p$，その分布について *2* 次のモーメントが有界 $\sup_k \mathrm{E}[\ X_k^2\ ] < \infty$ ならば，$Y_n = \dfrac{1}{n}\sum_{k=1}^{n} X_k$ は $p$ に概収束する．即ち $\mathrm{P}(\ \{\omega \mid \lim_{n\to\infty} Y_n(\omega) = p\}\ ) = 1$.

証明の詳細は [西尾, pp.139–141]．概略は次の通り．まず，次を証明する．

**定理 33 (kolmogorov の不等式)** 期待値 0 の独立確率変数列 $\{X_n\}$ に対して $S_n = \sum_{k=1}^{n} X_k$ とおくと $a > 0$ ならば $\mathrm{P}(\ \max\{|S_k| \mid k = 1, 2, \cdots, n\} \geq a\ ) \leq \mathrm{E}[\ S_n^2\ ]/a^2$

これの証明は $|S_k(\omega)|$ が初めて $a$ を超える $k$ によって $\Omega$ を分類.

**補題 34 (Kronecker)** 正項数列 $u_k$ が正の無限大に発散しているとする．このとき $\lim_{k\to\infty} \sum_{j=1}^{k} x_j/u_j$ が存在すれば $\lim_{k\to\infty} \dfrac{1}{u_k}\sum_{j=1}^{k} x_j = 0$.

これは初等解析学的.

定理 33 と 定理 3 から，次を得る.

**定理 35 (Kolmogorovconv)** 期待値 0 の独立確率変数列 $\{X_n\}$ に対して $S_n = \sum_{k=1}^{n} X_k$ とおく．

$$\sum_{k=1}^{\infty} \mathrm{V}[\ X_k\ ] < \infty$$

ならば $S_n$ は概収束かつ二次平均収束する.

補題 34 と 定理 35 から次を得る.

**定理 36** 期待値 0 の独立確率変数列 $\{X_n\}$ が $\sum_{k=1}^{\infty} \mathrm{E}[\ X_k^2\ ]/k^2 < \infty$ を満たせば，$\lim_{n\to\infty} \dfrac{1}{n}\sum_{k=1}^{n} X_k = 0,\ a.s.$

これから大数の法則 定理 32 を得るのは容易.

## 8.2 中心極限定理

各確率変数から $m$ を引いて和を取り，$\sqrt{n}$ で割った量は平均 0 分散が一変数の分布の分散 $v$ の正規分布 $N_{0,v}$ に法則収束する．乱暴に書けば $\sum_{k=1}^{n} X_k \sim N_{nm,nv}$ ということ．期待値と分散に関しては 補題 39 通り．それ以外の量の効果が消えることの意味を正確にするのが中心極限定理.

**定理 37 (中心極限定理)** $\{X_n\}$ は独立同分布確率変数列で $\mathrm{E}[\ X_1\ ] = m,\ \mathrm{V}[\ X_1\ ] = v < \infty$ とすると，$Z_n = \dfrac{1}{\sqrt{n}}\sum_{k=1}^{n}(X_k - m)$ の分布は $N_{0,v}$ に弱収束する.

弱収束（確率変数の法則収束）を言うには特性関数を調べて定理 25 等を利用するのが便利.

**証明**. 定理 22 と 定理 23 より

$$\mathrm{E}[\ \exp(\sqrt{-1}\xi Z_n)\ ] = \mathrm{E}[\ \exp(\sqrt{-1}(\xi/\sqrt{n})(X_1 - m))\ ]^n = (1 - \frac{v}{2}\,\xi^2/n + o(1/n))^n$$
$$= \exp(-\frac{v}{2}\,\xi^2)(1 + o(1))\,.$$

定理 25, 定理 28 より，定理を得る.

いくつかの形（結果が成り立つために十分な仮定の組み合わせ）がある．

**定理 38 (中心極限定理２)** $\{X_n\}$ は独立（同分布と限らない）確率変数列で各々３次モーメントを持ち，$\mathrm{E}[\, X_n\, ] = 0$, $n \in \mathbf{N}$, とする．さらに，$v_n = \sum_{k=1}^{n} \mathrm{V}[\, X_k\, ]$, $C_n = \sum_{k=1}^{n} \mathrm{E}[\, |X_k|^3\, ]$, と置くときに，

$$\lim_{n\to\infty} v_n^{-1} \max\{\mathrm{V}[\, X_k\, ] \mid k = 1, 2, \cdots, n\} = 0, \quad \lim_{n\to\infty} v_n^{-3/2} C_n = 0$$

とする．このとき，$\dfrac{1}{v_n} \sum_{k=1}^{n} X_k$ の分布は $N_{0,1}$ に弱収束する．

中心極限定理は分散や正規分布の重要性の根拠となる定理．既に注意したことだが，独立確率変数についての分散の加法性を思い出しておく．

**補題 39** $X_k$, $k = 1, 2, \cdots, n$, は２次モーメントを持つ独立な確率変数とし，$Y = \sum_{k=1}^{n} X_k$ とおく．このとき $Y$ も２次モーメントを持ち，

$$\mathrm{E}[\, Y\, ] = \sum_{k=1}^{n} \mathrm{E}[\, X_k\, ], \quad \mathrm{V}[\, Y\, ] = \sum_{k=1}^{n} \mathrm{V}[\, X_k\, ],$$

が成り立つ．

期待値の加法性は独立でなくても成り立つ．分散の加法性は独立性から導かれる $\mathrm{E}[\, X_i X_j\, ] = \mathrm{E}[\, X_i\, ]\mathrm{E}[\, X_j\, ]$, $i \neq j$, によって証明できる．

## 8.3 Kolmogorov の 0–1 law

発展的話題として Kolmogorov の 0–1 law を取上げる．

**定義 4** $\sigma$-加法族列 $\mathcal{B}_k$ に対して末尾事象とは $\bigcap_{k=1}^{\infty} \sigma\left(\bigcup_{j=k}^{\infty} \mathcal{B}_j\right)$ の元のこと．この加法族に関して可測な関数を末尾関数と呼ぶ．

**定理 40 (Kolmogorov's 0–1 law)** 独立な $\sigma$-加法族列の末尾事象の確率は 0 または 1 である．

**証明**. 詳しくは [西尾, 六章 §1 定理 2] 参照. 独立 $\sigma$-加法族列 $\mathcal{B}_k$ に対して末尾事象 $\Lambda$ をとる．$\mathcal{A} = \bigcup_{k=1}^{\infty} \sigma\left(\bigcup_{j=1}^{k} \mathcal{B}_j\right)$ とおくと $\Lambda \in \sigma(\mathcal{A})$ となるので近似定理（[西尾, 二章 §3 定理 5]）により $\Lambda' \in \mathcal{A}$ で近似できる．しかも末尾事象だから $\Lambda$ と $\Lambda'$ は独立．よって $P(\Lambda) \approx P(\Lambda)P(\Lambda')$ となる．

**系 41** 独立 $\sigma$-加法族列 $\mathcal{B}_k$ に対する末尾関数 $X$ は定数である．

**証明**. 定理 40 を $\Lambda = X^{-1}(-\infty, x])$ に適用すると分布関数 $F(x) = \mathrm{P}(\, \Lambda\, )$ はある 1 点で 0 から 1 にとぶので $X$ は確率 1 で定数.

これを用いるとたとえば次の事実が証明できる．

**命題 42** 独立確率変数列 $X_n$ と，収束または正の無限大に発散する正数列 $b_n$ に対して確率変数列 $Y_k = \sum_{j=1}^{k} X_j / b_k$ が収束する確率は 0 または 1 であり，$\lim_{n\to\infty} b_n = \infty$ ならば収束するときは極限は定数である．

証明は [西尾, p.136] 参照.

# 9　正規分布（続）

$m \in \mathbf{R}$, $v > 0$, に対して $\mathrm{P}(\ A\ ) = \sqrt{2\pi v}^{-1} \int_A \exp(-(x-m)^2/(2v))\,dx$, $A \in \mathcal{B}$, で定義される分布 $N_{m,v}$. $X$ が $N_{m,v}$ に従えば

(16)
$$\mathrm{P}(\ X \in A\ ) = \frac{1}{\sqrt{2\pi v}} \int_A \exp(-(x-m)^2/(2v))\,dx\,,$$
$$\mathrm{E}[\ f(X)\ ] = \frac{1}{\sqrt{2\pi v}} \int_{\mathbf{R}} f(x) \exp(-(x-m)^2/(2v))\,dx\,.$$

正規分布に基づく推定・検定が広く用いられる．分散までで決まる単純さの割に比較的実用性あり．よく使われる（周知性）．その他に論理的根拠．独立同分布確率変数列 $\{X_k\}$ について（適当な条件下で），和 $S_n \stackrel{\mathrm{def}}{=} \sum_{k=1}^{n} X_k$ は $n \to \infty$ で

- 大数の法則 定理 17：$S_n = m\,n(1+o(1))$, a.s.,
- 中心極限定理 定理 37：$S_n \sim N_{nm,nv}$.

**注**. 大部分の統計学の教科書は $X$ の従う分布や $Y = (X-t)/u$ の従う分布と書くべきところを積分変数を使って $x$ や $y$ の従う分布と書いている．これは $X$ の値域を $\Omega$ に取り直して $X(x) = x$ と考えていると思えばよい．

## 9.1　Gauss 積分の諸公式

後半で具体計算や表を利用するとき必要なので，$X$ の分布が $N_{m,v}$ のとき (16) を用いて具体的に計算してみる．

### 9.1.1　モーメント

**規格化**　先ず，$\mathrm{E}[\ 1\ ] = 1$ でないといけない．これは定義式の定数 $\frac{1}{\sqrt{2\pi v}}$ （規格化定数）を決める定義式と見ることができる．規格化が正しいことを確認する．

$$\mathrm{E}[\ 1\ ] = \frac{1}{\sqrt{2\pi v}} \int_{\mathbf{R}} \exp(-(x-m)^2/(2v))\,dx$$

なので，次の命題を証明すればよい．

**命題 43 ([小針, p.111, 命題 5.4])** $\int_{\mathbf{R}} \exp(-(x-m)^2/(2v))\,dx = \sqrt{2\pi v}$. 特に，$I \stackrel{\mathrm{def}}{=} \int_0^{\infty} \exp(-x^2)\,dx = \frac{\sqrt{\pi}}{2}$.

**補題 44** $J_n = \int_0^{\pi/2} \sin^n\theta\,d\theta$ とおくと，$n \in \mathbf{Z}_+$ に対して以下が成立.

*(i)* $J_n$ は $n$ について単調減少,
*(ii)* $J_{n+2} = \dfrac{n+1}{n+2} J_n$,
*(iii)* $(n+1)J_n J_{n+1} = \dfrac{\pi}{2}$.

**証明**. $J_n$ が減少することは定義から．部分積分で

$$J_{n+2} = -\cos\theta \sin^{n+1}\theta\Big|_0^{\pi/2} + (n+1)\,(J_n - J_{n+2})\,,$$

より第 2 の結果を得る．これから $(n+2)J_{n+2}J_{n+1} = (n+1)J_nJ_{n+1}$ を得るので，$(n+1)J_{n+1}J_n = J_1J_0$. $J_0 = \pi/2$, $J_1 = 1$ は明らかだから，第 3 式を得る．

**命題** *43* **の証明**. 第 1 の性質については，変数変換 $x = m + t\sqrt{2v}$ で

$$\frac{1}{\sqrt{2\pi v}}\int_{\mathbf{R}} \exp(-(x-m)^2/(2v))\,dx = \frac{1}{\sqrt{\pi}}\int_{-\infty}^{\infty} \exp(-t^2)\,dt.$$

$x \to -x$ 対称性から

$$\int_{-\infty}^{\infty} \exp(-x^2)\,dx = 2\int_0^{\infty} \exp(-x^2)\,dx$$

よって $I$ についての結果を証明すればよい.

(17) $1 - x^2 \le \exp(-x^2) \le \dfrac{1}{1+x^2}$， $x \in \mathbf{R}$,

より，

$$\sqrt{n}\int_0^1 (1-x^2)^n\,dx = \int_0^{\sqrt{n}} \left(1 - \frac{x^2}{n}\right)^n dx \le I \le \int_0^{\infty} \frac{1}{\left(1+\frac{x^2}{n}\right)^n}\,dx = \sqrt{n}\int_0^{\infty} \frac{1}{(1+x^2)^n}\,dx\,.$$

（両辺の等号は変数変換 $x\sqrt{n} \to x$ から．不等号は (17) で $x = x'/\sqrt{n}$ から）．最左辺で $x = \cos\theta$, 最右辺で $x = \cot\theta$, とおいて $\sqrt{n}\,J_{2n+1} \le I \le \sqrt{n}\,J_{2n-2}$. ここで，$J_n = \displaystyle\int_0^{\pi/2} \sin^n\theta\,d\theta$. 補題 44 より $\sqrt{n}\,J_{2n+2} \le I \le \sqrt{n}\,J_{2n-3}$. 故に $nJ_{2n+1}J_{2n+2} \le I^2 \le nJ_{2n-2}J_{2n-3}$. 再び 補題 44 より ($n \to \infty$ を考えて) $I^2 = \pi/4$ を得る.

**注**. 別解：

$$I^2 = \int_0^{\infty}\int_0^{\infty} \exp(-x^2-y^2)\,dx\,dy = \int_0^{\infty} \exp(-r^2)r\,dr\frac{\pi}{2} = \frac{\pi}{4}\,,$$

の方が早い.

補題 44 のついでに，$J_n$ がらみの補題を用意しておく.

**補題 45** $\displaystyle\lim_{n\to\infty} \frac{2^{2n}\sqrt{\frac{2}{n}}}{{}_{2n}C_n} = \sqrt{2\pi}.$

**証明**. $J_n = \displaystyle\int_0^{\pi/2} \sin^n\theta\,d\theta$ とおくと，補題 44 より，

(18) $J_{2n} = {}_{2n}C_n\,2^{-2n}J_0 = {}_{2n}C_n\,2^{-2n}\pi/2.$

他方，補題 44 より $J_{2n} \le J_{2n-1} \le J_{2n-2} = \dfrac{2n}{2n-1}J_{2n}$ なので，$J_{2n}^2 \le J_{2n}J_{2n-1} \le \frac{2n}{2n-1}J_{2n}^2$. 再度 補題 44 より $\frac{\pi}{2} = 2nJ_{2n-1}J_{2n}$ よって

$$\lim_{n\to\infty} \sqrt{2n}J_{2n} = \sqrt{\frac{\pi}{2}}\,.$$

(18) と組み合わせて

$$\lim_{n\to\infty} \sqrt{2n}{}_{2n}C_n 2^{-2n}\frac{\pi}{2} = \sqrt{\pi/2}\,.$$

$\pi$ で割って逆数を取ればよい.

**期待値**

$$\mathrm{E}[\,(X-m)^n\,] = \frac{1}{\sqrt{2\pi v}}\int (x-m)^n \exp(-(x-m)^2/(2v))\,dx = \frac{1}{\sqrt{2\pi v}}\int x^n \exp(-x^2/(2v))\,dx.$$

よって，$n$ が奇数ならば $\mathrm{E}[\,(X-m)^n\,] = 0$ 特に $n = 1$ のとき $\mathrm{E}[\,X\,] = m$.

**分散**　$n$ が偶数の時：

$$M_n = \mathrm{E}[\ (X-m)^n\ ] = \frac{1}{\sqrt{2\pi v}} \int x^n \exp(-x^2/(2v))\, dx$$

とおくと，部分積分から

$$M_{2m+2} = -v\, \frac{1}{\sqrt{2\pi v}} \int x^{2m+1} \left(\exp(-x^2/(2v))\right)'\, dx = (2m+1) v M_{2m}\,.$$

$M_0 = \mathrm{E}[\ 1\ ] = 1$ だから $M_{2m} = (2m-1)!!v^m$. 特に $M_2$ をみると $\mathrm{V}[\ X\ ] = v$.

### 9.1.2　和の分布

**定理 46** $X_k,\ k = 1, 2\cdots, n,$ がそれぞれ $N_{m_k, v_k}$ に従う独立確率変数ならば，$c_0 + \sum_{k=1}^{n} c_k X_k$ は平均 $c_0 + \sum_{k=1}^{n} c_k m_k$，分散 $\sum_{k=1}^{n} c_k^2 v_k$ の正規分布に従う.

これは次の線型変換と和の分布，及び独立確率変数の定数倍や排他部分和が再び独立確率変数になること，を用いれば証明できる.

**線型変換**

**命題 47** $Y = (X-t)/u$ の従う分布は $N_{(m-t)/u, v/u^2}$. 特に $t = m,\ u = \sqrt{v} = \sigma,$ として，$X$ が $N_{m,\sigma^2}$ に従うとき $Y = (X-m)/\sigma$ は $N_{0,1}$ に従う.

**証明**. 上でも用いたが，積分の変数変換 $x = uz + t$ で容易に分かるように，

$$\mathrm{P}(\ Y \in A\ ) = \frac{1}{\sqrt{2\pi v}} \int_A \exp(-(uz+t-m)^2/(2v)) u\, dz$$

だから.

**和の分布**

**命題 48** $X_k,\ k = 1, 2,$ がそれぞれ $N_{m_k, v_k}$ に従う独立確率変数ならば，$X_1 + X_2$ は $N_{m_1+m_2, v_1+v_2}$ に従う.

**注**. 期待値と分散がそれぞれ和で表されるのは，期待値の線型性と独立確率変数の分散の加法性から直ちに分かる.

正規分布でなくても一般に次の公式が成り立つ.

**補題 49** $X_k,\ k = 1, 2,$ がそれぞれ密度 $\rho_k$ を持つ分布に従う独立確率変数ならば，$X_1 + X_2$ は密度 $\rho(x) = \int \rho_1(t)\, \rho_2(x-t)\, dt$ を持つ.

**証明**. 独立確率変数なので (11) より，$(X_1, X_2)$ の従う分布 $\tilde{Q}$ は $B \in \mathcal{B}_2$ に対して

$$\tilde{Q}(B) = \int_{(x_1,x_2)\in B} \rho_1(x_1) \cdot \rho_2(x_2)\, dx_1 dx_2\,.$$

従って $X_1 + X_2$ の従う分布 $Q$ は $A \in \mathcal{B}$ に対して $B = \{(x_1, x_2) \mid x_1 + x_2 \in A\}$ とおけば

$$Q(A) = \mathrm{P}(\ X_1 + X_2 \in A\ ) = \int_{x_1+x_2\in A} \rho_1(x_1) \cdot \rho_2(x_2)\, dx_1 dx_2\,.$$

変数変換 $x_1 + x_2 = x,\ x_1 = t$ によって

$$Q(A) = \int_{x\in A} \left( \int \rho_1(t)\, \rho_2(x-t)\, dt \right) dx\,.$$

**命題 *48* の証明**. 補題 49 で $\rho_k(x)=\dfrac{1}{\sqrt{2\pi v_k}}\exp(-(x-m_k)^2/(2v_k))$, $k=1,2$, とおけば $X_1+X_2$ の分布の密度 $\rho$ を得る．$v=v_1v_2/(v_1+v_2)$ とおくと,

$$
\begin{aligned}
\rho(x) &= \int \rho_1(t)\,\rho_2(x-t)\,dt \\
&= \frac{1}{2\pi\sqrt{v_1v_2}}\int dt\,\exp-\left[\frac{1}{2v}(t-v(\frac{m_1}{v_1}+\frac{x-m_2}{v_2}))^2\right. \\
&\quad \left.+\frac{m_1^2}{2v_1}(1-\frac{v}{v_1})-\frac{vm_1(x-m_2)}{v_1v_2}+\frac{(x-m_2)^2}{2v_2}(1-\frac{v}{v_2})\right] \\
&= \frac{1}{\sqrt{2\pi(v_1+v_2)}}\exp\left[-\frac{1}{2(v_1+v_2)}\left(m_1^2-2m_1(x-m_2)+(x-m_2)^2\right)\right] \\
&= \frac{1}{\sqrt{2\pi(v_1+v_2)}}\exp\left[-\frac{(x-m_1-m_2)^2}{2(v_1+v_2)}\right].
\end{aligned}
$$

ここで 命題 43 を用いた.

**注**. $X_1$ と $X_2$ がそれぞれノルウェー人と日本人の身長の分布を表すとき和 $X_1+X_2$ の分布はノルウェー人と日本人の和集合の身長の分布ではない [小針, p.118]．$X_1$ と $X_2$ が独立という仮定なので，$X_1+X_2$ はノルウェー人と日本人のペア毎にノルウェー人の上に日本人が乗った時の全長を測り，全てのペアにわたる測定を行ったときの分布である.

### 9.1.3 特性関数

§6.4 より，正規分布 $N_{m,v}$ の特性関数は $\phi(\xi)\overset{\text{def}}{=}\mathrm{E}[\,\exp(\sqrt{-1}\xi X)\,]=\dfrac{1}{\sqrt{2\pi v}}\displaystyle\int\exp(\sqrt{-1}x\xi)\exp(-(x-m)^2/(2v))\,dx$ で与えられる．指数部を平方完成

$$\sqrt{-1}x\xi-(x-m)^2/(2v)=-(x-m-\sqrt{-1}\xi v)^2/(2v)+(\sqrt{-1}2m\xi v-\xi^2v^2)/(2v)$$

して，複素積分（正則関数）に持ち込めば計算できる．計算すると

(19) $\phi(\xi)=\exp(\sqrt{-1}m\xi-v\xi^2/2)$.

### 9.1.4 中心極限定理（例）

確率空間 $(\Omega,\mathcal{F},P)$ 上の確率変数の無限列 $\{X_k\}$ が $\{0,1\}$ に値を取る独立同分布の確率変数で，その分布は $P_p$ (§6.2.1) とし，和を $Y_n=\displaystyle\sum_{k=1}^{n}X_k$ とおく．$\mathrm{E}[\,X_1\,]=p$, $\mathrm{V}[\,X_1\,]=\mathrm{E}[\,X_1^2\,]-p^2=p-p^2$．平均 $\dfrac{1}{n}Y_n$ は大数の法則 命題 16 (定理 17) から確率 1 で $p$ に収束する．その収束の速さを記述するのが中心極限定理 定理 37. それによれば，$\dfrac{1}{\sqrt{n}}(Y_n-np)$ の分布は $N_{0,p-p^2}$ に弱収束する．線型変換すれば，$\dfrac{1}{\sqrt{np(1-p)}}(Y_n-np)$ の分布が $N_{0,1}$ に弱収束することを言えばよい．このことを直接計算で確かめてみる.

命題 13 より，$Y_n$ の分布は 2 項分布 $B_{n,p}$．$B_{n,p}(\{k\})={}_nC_kp^k(1-p)^{n-k}$, $N_{0,1}(dx)=\dfrac{1}{\sqrt{2\pi}}\exp(-\dfrac{x^2}{2})\,dx$.
$F(x)\overset{\text{def}}{=}N_{0,1}((-\infty,x])=\displaystyle\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp(-\frac{x^2}{2})\,dx$ を（正規分布の）分布関数と呼ぶ.

定理 25 より次の命題が成り立てばよい.（特性関数を使うこともできるが，それは 定理 37 の証明そのものなので検証は容易.）

**命題 50**

$$\lim_{n\to\infty}\sum_{k=0}^{np+x\sqrt{np(1-p)}} B_{n,p}(\{k\}) = F(x)\,.$$

**補題 51 (Stirling)** $\displaystyle\lim_{n\to\infty}\frac{n!}{\sqrt{2\pi n}n^n e^{-n}} = 1.$

**注**. $n!$ の近似式として極めて良く用いられる．$n=1$ でも誤差は 8.5% 以内.

**証明**. ([小針, pp.109–110]).

$$n\log n - n + 1 = \int_1^n \log x\,dx = \sum_{k=1}^{n-1}\log k + \frac{1}{2}\log n + \alpha_n$$

とおく．$\alpha_n$ は $[1,n]$ で $x$ 軸と $\log x$ が囲む面積から，$(n-1/2, n+1/2]$ で $\log n$ となる階段関数 $g(x)$ が囲む面積を引いた値．$k = 1, 2, \cdots, n-1$ に対して $(k, k+1/2]$ と $\log x$ と $g(x)$ が囲む面積を $a_{2k-1}$, $(k+1/2, k+1]$ と $\log x$ と $g(x)$ が囲む面積を $a_{2k}$, とおくと $\alpha_n = \displaystyle\sum_{k=1}^{2n-2}(-1)^{k-1}a_k$. $\log x$ は単調増加上に凸なので，図を描けば明らかなように $a_1 > a_2 > \cdots > a_{2n-2}$. また，$0 < a_{2n} < (\log n - \log(n-1))/2 \to 0$, $n\to\infty$. よって $\{\alpha_n\}$ は収束．極限を $\alpha$ とする．$A = \exp(1-\alpha)$ とおくと，

$$A = \lim_{n\to\infty} n!/(n^{n+1/2}e^{-n}) = \lim_{n\to\infty}(2n)!/((2n)^{2n+1/2}e^{-2n})\,.$$

よって 補題 45 より

$$A = \lim_{n\to\infty} 2^{2n}\sqrt{2/n}/{}_{2n}C_n = \sqrt{2\pi}\,.$$

**命題** *50* **の証明**. 詳しくは [楠岡 2]．$q = 1-p$ とおく．$0 < a < \min\{p, q\}$ を固定し，$k$ の和を 3 つの領域に分ける．

(i) $\tilde{A}_n(M) = \{0 \le k \le n \mid |\frac{k}{n} - p| < \frac{M}{\sqrt{n}}\}$,

(ii) $I_n(M) = \{0 \le k \le n \mid \frac{M}{\sqrt{n}} \le |\frac{k}{n} - p| < a\}$,

(iii) $Z_n = \{0 \le k \le n \mid |\frac{k}{n} - p| \ge a\}$.

$d_n = \dfrac{n!}{\sqrt{2\pi n}n^n e^{-n}}$ とおくと補題 51 より $d_n$ は 0 と $\infty$ から bounded away. $f(x) = -x\log\dfrac{x}{p} - (1-x)\log\dfrac{1-x}{1-p}$ とおくと $f(p) = 0$ かつ，$x \le p$ で狭義単調増加，$x \ge p$ で狭義単調減少．さらに，

$$B_{n,p}(k) = {}_nC_k p^k q^{n-k} = \frac{d_n}{d_k d_{n-k}}\frac{1}{2\pi n\frac{k}{n}(1-\frac{k}{n})}e^{nf(k/n)}\,.$$

これらを用いると，

$$\lim_{n\to\infty}\sum_{k\in Z_n} B_{n,p}(k) = 0,$$

$$\lim_{M\to\infty}\sup_{n>0}\sum_{k\in I_n(M)} B_{n,p}(k) = 0,$$

$$\sup_{M>0}\lim_{n\to\infty}\sup_{x\in[-M/\sqrt{pq}, M/\sqrt{pq}]}|\sqrt{npq}B_{n,p}([np + x\sqrt{npq}) - \frac{1}{\sqrt{2\pi}}e^{-x^2/2}| = 0\,.$$

以上より

$$\lim_{n\to\infty}\sum_{k=0}^{np+x\sqrt{n}} B_{n,p}(\{k\}) = F(x)\,,$$

を得る．

# 第 2 部
# データ

データ：立論・計算の基礎となる既知のあるいは認容された事実・数値．[広辞苑]

## 10 母集団と標本 09

### 10.1 データ

確率論に限らず数学を現実に応用するには，現実を数学的に適切な形で表現しないといけない．データとは適切に表現された現実（の一断面）ということである．

ここでは，確率論の視点から見たデータを考える．確率論で不問にした確率とは何か（例えば全体集合 $\Omega$ は現実には何の集合か）が問われる．

確率論の視点からみたデータ（現実）は確率空間 $(\Omega, \mathcal{F}, P)$ の全体集合 $\Omega$ の元の一つ $\omega \in \Omega$ が実現したと考える．確率論ではこれを標本 (sample) と呼ぶ．

### 10.2 母集団と標本

もう一つの見方.「現実はただ一つのサンプル」と見るのでなく，現実の中に分布のサンプル列を見る．例：実験データの集積．統計調査（アンケート，国勢調査，試験成績分布）．

架空の世界の集合 $(\Omega, \mathcal{F}, P)$ の下で見るときは，測定あるいは調査は確率変数 $X_k,\ k = 1, 2, \cdots, N$.

**定義 5 無作為抽出** 調査が独立同分布確率変数列になるように調査対象を選ぶこと．例：乱数表を見ながらリストから選ぶ.

**標本** 独立同分布確率変数列 $X_k,\ k = 1, 2, \cdots, n$ のこと．実現した $\omega \in \Omega$ に対する独立確率変数列の値の数列 $X_k(\omega),\ k = 1, 2, \cdots, n$ のことも標本という（区別のため標本値 *sample value* ともいう）．列の長さ $n$ を標本の大きさという.（統計学では標本 *(sample)* は確率論と別の意味に使う.）

**母集団** 標本（確率変数）の（$(\mathbf{R}, \mathcal{B})$ 上の）確率分布．名前の付いている分布の場合，その名前をつけて，正規母集団などと呼ぶ．調査の対象（データ）の確率空間.

**経験分布** 標本の度数分布 $\displaystyle\frac{1}{n}\sum_{k=1}^{n}\delta_{X_k(\omega)}$.（$(\mathbf{R}, \mathcal{B})$ 上の離散確率分布になっている.）

**標本の平均** $\displaystyle\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$.

**標本の分散** $\displaystyle S^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2$.

標本の平均と分散はどちらも確率変数である．経験分布も分布を値域とする確率変数である．

**命題 52** 大きさ $n$ の標本の平均について

*(i)* $\mathrm{E}[\ \bar{X}\ ] = \mathrm{E}[\ X_1\ ]$,
*(ii)* $\mathrm{E}[\ S^2\ ] = \dfrac{n-1}{n}\mathrm{V}[\ X_1\ ]$.

**証明**. 分散だけが問題.

$$\mathrm{E}[\ S^2\ ] = \mathrm{V}[\ X_1\ ] - \mathrm{E}[\ (\bar{X} - m)^2\ ] = \mathrm{V}[\ X_1\ ](1 - \frac{1}{n}).$$

特に，定理 46 より次を得る．

**命題 53** 正規母集団 $N_{m,\sigma^2}$ から大きさ $n$ の標本 $X_k,\ k = 1, 2, \cdots, n,$ の平均を作るとその分布は $N_{m,\sigma^2/n}$ に従う．

### 10.3 経験分布 10

確率論から統計調査の用語を眺めたときの疑問.

(i) サンプル $X_1(\omega)$, $X_2(\omega)$, $\cdots$, $X_n(\omega)$, なる値の列．サンプルをたくさん集めることの意義は何か？
(ii) 母集団 $P \circ X_1^{-1}$ とは何か？実体がない架空の世界の投影を持ち込む意味は何か．現実に観測することができない分布ならば現実（データ，サンプル）だけを問題にすれば十分ではないか？
(iii) 経験分布は実現した値の分布である．確率論としては公理を満たせば構わないとはいえ，制御できない撹乱要因を $\Omega$ とする見方との関連はあるのか？

大数の法則から，$\bar{X}$ は $n \to \infty$ で母集団の平均に概収束するのでたくさんデータを集めれば母集団の平均値の推測の精度が（何らかの意味で）上がる，という期待が持てる．実はデータを十分集めれば母集団そのものが推測できることが知られている．

**定義 6** *(i)* 分布 $(\mathbf{R}, \mathcal{B}, P)$ に対して，$F(x) = \mathrm{P}(\ (-\infty, x]\ )$ を $P$ の分布関数と呼ぶ.（確率論の概念．実数の時たいへん便利な概念.）
*(ii)* 経験分布 $\displaystyle\frac{1}{n}\sum_{k=1}^{n}\delta_{X_k}$ の分布関数 $F_n(x)$ を経験分布関数 *(empirical distribution function)* と呼ぶ.（統計学の概念.）

$F_n(x) = \dfrac{1}{n}\sharp\{k \mid 1 \le k \le n,\ X_k \le x\}$
次の定理も証明しない.

**定理 54 (Glivenko–Cantelli, [数学辞典, 48F, 338E])** 経験分布 $\displaystyle\frac{1}{n}\sum_{k=1}^{n}\delta_{X_k}$ の分布関数 $F_n(x)$ は母集団 $P \circ X_1^{-1}$ の分布関数 $F(x)$ に概収束する．

（値域が実数でない場合の概収束の定義は同様に考える．値域空間の収束の定義が必要だが，一様収束位相による収束とする.）

従って，モーメント（平均や分散）も母集団の対応する量に概収束する．近づくスピードの速い量を選んで推定や検定を行うと少量のデータで強い結論を得る．通常，平均や分散は近づくスピードが速いので，サンプルの（分布ではなく）平均や分散に注目する．

未勉強：*higher order moment* は標本平均の母集団値の回りの分散の $n \to \infty$ での 0 への収束が遅いはず.

## 11 乱数

経験分布の分布関数は母集団の分布関数に概収束するので，サンプルの列を十分長く得れば元の分布をほぼ表している．

確率論的世界観では世界は確率空間 $(\Omega, \mathcal{F}, P)$ で理論的には記述される．他方，実験や観測はサンプル $\omega \in \Omega$ に対する観測の列 $X_k(\omega)$, $k = 1, 2, \cdots, n$. $k$ は固有名詞を意味する．統計的性質，即ち固有名詞を忘れれば，経験分布は元の分布に近づく．つまり $X_k$ が独立同分布確率変数（観測）ならばその経験分布は母集団に分布関数の意味で概収束.

理論と実験がこのような対応なので，何らかの理論と観測によって母集団の分布を推定したとき，そこからデータを取出したときどんな結果が期待されるかというサンプルを得てみたい．確率論的な法則をテストしたりそれに基づく予測をするための実験に相当.「感じをつかむ，直感の助け」として非常に有効である.

確率論的世界観の特徴：サンプルだから推定が正しくても実測がそのとおりになるわけではない.

確率空間上の確率変数列のサンプルを発生させる手続またはその手続の結果得られたサンプルを乱数と呼ぶ.

基本としてまず，独立同分布確率変数列のサンプル $X_k(\omega)$ を確保したい．数字の列

むかしは機械的方法（本当にさいころやコインを振る）で数字の列を得て表にして，化学便覧などの表に載せていた．現代は計算機のプログラムで毎回発生させる擬似乱数.

## 11.1 乱数につきまとう問題

(i) さいころを振って 3, 1, 2, 5, 1, 1, 6, 5, 3, 4 と出たとしよう．即ち $X_1(\omega) = 3, \cdots,$ なる $\omega \in \Omega$ が実現した（サンプル）ということである．独立同分布だからこれが起こる確率は $2^{-10} \approx 0.1\%$ だが，インチキとは思わない．なぜか？

(ii) さいころを振って 1, 1, 1, 1, 1, 1, 1, 1, 1, 1 と出たとしよう．これが起こる確率も $2^{-10} \approx 0.1\%$ だが，インチキと思う．なぜか？

このサンプルを一様分布の乱数とみると，両者は同じ資格がある．乱数の価値が実用上の目的から決ることを示唆している.

(i) 対立仮説（他のもっともらしい理論的可能性）との比較．自然科学の発展：迷信の打破または既存理論の修正．常に比較が行われる.「理論の特徴」が見えるサンプルがほしい．§18.2 も参照.

(ii) システムの構築．予想される間違いや人間の恣意．例：さいころが一様でないとすると，特定の目がよく出る間違いが可能性が高い (一方だけ重い)．人間にとって作りやすい数列（人間にとって分りやすい）は恣意が入りやすい.

この問題には立入らない.

## 11.2 擬似乱数につきまとう問題

擬似乱数：適当な確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数列 $\{X_n\}$ が $n$ の漸化式で与えられる．確率空間は通常漸化式の初期値で与える (たとえば $X_0(\omega) = \omega$. ) **決定論的なので独立ではあり得ない.**

簡単な例：$X_{n+1} = f(X_n)$.

**命題 55** $(\Omega, \mathcal{F}, P)$ 上の確率変数 $X_1 : \ \Omega \to \mathbf{R}$ が $\mathcal{B}/\mathcal{F}$ *measureable,* $f : \ \mathbf{R} \to \mathbf{R}$ が *Borel measureable,* とする．$X_2 = f(X_1)$ とおくと，$X_2$ が単位分布でないならば $X_1$, $X_2$ は独立ではない.

**証明**. 独立ならば任意の Borel 可測関数 $g$ に対して

$$\mathrm{E}[\, g(X_2)^2 \,] = \mathrm{E}[\, g(f(X_1))g(X_2) \,] = \mathrm{E}[\, g(f(X_1)) \,]\mathrm{E}[\, g(X_2) \,] = \mathrm{E}[\, g(X_2) \,]^2 .$$

$g(x) = \chi(x \geq a)$ とおくと $F(a) = \mathrm{E}[\, g(X_2) \,] = \mathrm{P}(\, X_2 \geq a \,)$ なので $X_1$ の分布の分布関数 $F(x)$ は各点で 0 または 1. $F(x)$ は単調増加なのである 1 点 $x_0$ があって，$F(x) = \chi(x \geq x_0)$. すなわち，$X_2$ の分布は単位分布.

非常に便利なので，原理的問題にもかかわらず普通は使われる.

any sequence of numbers produced within a computer will be deterministic. Indeed, one of the advantages in using such sequences in simulation is that they can be repeated if necessary, ([Ripley])

アルゴリズムに対する条件.

(i) $\{X_n\}$, $n = 1, 2, \cdots, N$, の経験分布が $N \to \infty$ で求める分布に近づく.
(ii) 全ての $n$ で $X_n$ の分布が求めるものになる
(iii) できるだけ「独立」らしく振舞う
(iv) 高速に生成できる

「独立」らしさとして何を検査するべきかは議論の余地がある．きりがないので，深い一般論には立入らない.

## 11.3　乱数の生成（一様分布） 11

通常の擬似乱数アルゴリズム論 [Devroye].

(i) $[0, 1)$ 上の一様分布に従う擬似乱数 (一様乱数) を定義．アルゴリズムに対する条件 §11.2 を注意深く検討しておく.
(ii) ほしい分布に対する擬似乱数を，一様分布に従う独立同分布確率変数列が定義されているとの仮定の下で，定義．実際にはサンプル列として一様乱数を用いる.

[Devroye] は後者のみ扱う．前者 (一様乱数) の初等的解説については，たとえば，[Knuth] や [Ripley] など. $(X_n, X_{n+1})$ を 2 次元に図示すると，格子状の構造が見える（相関の存在）. [Ripley] では，これがなるべく小さな正方形に近いものが均等分布しているので望ましいと考える．この立場から，いくつかの推奨値が調べられている.

### 11.3.1　一様擬似乱数生成プログラムと検査例

一様分布に従う独立確率変数 $U_k$, $k = 1, 2, \cdots$, を一様乱数と呼ぶことにする．あらゆる疑似乱数の生成に用いるので，独立性とスピードを両立させなければならない．ここでは一例のみ取り上げる.

**生成方法**　Multiplicative congruential generator （乗法的合同法乱数生成手続き）.

$$U_{n+1} = aU_n \bmod M,\ \ M = 2^{32},\ \ a = 48828125,\ \ U_0 = \omega = 1000001.$$

この選択は次の事実に基づく.

**命題 56 ([Knuth, pp.19–21])** $M = 2^b$ のとき最大周期が $M/4$ になり，初期条件 $X_0$ が奇数で $a = 3 \bmod 8$ *or* $5 \bmod 8$ のとき，そのときに限り最大周期を取る.

**検査項目**

(i) 期待値
(ii) 分散
(iii) $U_n$ と $U_{n+1}$ の共分散 $\displaystyle\frac{1}{n}\sum_{k=1}^{n}(U_1(f^k(\omega)) - m_1)(U_2(f^k(\omega)) - m_2)$ 及び標準偏差の積で割った相関係数
(iv) 経験分布の分布関数 $\displaystyle F_n(x) = \sum_{1 \le k \le n:\ U_k \le x} 1/n$ と一様分布の分布関数の差 $\displaystyle\sup_x |F_n(x) - x|$.

期待値の収束のスピード：中心極限定理 $(\bar{U} - m)/\sqrt{n}$ が収束ということは $m$ からのばらつきが $V/n^{1/2}$ の程度ということ.

$$(n, |0.5 - E|\sqrt{n}) = (4^1, 0.14),\ (4^2, 0.31),\ (4^3, 0.08),\ (4^4, 0.20),\ (4^5, 0.31),\ (4^6, 0.03),\ (4^7, 0.30).$$

$V^{1/2} = 0.289$ だから係数まで込めて理論通り.

**注**. なぜ分布関数か？分布を直接調べたい.

(i) 分布 $Q$: 集合関数なので例えば図示しにくい．$\mathcal{F} = 2^{\Omega}$ つまり各点で $Q$ が定義されていれば図示できるが，連続分布ではこれは不可能.

(ii) 特性関数 $\phi$: 複素関数なので図示しにくい．意味が分かりにくい.

(iii) 密度 $\rho$: あれば望ましい．離散分布は密度がない．経験分布の収束のように離散分布が連続分布に収束する場合は $\rho$ も $Q$ も使えない.

(iv) 分布関数 $F$: 実数上の確率測度（分布）の場合のみ図示できる.

一長一短があるが，分布（実数上の確率測度）では分布関数が理論的にも非常に有効.

検査プログラムは §A.

```
function rnd
real*8 r32,rnd
integer rrnd,irnd
parameter (r32=2d0**32, rrnd=48828125)
data irnd/1000001/
rnd=irnd/r32+0.5d0
irnd=irnd*rrnd
return
end
```

| data# | E[Uk;1 ≦ k<N]<br>E[Uk;1<k ≦ N] | V[Uk;1 ≦ k<N]<br>V[Uk;1<k ≦ N] | ρ [Uk,U{k+1}]<br>cov[Uk,U{k+1}] | sup\|x-Fn(x)\| |
|---|---|---|---|---|
| 4^ 1 | .4286521554 | .0286242852 | -.8911578013D+00 | .4094172099D+00 |
| 4 | .3609646145 | .0280972518 | -.2527282904D-01 | |
| 4^ 2 | .4215964326 | .0663321115 | -.6995458436D-01 | .2113944150D+00 |
| 16 | .4462228971 | .0786956629 | -.5054212419D-02 | |
| 4^ 3 | .4896514937 | .0751253890 | -.8512888788D-01 | .6061214139D-01 |
| 64 | .4971359519 | .0787568413 | -.6548087461D-02 | |
| 4^ 4 | .4875027783 | .0775528506 | -.3400480152D-01 | .4844151321D-01 |
| 256 | .4891295379 | .0782664401 | -.2649274246D-02 | |
| 4^ 5 | .5096296231 | .0797432234 | .3117112503D-01 | .2823329647D-01 |
| 1024 | .5095349538 | .0797541620 | .2485856466D-02 | |
| 4^ 6 | .4995255809 | .0834541058 | -.2494295474D-02 | .7477757288D-02 |
| 4096 | .4996289671 | .0834980116 | -.2082139481D-03 | |
| 4^ 7 | .5023655556 | .0839676313 | -.2770665070D-02 | .8970827097D-02 |
| 16384 | .5023395855 | .0839787909 | -.2326616423D-03 | |

理想一様乱数：E=0.5, V=1/12=0.0833, ρ=cov=0, sup=0

**注**. FORTRAN では整数型変数では足算および掛算で $\bmod 2^{32}$ で $[-2^{31}, 2^{31})$ に入るように定義していることを用いる．言語や処理系によって異なる．Debug option によっても異なる.

## 11.4　乱数の生成（非一様連続分布） 12

一様乱数 $U_k,\ k=1,2,\cdots,$ が与えられていると仮定して,（一様分布以外の）分布 $(\mathbf{R},\mathcal{B},Q)$ に従う乱数, 即ち，独立確率変数列 $X_1,\ X_2,\ \cdots,$ のサンプルを求める [Devroye].

この節 §11.4 では $U$ を $[0,1)$ 上の一様分布確率変数とする．$U$ を用いて $Q$ を分布に持つ確率変数 $X$ が定義できれば，$F^{-1}(U_k),\ k=1,2,\cdots,$ は $Q$ に従う乱数である.（独立性は $U_k$ の独立性に帰着する.）よって $Q$ を分布に持つ確率変数 $X$ を一様分布確率変数 $U$ を用いて定義すればよい.

### 11.4.1　Inversion method (逆関数法)

分布関数 $F(x)=\mathrm{P}(\ X\le x\ )$ は定義によって $F(-\infty)=0,\ F(\infty)=1,$ かつ単調増加．よって $(0,1)$ 上の逆関数 $F^{-1}(u)=\inf\{x\in\mathbf{R}\mid F(x)\ge u,\ 0<u<1\}$ が存在.

**定理 57 (Inversion method)** $F^{-1}(U)$ の分布は分布関数 $F$ に対応する分布 $Q$ である.

**証明**. $\mathrm{P}(\ F^{-1}(U)\le x\ )=\mathrm{P}(\ F(x)\ge U\ )=F(x)$. 最後の等号は $[0,1)$ 上の一様分布の定義である.

**注**. うまく行けば極めて簡明・強力だが，逆関数 $F^{-1}$ が簡単に計算できる形で求められないといけない．それが自明ではないところが限界.

#### 指数分布

**定理 58 (指数分布乱数)** $X=\lambda^{-1}\log(1/U)$ で与えられる乱数は，平均 $1/\lambda$ の指数分布 *(14)*

$$\mathrm{P}(\ A\ )=\lambda\int_{A\cap\{x>0\}}\exp(-x\lambda)\,dx,\quad A\in\mathcal{B},$$

に従う乱数.

**証明**. $1-U$ を改めて $U$ と見れば（一様分布なので）$X$ の分布の分布関数 $F$ は 定理 57 より，$F^{-1}(x)=\lambda^{-1}\log(1/(1-x))$, 即ち，$F(x)=1-\exp(-\lambda x)=\displaystyle\int_0^x\lambda\exp(-y\lambda)\,dy$. $F(x)=\mathrm{P}(\ X\le x\ )=\int_0^x\rho(y)\,dy$ と見比べれば，(14) より，$F$ は指数分布の分布関数になっている.

**注**. $1-U$ を $U$ にするのは $F^{-1}$ が増加関数でないと 定理 57 が使えないから.

### 11.4.2　Rejection method (棄却法)

密度 $f$; $\mathrm{P}(\ X\in A\ )=\displaystyle\int_A f(x)\,dx$ を持つ分布 $Q$ に従う確率変数 $X$ を考える．別に密度 $\tilde f$ を持つ分布 $\tilde Q$ があって，$\tilde X$ の分布は $\tilde Q$ とする．$0<R\le\displaystyle\inf_{x\in\mathbf{R}}\frac{\tilde f}{f}(x)$ なる定数 $R$ が取れるとする.

**定理 59 (Rejection method)** 一様乱数 $U$ と $\tilde X$ が独立ならば $\tilde\Omega=\{\omega\in\Omega\mid U'(\omega)<R\dfrac{f}{\tilde f}(\tilde X(\omega))\}$ とおくと $\mathrm{P}(\ \tilde\Omega\ )=R$ であって，$\tilde\Omega$ に制限したときの $\tilde X$ の分布（条件付き確率）は $Q$ に等しい：$Q(B)=\mathrm{P}(\ \tilde\Omega\cap\{\tilde X\in B\}\ )/R,\ B\in\mathcal{B}$.

実際に乱数列を生成するときは次のようにすればよい．$\tilde{Q}$ の分布関数 $\tilde{F}(x) = \int_{-\infty}^{x} \tilde{f}(y)\,dy$ の逆関数を $\tilde{h}(x) = \tilde{F}^{-1}(x)$ とする．定理 57 より，$\tilde{X} = \tilde{h}(U)$ の分布は $\tilde{Q}$.

**系 60** $u_k$, $u'_k$, $k = 1, 2, \cdots$, を独立な一様乱数のサンプル（実際は一列の一様乱数を順次使う），$\tilde{x}_k = \tilde{h}(u_k)$ とする．$R\,\dfrac{f}{\tilde{f}}(\tilde{x}_k) \geq u'_k$ を満たす $k = k_j$, $j = 1, 2, , \cdots$, を集めて部分列を作ると，$\tilde{x}_{k_j}$, $j = 1, 2, , \cdots$, は $Q$ に従う乱数のサンプルである．

**定理** *59* **の証明**. $U$, $X$ が独立だから (11) が使える．$R$ についての仮定より $Rf(x)/\tilde{f}(x) \leq 1$ なので

$$\mathrm{P}(\,\tilde{\Omega}\,) = \int_0^1 du \int_{u \leq Rf(x)/\tilde{f}(x)} \tilde{f}(x)\,dx = \int R\frac{f}{\tilde{f}}(x)\,\tilde{f}(x)\,dx = R\,.$$

このことを使うと，同様に,

$$\mathrm{P}(\,\tilde{\Omega} \cap \{\tilde{X} \in B\}\,)/R = \frac{1}{R}\int_0^1 du \int_{u \leq Rf(x)/\tilde{f}(x),\ x \in B} \tilde{f}(x)\,dx = \int_{x \in B} f(x)\,dx = Q(B)\,.$$

**注**. (i) 定理 59 の幾何学的な意味は [Devroye, II.3.1,Theorem 3.1]: $\tilde{X}$ と $U$ が独立ならば $(\tilde{X}, \tilde{f}(\tilde{X})\,U/R)$ の分布は 2 次元領域 $\tilde{\Omega} = \{(x, y) \mid 0 \leq y < \tilde{f}(x)/R\}$ 上の一様分布である．($\tilde{\Omega}$ は 2 次元ルベーグ測度 $1/R$ なので密度は $R$.)

(ii) 分布関数 $F$ の代わりに密度 $f = F'$ で $h = F^{-1}$ を表せば $f(h(x)) \times h'(x) = 1$.

(iii) 発生した列と選んだ列の比は $R$ (Acceptance rate)．これが 0 に近いと乱数列生成に時間がかかって役に立たない．

**von-Mises 分布**　$a > 0$ を定数とするとき $f(\theta) = N\exp(a\cos\theta)$, $N^{-1} = \int_{-\pi}^{\pi} \exp(a\cos\theta)\,d\theta$ を密度とする $[-\pi, \pi)$ 上の分布．統計力学の XY-model などのシミュレーションで用いられる．Inversion method が適用できない例．$a$ が広い範囲で動くような乱数列が必要なとき，分布のピークの鋭さが大きく変わるので，それに対応できる高速乱数生成法が必要.

単峰でピークの鋭さが乱数列の中で大きく変わるときに有効な方法 [服部中島]．$\tilde{f}(\theta) = \tilde{N}/(2\cosh(\alpha\theta)+2\beta)$ を用いる．$\tilde{N}$ は規格化定数．$\alpha, \beta$ は $a$ に応じて決める．(予め解析的考察により，高速計算できて $R$ が大きくなるように適当な関数形を探しておく.)

$$h(x) = \begin{cases} 2\alpha^{-1}\,\mathrm{arc\,tanh}(B^{-1}\tanh((2x-1)\,\mathrm{arc\,tanh}(AB))), & \beta > 1, \\ 2\alpha^{-1}\,\mathrm{arc\,tanh}((2x-1)A), & \beta = 1, \\ 2\alpha^{-1}\,\mathrm{arc\,tanh}(B^{-1}\tan((2x-1)\,\mathrm{arc\,tan}(AB))) & -1 < \beta < 1, \end{cases} \qquad \begin{aligned} A &= \tanh(\pi\alpha/2)\,, \\ B &= \sqrt{\tfrac{|\beta-1|}{\beta+1}}\,. \end{aligned}$$

[服部中島] で実際に用いたプログラムは §B を参照.

**正規分布**　大事な分布なので種々の方法が考案されているが polar method [Devroye, p.235, V.4.4] を取り上げる.

棄却法と変数変換と指数分布に従う乱数を巧みに用いて，正規分布に従う乱数を生成できる.

**定理 61** $U$, $U'$, $E$ は独立確率変数で，$U$, $U'$ の分布は $[0, 1)$ 一様分布，$E$ のは平均 1 の指数分布とする．$V = (2U-1)^2 + (2U'-1)^2$ とおく．$\tilde{\Omega} = \{V \leq 1\}$ に制限したときの $\sqrt{\dfrac{2E}{V}}(2U-1)$, $\sqrt{\dfrac{2E}{V}}(2U'-1)$ は正規分布 $N_{0,1}$ を分布に持つ独立な確率変数である.

定理 59 のようにすれば証明できるが証明は略.

### 11.5 乱数の生成 (Poisson 分布)

離散分布に従う乱数の例として Poisson 分布 §6.2.2 を取り上げる．命題 18 より，次を得る．

**定理 62** 独立同分布確率変数列 $\{E_k\}$, $k = 1, 2, 3, \cdots,$ の分布が平均 1 の指数分布の時 $X = \max\{k \in \mathbf{Z} \mid \sum_{i=1}^{k} E_i \leq \lambda\}$ の分布は平均 $\lambda$ の *Poisson* 分布．

アルゴリズムは次のようにするのが普通．

**系 63** $[0, 1)$ 一様乱数サンプル列 $\{u_k\}$, $k = 1, 2, 3, \cdots,$ に対して $x = \min\{k \in \mathbf{Z}_+ \mid \prod_{i=1}^{k+1} u_i < e^{-\lambda}\}$ で逐次生成される非負整数 $x$ の列は *Poisson* 分布に従う乱数サンプルである．

**注**. この方法は $\lambda$ に比例して時間がかかり，$\lambda$ が大きくなるほど誤差が増大するので，大きい $\lambda$ では他の方法が用いられる [Devroye, X.3].

# 第 3 部
# 統計学入門

13

## 12 統計的推測

**統計法則を得る手続き**　確率論：確率空間 $(\Omega, \mathcal{F}, P)$ が与えられているときに，確率変数 $X$ に関する統計量，分布 $P \circ X^{-1}$ や期待値 $\mathrm{E}[\, f(X)\,]$ など，が定義される．

これを現実へ応用すると，**現象の生起確率（確率空間）が分かっていれば**，測定・観測（i.i.d. 確率変数列 $Z = \{X_k \mid k \in \Lambda\}$）の結果即ちデータ（標本値 $Z(\omega)$）についての統計的な予測（分布，期待値）ができる，ということである．

しかし，このためには現象の原因，即ち統計法則（元の確率空間）を知らなければならない．そのためには，先ず，現象，即ち標本値が与えられていて，そこから統計法則を推測し，その法則（候補）から予測を行って，実験・観察と比較し，法則の正当性・有効性を確認しなければならない．

**推定，検定**　決定論的な法則（例えば微分方程式で書かれる力学の法則）は，一意的な結論が得られるので，これを予言として実験を行い，結果を比較して，法則の結論と実験結果が一致すれば法則の正当性（有効性）を認めることができる．

統計法則は原理的に一意的な予言はできないので，法則をどのように**推定**するか法則が正しいことをどのように**検定**するか，それ自体が考察を要する．

統計的推測を一般に推定と検定に分ける．

**検定**　無作為抽出されたサンプルをもとに，母集団に関する，ある程度信頼のおける命題を立てること．

母集団に関する適当な仮説（帰無仮説）を選び，その仮説下で適当な統計量（サンプルの関数）を選んで計算し，サンプル値（統計量の実現値）に基づいて，仮説の正否を判断する．帰無仮説が正しくないことを主張する形をとることが多い．[2]

例：帰無仮説 H: 「100 年前と現在で，日本人の平均身長は変化していない」

統計量：100 年前と現在それぞれの 50 例についての平均身長の標本平均と不偏分散に基づいて，偏差と標本平均の比．

比が非常に大きいというデータが得られたとする．H が正しいと仮定すると，そのようなことが起きる確率 $p$ は大変小さい．あらかじめ定められた数値 $\alpha$ （危険率）より $p$ が小さいときに，帰無仮説を棄却する．

[2] [渡辺浩] 一部改変.

## 13 標本平均と不偏分散

大きさ $n$ の標本，即ち (§10.2)，独立同分布確率変数列 $X_k,\ k=1,2,\cdots,n$, あるいは実現した $X_k(\omega)$, $k=1,2,\cdots,n$, が与えられたとき，母集団（$X$ の分布 $Q=P\circ X^{-1}$）を推定したい.

**母集団特性値 (population characteristic)** 典型的な（目安となる）量（分布から計算される量）：平均 $\mathrm{E}[\,X\,]$，分散 $\mathrm{V}[\,X\,]$，モーメント等.

**統計量** 標本から計算される量（$\{X_k\}$ の可測関数）：

**標本平均**

$$(20)\quad \bar{X}_n=\frac{1}{n}\sum_{k=1}^{n}X_k\,.$$

例：$\bar{X}_1=X_1$.

**不偏分散**

$$(21)\quad V_n=\frac{1}{n-1}\sum_{k=1}^{n}(X_k-\bar{X}_n)^2.$$

**点推定** 母集団特性値の推定値としての統計量．$\bar{X}_n$, $V_n$ を $\mathrm{E}[\,X\,]$, $\mathrm{V}[\,X\,]$ の推定値とする.（cf. 区間推定.）主な根拠 [林周二, 14.1],[数学辞典, 287C]：（当然 sample だけで書けなければならないが，さらに,）

(i) 一致性

$$(22)\qquad \lim_{n\to\infty}\bar{X}_n=\mathrm{E}[\,X\,], a.e.,$$

$$(23)\qquad \lim_{n\to\infty}\frac{1}{n-1}\sum_{j=1}^{n}(X_j-\bar{X}_n)^2=\mathrm{V}[\,X\,], a.e.$$

（証明：前者は $\mathrm{V}[\,X\,]<\infty$ くらいあれば大数の法則，後者は左辺 lim の中を変形すると，かってな定数 $\mu$ に対して

$$\frac{1}{n-1}\sum_{j=1}^{n}(X_j-\mu)^2-\frac{1}{n(n-1)}\left(\sum_{j=1}^{n}(X_j-\mu)\right)^2=\frac{n}{n-1}\left(\frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)^2-(\bar{X}_n-\mu)^2\right).$$

$\mu=\mathrm{E}[\,X\,]$ ととれば第 1 項は大数の法則から $\mathrm{E}[\,X^4\,]<\infty$ くらいで $\mathrm{V}[\,X\,]$ に概収束．第 2 項は大数の法則から 2 乗の中身が 0 に概収束[3] .）

(ii) 不偏性 $\mathrm{E}[\,\bar{X}_n\,]=\mathrm{E}[\,X\,]$, $\mathrm{E}[\,V_n\,]=\mathrm{V}[\,X\,]$.

(iii) その範囲で分散最小の量が望ましい（有効性）．適当な条件があると,（下記 $\theta$ に関して一様に）分散が最小の不偏推定量が一意に存在する.

**注**. (i) 母数（母集団の parameter）の推定値とするのが点推定のより現代的な用語定義 [数学辞典, 285A]. Sample space: $(\mathbf{R},\mathcal{B})$. 確率変数 $X:\ \Omega\to\mathbf{R}$ の定める sample space 上の分布 $Q=P\circ X^{-1}$ は確率測度の族 $\mathcal{P}=\{P_\theta\mid\theta\in\Theta\}$ に属していることは分かっているとする．$Q=P_\theta$ なる $\theta$ を母数 (parameter) の真の値という．$\Theta$: parameter space.

(ii) 分散が一様に最小の不偏推定量の一意存在の条件は十分性を含む．Fisher は一致性，有効性，十分性を推定量選択の根拠とした [林周二, 14.1], [数学辞典, 286F].

[3] 19960527 楠岡成雄氏証明.

### 13.1 最尤法

平均と分散以外の一般的な母数の点推定の古典的な方法 [林周二, 14.2].

$Q = P \circ X^{-1}$ の属する確率測度の空間の元 $P_\theta$ の密度 $f_\theta$; $P_\theta(A) = \int_{x \in A} f_\theta(x)\,dx$. 尤度関数 $L(\theta) = \prod_{j=1}^{n} f_\theta(X_j(\omega))$

**最尤法による点推定の原理**　$L$ の最大値を与える $\theta$ を採る：$\dfrac{\partial \log L}{\partial \theta} = 0$.

同じ標本値 $\{X_j(\omega)\}$ でも $P_\theta$ の選び方で結果は異なる．例えば $\mathrm{E}[\,X\,]$, $\mathrm{V}[\,X\,]$ については $\bar{X}_n, V_n$ とは異なる推定値を得る可能性がある．

例：正規分布 $N_{\mu,v}$ に従うことが分かっている（予想される）場合に，$\mu$, $v$ を標本値 $x_j = X_j(\omega)$, $j = 1, 2, \cdots, n$, から最尤法を用いて推測する．密度 $f_{\mu,v}$ は

$$f_{\mu,v}(x) = \frac{1}{\sqrt{2\pi v}} \exp(-(x-\mu)^2/(2v))$$

である．

$$\frac{\partial \log L}{\partial \mu} = -\frac{1}{v}\,(n\mu - \sum_{i=1}^{n} x_i) = 0$$

から $\mu^* = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} x_i$,

$$\frac{\partial \log L}{\partial (1/v)} = -\frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^2 + \frac{nv}{2} = 0$$

から $v^* = \dfrac{1}{n} \displaystyle\sum_{i=1}^{n} (\mu^* - x_i)^2$, を得る．

平均値の推定は標本平均だが，分散の推定値は不偏分散の $(n-1)/n$ 倍である．これはどちらかが間違っているのではなく，推定の方法によって推定値が異なることを言う．

**推測は特定の手続きを指す言葉ではなく，考え方の指導原理を述べる言葉．具体的な方法によって同じデータの下での同じ母数の推定値が異なることがある．どの方法を選ぶかは，先に結果についての根拠または予想が必要で，それによる．従って，問題毎に異なるし，同じ問題でも立場や考え方で異なる．**

以下では方法の選択については素朴な原理のみを例示する．殆どの議論は，よく用いられる推測の方法の説明，即ち，選択肢の提示である．

最尤法に関する一般的結果は [数学辞典, 287M] 参照.

## 14 正規母集団に関する推測

古典的な数理統計学は母集団が正規分布に従うことが予め分かっている場合の理論から始まった．非常に整備されていて，簡単な方法としてよく用いられる．

**重大な欠陥**　予め母集団がどんな分布に従うか分かっていれば統計的推測は必要ない！

**数学と現実の関係**　データは有限で，母集団（確率測度）は無限の可能性があるから，統計的推測を行うのには必ず母集団について仮定が必要．統計学でなくても，他の数学でも現象に当てはめるときは**必ず**とてつもなく強い仮定をおいて考察する．非常によく統制された自然科学の法則検証の実験を除いて，たいていの現実的な問題への数学の応用では厳密には成り立たない仮定をおく．統計学ではこの問題が直感的でない形で現れるので特別な考察が必要.

**適切な使い方** 母集団の分布を仮定すると，サンプルの分布が決まる．サンプル（確率変数列）は独立性の仮定から，中心極限定理が成り立つので，統計量の分布は正規分布または正規分布母集団に基づく導出で計算できる（具体例参照）．それに対して正規母集団に関する検定を適用すれば，遡って母集団に対する仮定を（統計的に）棄却できる．あるいは，母集団に対する仮定の中にパラメータを含めておけば，上記の手続きを経て，母集団に関する推定が可能になる．

§14 では母集団が正規分布に従うことが分かっている，または，近似できることがわかっている，ときに成立する統計的推測（推定・検定）の方法を紹介する．

## 14.1 正規分布

母集団（確率変数 $X$ の分布）が平均 $\mu$, 分散 $\sigma^2$ の正規分布 $N(\mu,\sigma^2)$ に従うとは密度が

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

で表されること，即ち，

$$Q(A)=\mathrm{P}(\ X\in A\ )=\int_A f_{\mu,\sigma^2}(x)\,dx$$

のこととする．

**規格化** $\displaystyle\int_{\mathbf{R}} f_{\mu,\sigma}(x)\,dx=1$ は既出 命題 43.

**平均** $\displaystyle\mathrm{E}[\ X\ ]=\int x\,f_{\mu,\sigma^2}(x)\,dx=\mu.$

**分散** $\displaystyle\mathrm{V}[\ X\ ]=\int (x-\mu)^2\,f_{\mu,\sigma^2}(x)\,dx=\sigma^2.$

分布関数

(25) $$F(x)=\int_{-\infty}^{x} f_{0,1}(y)\,dy=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{y^2}{2}\right)\,dy$$

を表にしておくと統計的推測では便利 [小針, pp.286], [林周二, 11.1]．$f(-x)=f(x)$ より $F(x)=1-F(1-x)$ なので $x\geq 1/2$ のみ知っていればよい．$F$ は狭義単調増加で $F(\infty)=1$.

| x | F(x) |
|---|---|
| 0 | 1/2 |
| 1 | 0.8413 |
| 1.6448 | 0.9500 |
| 1.9600 | 0.9750 |
| 2 | 0.9772 |
| 2.326 | 0.9900 |
| 2.5758 | 0.9950 |
| 3 | 0.9986 |
| 4 | 0.99997 |

.1[11]

**不偏分散と標本平均の独立性** 正規母集団のサンプルについては，標本平均と不偏分散は独立である．このことは §14.4.5 で用いるので，ここに取り上げる．

**命題 64** $X_i,\ i=1,2,\cdots,n,$ が正規母集団 $N(\mu,\sigma^2)$ の大きさ $n$ のサンプル，即ち，*i.i.d.*, ならば，不偏分散 $V_n$ と標本平均 $\bar{X}_n$ は独立である．

[11] Hayashi

**証明**. 結合分布の密度は

$$\mathrm{P}(\ \{X_i\} \in A\ ) = \int_A \exp(-\sum_{i=1}^n \frac{(x_i-\mu)^2}{2\sigma^2})\, d^n x = \int_A \rho(\{x_i\})\, d^n x\,.$$

他方，$\bar{x} = \dfrac{1}{n}\displaystyle\sum_{j=1}^n x_j$ とおくと，

$$n(\bar{x}-\mu)^2 + \sum_{i=1}^n (x_i-\bar{x})^2 = \sum_{i=1}^n (x_i-\mu)^2\,.$$

よって，

$$-\log\rho(\{x_i\}) = \frac{1}{2\sigma^2}\left(n(\bar{x}-\mu)^2 + \sum_{i=1}^n (x_i-\bar{x})^2\right).$$

また，変数変換 $(\{x_i\}) \to (\bar{x}, \{\delta x_j\})$ を，$\bar{x} = \dfrac{1}{n}\displaystyle\sum_{j=1}^n x_j$ と $\delta x_j = x_j - x_{j+1}$, $j = 1, 2, \cdots, n-1$, で定義すると，$d^n x = \dfrac{2}{n}\, d\bar{x}\, d^{n-1}\delta x$ となるので，従って，

$$\begin{aligned}&\mathrm{P}(\ V_n \in B,\ \bar{X}_n \in C\ ) = \frac{2}{n}\int_{\bar{x}\in C} \exp\left(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2})\right) d\bar{x}\ \times \int_{v(\delta x)\in B} \exp\left(-\frac{(n-1)v(\delta x)}{2\sigma^2})\right) d^{n-1}\delta x \\ &= \mathrm{P}(\ V_n \in B\ )\mathrm{P}(\ \bar{X}_n \in C\ )\,.\end{aligned}$$

ここで，$v(\delta x) = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^n (x_i-\bar{x})^2$ が $\bar{x}$ によらない，$\delta x$ の 2 次形式であることを用いて結合分布を二つの積分の積に書いた.

## 14.2 区間推定

点推定（$\bar{X}_n, V_n$）は平均及び分散の目安としては優れている．しかし，ちょうどその値が本当の値とは思えない．区間で推定するのがより自然.

[林周二, 13.5–8], [小針, §8.2], 母集団の平均を $\mu$ 分散を $\sigma^2$ とおく．さらに，$\sigma^2$ **の値が有限で，分かっている** と仮定する.

**注**. $\sigma^2$ が分からない場合の平均の推定は §14.4.5, §14.4.5.

$\mu$ のみが未知の時，標本値 $x_i = X_i(\omega)$, $i = 1, 2, \cdots, n$, から推定する.

サンプル $\{X_i\}$ は仮定により独立同分布確率変数列で，$\mathrm{E}[\ X_1\ ] = \mu$, $\mathrm{V}[\ X_1\ ] = \sigma^2$ である．定理 37 の仮定が満たされるので

**中心極限定理 (定理 37)**　$Z_n = \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^n (X_i-\mu)$ の分布は $N_{0,\sigma^2}$ に弱収束する.

中心極限定理より，サンプルサイズ $n$ が十分大きければ標本平均 $\bar{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^n X_i$ は $N(\mu, \sigma^2/n)$ にほぼ従う．特に，母集団が正規分布 $N_{\mu,\sigma^2}$ に従うことが分かっていれば $n$ が小さくても，命題 53 より，正確に $N(\mu, \sigma^2/n)$ に従う.

従って，(24) より，

$$\begin{aligned}
(26)\quad & \mathrm{P}(\ \mu - z'\frac{\sigma}{\sqrt{n}} < \bar{X}_n < \mu + z\frac{\sigma}{\sqrt{n}}\ ) \\
&= \frac{1}{\sqrt{2\pi\sigma^2/n}} \int_{\mu - z'\frac{\sigma}{\sqrt{n}} < x < \mu + z\frac{\sigma}{\sqrt{n}}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2/n}\right) dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-z'}^{z} \exp\left(-\frac{y^2}{2}\right) dy \\
&= F(z) - F(-z') .
\end{aligned}$$

最後の $F$ は標準正規分布の分布関数 (25).

**考察**　(26) より次の考察を得る.

(i) 連続分布なので，１点をとる確率は 0．よって，実測値による標本平均 $\bar{X}_n(\omega)$ が $\mu$ に等しい確率は 0.

(ii) 正規分布の密度は全ての $x$ で正である．よって，確率 1 で言えることは，$\bar{X}_n(\omega)$ はどれかの実数値になるということだけ.

(iii) よって，意味のある結論があるとすれば，それは，標本平均 $\bar{X}_n$ が区間 $(\mu - z'\frac{\sigma}{\sqrt{n}}, \mu + z\frac{\sigma}{\sqrt{n}})$ に入る確率が $F(z) - F(z')$ である，ということである.

**定義 7** 母集団の分布 $P_\theta$ が未知の *parameter* $\theta$ を含むとする．ある $\alpha > 0$ を固定するとき，$\alpha = P_\theta(x \notin A_\theta)$ となる *parameter dependent* な事象 $A_\theta$ を決めておく.

$x$ を，実測したサンプルとする.

*Parameter* の集合 $\Theta = \{\theta \mid x \in A_\theta\}$ を $\theta$ の **信頼区間** とよび，$\alpha$ を **危険率**, $1-\alpha$ を信頼率, $(1-\alpha)\times 100\%$ を信頼係数，と呼ぶ *[小針, §8.1]*.

信頼区間でもって parameter の推定値としようというのである.

典型的には危険率を $\alpha = 0.05$ あるいは 0.01 として，$F(z) - F(-z) = 1 - \alpha$ となる $z > 0$ を選ぶ. $F(-\infty) = 0$, $F(\infty) = 1$, $F(-z) = 1 - F(z)$ に注意して，$z = 1.960$ あるいは 2.575．一般論の $A_\mu$ は $(\mu - z\frac{\sigma}{\sqrt{n}}, \mu + z\frac{\sigma}{\sqrt{n}})$ とおいてよい．$\bar{x} = \bar{X}_n(\omega)$ が（観測によって）与えられたとき信頼区間

$$(27)\quad \Theta = \{\mu \mid \bar{x} \in A_\mu\} = (\bar{x} - z\frac{\sigma}{\sqrt{n}}, \bar{x} + z\frac{\sigma}{\sqrt{n}})$$

をもって $\mu$ の推定とする.

**例題－平均値の推定**　硬貨を 400 回投げて表 220 回，裏 180 回出た．この硬貨の表の出る確率を推定せよ.

表の出る確率を $p$ とする．サンプル（試行）の独立性は仮定する．平均 $\mu = p$，分散 $\sigma^2 = (-p)^2 \times (1-p) + (1-p)^2 \times p = p(1-p)$.

（$k$ 回表が出る確率は $n = 400$ として２項分布 $B_{n,p}$. §9.1.4 より，$n$ が十分大きければ，平均 $\bar{x}$ は正規分布 $N_{p,p(1-p)/n}$ に近い，と考えても良い.）

分散の $p$ は $\bar{x}$ で置き換える.

**注**. 分散が分からない場合の正しい平均の推定は §14.4.5, §14.4.5.

信頼区間は (27) より，$\bar{x} = 220/400 = 0.55$, $\frac{\sigma}{\sqrt{n}} = \sqrt{p(1-p)/n} = \sqrt{0.55 \times 0.45/400} = 0.02487$ （母集団分散を不偏分散で置き換える）なので，危険率 $\alpha = 0.01$ あるいは 0.05 に対して，$\mu = 0.55 \pm 0.064 = (0.486, 0.614)$ あるいは $\mu = 0.55 \pm 0.049 = (0.501, 0.599)$ となる.

$\bar{x}$ が等しくても，危険率 $\alpha$，sample size $n$，によって信頼区間は異なる．$\bar{x} = 0.55$ の場合.

| $n$ | $\alpha = 0.05$<br>$z = 1.960$ | $0.01$<br>$2.575$ |
|---|---|---|
| 100 | $(0.453, 0.647)$ | $(0.422, 0.678)$ |
| 200 | $(0.481, 0.619)$ | $(0.459, 0.641)$ |
| 300 | $(0.494, 0.606)$ | $(0.476, 0.624)$ |
| 400 | $(0.501, 0.599)$ | $(0.486, 0.614)$ |

**注**. (i) [小針, pp.199–203] に基本的な練習問題が多数ある．

(ii) $n$ 大なほど幅は狭い．標本を集めることの意味が分かりやすい．幅は $\sqrt{n}$ に反比例する．俗称で統計誤差と呼ばれているもの．

(iii) 危険率 $\alpha$: 本当は正しいのにうっかりしていて棄却してしまう確率 §18.2．$\mu$ が正しい値のときに，確率 $\alpha$ で $\mu$ が信頼区間の外に出る．危険率が小さいほど信頼区間は広がる（危険を冒さないので主張が弱くなる）．

(iv) 母集団のパラメータ $\mu$ が分かればサンプル値 $\bar{x}$ についての**確率**が計算できる．しかし，区間推定はサンプル値 $\bar{x}$ が一つ与えられて未知の母集団 $\mu$ について結論を出そうという．候補となる分布の集合（この場合は正規分布）の一つ一つについて，それが母集団だった場合にサンプル値があまり珍しくないならば，その分布を母集団の候補として残す，というのが信頼区間であり，サンプル値の珍しさの許容限界を示すのが危険率である．

**母集団を知らないのに確率は計算できない．**ここで出した区間推定の区間や危険率は **確率ではない**.

(v) 実際には分散を知らない．そこで不偏分散 $V_n$ (21) をもって $\sigma^2 = \mathrm{V}[\ X\ ]$ の代用とする．$n$ が十分大きければ (23) よりこの代用は正当化される．そこで，十分大きなサンプルを用意することが求められる（大標本理論）．§14.4.10 参照．次の検定でも同様．分散が分からない場合の正しい平均の推定は §14.4.5, §14.4.5.

## 14.3 検定

— 95% で棄却（否定）できない場合，たとえば 4.5cm の金魚を示されたときには何が言えるだろうか．

これは金魚です

という以外，何も言えないのである．相手が，さらに何か言うのを待っている様子なら,「美しいですね」とか何とか言って，その場をしのぐしか仕方ないのである．つまり，

「95% の確からしさで，これがA試験場のものでないと言える」わけではない

のである．否定できるときは意味があるが，否定できないときは，何も分からないのである．だから帰無仮説という名もついているのである．— 「返さないとは言わない」は「返します」と同義ではないのである．[4]

**検定と推定**　予想があるとき **検定**，ないとき **推定**．こんにちの理論では両者に殆ど差を考えない．既に注意したように，推定といえども母集団について暗黙の予想または仮定がある．古典的には，推定から自然に出てきた概念と検定から自然に出てきた概念が今日，用いられる．古典的な意味で解説する．

**定義 8** 母集団の分布 $P_\theta$ が未知の *parameter* $\theta$ を含むとする．ある $\alpha > 0$ を固定するとき，$\alpha = P_\theta(x \notin A_\theta)$ となる *parameter dependent* な事象 $A_\theta$ を決めておく．$A_\theta$ の補集合を **危険域（棄却域）** と呼ぶ.

仮説 $H$:「母集団の分布は $P_\theta$ である」をたてて実測し，サンプル $x$ を得たとする．$x \notin A_\theta$ ならば $H$ を棄却する．即ち,「母集団の分布は $P_\theta$ でない」と結論する．$x \in A_\theta$ ならば $H$ を棄却しない．即ち,「母集団の分布は $P_\theta$ でないとは言えない」と結論する．この手続きを $H$ の検定と呼び，仮説 $H$ を**帰無仮説** *(null hypothesis)* と呼ぶ.

[4] [小針, §8.1].

$\alpha$ を **危険率** または **有意水準**, $1-\alpha$ を信頼率, $(1-\alpha)\times 100\%$ を信頼係数, と呼ぶ *[小針, §8.1], [林周二, 13.3]*.

**注**. 主張したいことと逆のことを H に立てる．即ち，既存の常識または理論を H に立てて，実測によって常識を覆す，という使い方をする（現象の観測）．あるいはまた，問題のない正常状態を H とし，実測によって否定されれば異常が起こったと見て対策を立てる，という使い方をする（抜き取り検査）．

危険率 $\alpha$ は H が正しいのにこれを棄却する確率である．$\alpha$ は通常 1% や 5% にとるので，通常の使い方では H が正しいのにこれを棄却する確率は小さい．従って，H が検定で棄却されれば，それは強力な根拠がある．他方，H が間違っていても（例えば $\theta$ がごくわずかずれている場合），H を棄却できない可能性は高い場合がある．従って，厳密に言えば，H が棄却されない場合は何も言えない．あるいは，H が正しいという根拠が統計検査に内在してはいない (§18.2) [小針, §8.2]．棄却される場合のみ検査が独自性を発揮するという意味で，帰無仮説と呼ばれる．

**例題ー平均値の差の検定**　問 [林周二, 13.8]：二人の研究者がそれぞれある測定を行い，データ平均 $\bar{X}_n(\omega) = 175.18$, $\bar{Y}_n(\omega) = 174.29$, を得た．この二系統の測定結果は等しいと言えるか？有意水準 5% で検定せよ．実験誤差 $\sigma/\sqrt{n}$ はどちらも $1/\sqrt{10}$ であるとみてよいことが分かっているとする．

解：両者の母平均をそれぞれ $\mu, \mu'$ とする．

帰無仮説 H: $\mu = \mu'$.

$\bar{X}_n, \bar{Y}_n$ はそれぞれ $N(\mu, \sigma^2/n)$ , $N(\mu', \sigma^2/n)$ に従うので，$\bar{X}_n - \bar{Y}_n$ は §9.1.2 より，$N(\mu-\mu', 2\sigma^2/n)$ に従う．仮説 H の下で，$N(0, 0.2)$ に従うはずだから，$Z = (\bar{X}_n - \bar{Y}_n)/\sqrt{0.2}$ は $N(0,1)$ に従う．

§14.2 と同様に，$\alpha = 0.05$ に対して，$F(z) - F(-z) = 1-\alpha$ となる $z>0$ を選ぶと，$z = 1.960$. 即ち，$Z$ の信頼区間 $A = (-1.960, 1.960)$. 実測値は $Z(\omega) = (175.18 - 174.29)/\sqrt{0.2} = 1.77$. よって，H は棄却できない．二系統の測定結果に（同一対象の測定であるとして）矛盾はない．

**推定との比較**　硬貨を 400 回投げて表 220 回，裏 180 回出た．この硬貨は公平か？推定 §14.2 と同様に，$(\bar{X}_n - \mu)\sqrt{n}/\sigma$ の分布は $N(0,1)$ に近い．(二項分布で厳密にやっても良いが，数値的には殆ど変わらない．) $\alpha = 0.05, 0.01$ に対して，$F(z) - F(-z) = 1-\alpha$ となる $z>0$ を選ぶと，$z = 1.960, 2.575$. よって信頼区間（危険域の補集合） $A_\mu = (\mu - z\dfrac{\sigma}{\sqrt{n}}, \mu + z\dfrac{\sigma}{\sqrt{n}})$ とおいて，$\bar{X}_n(\omega) \notin A_\mu$ なら H を棄却する．

帰無仮説 H: 公平である．即ち，$\mu = p = 0.5$, $\sigma^2 = p(1-p) = 0.25$.

$A = (0.451, 0.549), (0.436, 0.564)$.

他方サンプル値は $220/400 = 0.55$. 即ち，信頼係数 99% ではこの硬貨は不公平とは言えず，信頼係数 95% ならばこの硬貨は不公平と結論する．

**注**. §14.2 と分散の取り扱いが異なる．$n$ が十分大きければ大数の法則でどちらも同じことになるので，大標本理論と呼ばれる．この点の不備を補正するのが §14.4.5, §14.4.5 の目標.

**危険域の設定の仕方**　上の例題に，もし，前提条件「この硬貨は表が出やすい細工があるという情報が入った．そこで硬貨を投げて検査した.」が入ったらどうなるか？表が出にくいケースを考える必要はない．すると，信頼区間を $(a,b)$ の形にとるのは不自然．当然，裏が多めに出たデータの場合は「細工の事実なし」と結論すべき．同じ信頼係数 $\alpha$ （つまり硬貨は公平なのに公平でないとする危険）ならば，表が多めに出た場合に集中させた方が細工の検出が起こりやすい．

（分布の両側を排除（危険域）した図と分布の片側を排除した図で，平均が大きい側にずれた分布を重ねて，どちらが危険域内の面積が広いかを比べると分かる）．

**$\alpha$ を固定しても信頼区間 $A_\theta$ の選び方には任意性がある．誤って H を捨てる危険は等しくても，誤って H を残す危険が変化しうる．** $F(z) = \mathrm{P}(\ X \le z\ ) = 1-\alpha$ となる $z$ を探すことになる．§14.1 の表に戻ると，$\alpha = 0.05, 0.01$ に対して，$z = 1.645, 2.327$.

信頼区間 $A_\mu = (-\infty, \mu + z\dfrac{\sigma}{\sqrt{n}})$ とおいて，$\bar{X}_n(\omega) \notin A_\mu$ なら H を棄却する．

帰無仮説 H: 公平である．即ち，$\mu = p = 0.5$, $\sigma^2 = p(1-p) = 0.25$.

$A = (-\infty, 0.541)$, $(-\infty, 0.548)$.

他方サンプル値は $220/400 = 0.55$．即ち，信頼係数 95% はもちろん 99% でもこの硬貨は不公平と結論する．不公平だという結論が強化されたのは「表が出やすい細工」という情報が追加されたからであるとも言える．

**注**. 危険域を素朴な直感で選んだ．この選択の数学的一般論は §18.2 を見よ．

## 14.4　カイ平方分布 14

[小針, §8.3, §7.3–4] [林周二, 15.1], [楠岡 2, p.212] 基準正規分布 $N(0,1)$ に従う独立な確率変数 $X_1$, $X_2$, $\cdots$ , $X_n$ の平方和

$$Y = \sum_{i=1}^{n} X_i^2$$

が従う分布を，自由度 $n$ のカイ平方分布（$\chi^2$ 分布）といい，${\chi^2}_n$ と書く．

**カイ平方分布の具体形**　定義より[5]，

$$\mathrm{P}(\ \sum_{i=1}^{n} X_i^2 \le a\ ) = \int_{\sum_{i=1}^n x_i^2 \le a} e^{-\frac{1}{2}\sum_{i=1}^n x_i^2} \frac{d^n x}{\sqrt{2\pi}^n}\,.$$

$F_n(z) = \displaystyle\int_{0 \le \sum_{i=1}^n x_i^2 \le z} \frac{d^n x}{\sqrt{2\pi}^n}$ とおくと（部分積分と Fubini より），

$$\begin{aligned}\int_0^a e^{-y/2} F_n'(y)\,dy &= e^{-a/2}F_n(a) + \frac{1}{2}\int_0^\infty e^{-y/2}F_n(y)\,dy = \int_{\sum_{i=1}^n x_i^2 \le a} e^{-\frac{1}{2}\sum_{i=1}^n x_i^2} \frac{d^n x}{\sqrt{2\pi}^n} \\ &= \mathrm{P}(\ \sum_{i=1}^{n} X_i^2 \le a\ )\,.\end{aligned}$$

$F_n$ の具体形を求める．$A_n = \displaystyle\int_{\sum_{i=1}^n x_i^2 \le 1} \frac{d^n x}{\sqrt{2\pi}^n}$ とおくと，$A_n$ は $z$ によらない定数で，$F_n(z) = A_n\, z^{n/2}$. 他方，

$$A_n\Gamma(\frac{n}{2}+1) = \int_0^\infty A_n z^{n/2} e^{-z}\,dz = \int_{\mathbf{R^n}} \left( \int_{\sum_{i=1}^n x_i^2}^\infty e^{-z} \right) \frac{d^n x}{\sqrt{2\pi}^n} = \left( \int_{\mathbf{R}} e^{-x^2} \frac{dx}{\sqrt{2\pi}} \right)^n = 2^{-n/2}\,.$$

よって，$F_n(z) = \left(\dfrac{z}{2}\right)^{n/2} \Gamma(\dfrac{n}{2}+1)$．これまでの式に代入して，$\Gamma(x+1) = x\Gamma(x)$ を用いると，

$$\mathrm{P}(\ \sum_{i=1}^{n} X_i^2 \le a\ ) = \frac{1}{\left(\frac{z}{2}\right)^{n/2}\Gamma(n/2)} \int_0^a e^{-y/2} y^{n/2-1}\,dy\,.$$

分布の密度 $f_n$ は

(28)　$$f_n(x) = \frac{1}{2^{n/2}\Gamma(n/2)}\, x^{n/2-1} e^{-x/2}\,.$$

$X$ が自由度 $n$ のカイ平方分布に従うとき，$\alpha = \mathrm{P}(\ X > a\ ) = \displaystyle\int_a^\infty f_n(x)\,dx$ となる $a = a(n, \alpha)$ の表を掲げる [小針, pp.287], [林周二, 15.1].

[5] 以下の変形の意味はデルタ関数を使えば明らかになる．

| $n \setminus \alpha$ | 0.99 | 0.95 | 0.05 | 0.01 |
|---|---|---|---|---|
| 1 | 0.00 | 0.00 | 3.84 | 6.63 |
| 2 | 0.02 | 0.10 | 5.99 | 9.21 |
| 3 | 0.12 | 0.35 | 7.81 | 11.34 |
| 4 | 0.30 | 0.71 | 9.49 | 13.28 |
| 5 | 0.55 | 1.15 | 11.07 | 15.09 |
| 6 | 0.87 | 1.64 | 12.59 | 16.81 |
| 7 | 1.24 | 2.17 | 14.07 | 18.48 |
| 8 | 1.65 | 2.73 | 15.51 | 20.09 |
| 9 | 2.09 | 3.33 | 16.92 | 21.67 |
| 10 | 2.56 | 3.94 | 18.31 | 23.21 |
| 12 | 3.57 | 5.23 | 21.03 | 26.22 |
| 14 | 4.66 | 6.57 | 23.68 | 29.14 |
| 16 | 5.81 | 7.96 | 26.30 | 32.00 |
| 18 | 7.01 | 9.39 | 28.87 | 34.81 |
| 20 | 8.26 | 10.85 | 31.41 | 37.57 |
| 50 | 29.7 | 34.76 | 67.50 | 76.15 |
| 100 | 70.1 | 77.93 | 124.34 | 135.81 |

**$\chi^2$ 分布の平均と分散**　具体形と，ガンマ関数の性質から，以下が直ちに求まる．

**平均** $m = n$.

**分散** $v = 2n$.

**分散の分布**　定義から明らかに，一般の正規分布 $N(\mu, \sigma^2)$ に従う独立な確率変数 $X_1, X_2, \cdots, X_n$ に対して，$\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(X_i - \mu)^2$ は，自由度 $n$ のカイ平方分布に従う．

**不偏分散の分布**　$\bar{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$ に対して，$V_n = \dfrac{1}{n-1}\displaystyle\sum_{j=1}^{n}(X_j - \bar{X}_n)^2$ の分布．

$$W_j = \frac{X_j - \mu}{\sigma}\,,\quad \bar{W}_n = \frac{1}{n}\sum_{j=1}^{n} W_j\,,$$

とおくと，$W_j$ の分布は $N(0,1)$ で，

$$\begin{aligned} &\mathrm{P}(\ \frac{n-1}{\sigma^2}V_n \in A\ ) = \mathrm{P}(\ \sum_{j=1}^{n}(W_j - \bar{W}_n)^2 \in A\ ) = \int_{w^T S\, w \in A} \exp(-\frac{1}{2}\sum_{j=1}^{n} w_j^2)\, \frac{d^n w}{\sqrt{2\pi}^n} \\ &= \int_{v^T D\, v \in A} \exp(-\frac{1}{2} v^T\, v)\, \frac{d^n v}{\sqrt{2\pi}^n}\,. \end{aligned}$$

ここで $w^T S\, w = \displaystyle\sum_{j=1}^{n}(w_j - \bar{w}_n)^2$, 即ち，$S_{ij} = S_{ji} = \delta_{ij} - 1/n$, $OSO^T = D$, $OO^T = 1$, $v = O\, w$. $S$ の固有値は $\displaystyle\sum_j S_{ij}O_{kj} = d_k O_{ki}$ を解くと，$d_k = 0$ $O_{ki} = O_k$, or $d_k = 1$, $\displaystyle\sum_j O_{kj} = 0$. 即ち，$D = \mathrm{diag}(1, 1, \cdots, 1, 0)$, $O^T = (v^{(1)}, \cdots, v^{(n-1)}, \mathbf{1})$. $v_n$ については全範囲積分なので，

$$\mathrm{P}(\ \frac{n-1}{\sigma^2}V_n \in A\ ) = \int_{\sum_{i=1}^{n-1} v_i^2 \in A} \exp(-\frac{1}{2}\sum_{i=1}^{n-1} v_i^2)\, \frac{d^{n-1} v}{\sqrt{2\pi}^{n-1}}\,.$$

よって，$\frac{n-1}{\sigma^2}V_n$ の従う分布は自由度 $n-1$ の $\chi^2$ 分布．即ち，不偏分散 $V_n$ は自由度が見かけ上一つ少ない．これは平均を $\bar{X}$ で推定したために，自由度が一つ減ったと理解すると覚えやすい．

#### 14.4.1 分散の推定，検定

[林周二, 15.2], [小針, §8.3]. **カイ平方分布を用いる際は，母集団に正規分布（または，正規分布がよい近似であること）を仮定する必要がある．** $n$ を大きくしても，カイ平方分布への中心極限定理があるのではない．

比較の対象の分散に関して正確な知識がある場合は，データから求めた分散の和との比をカイ平方分布に当てはめることにより，検定ができる．例えば古い作業方法による製品の分散データが多く，その平均が大数の法則で母分散と思える場合など．母分散 $\sigma^2$ が変わっていなければ，新しい方法による製品のデータの不偏分散 $V_n$ に対して，$(n-1)V_n/\sigma^2$ 倍がカイ平方分布に従うはず．その手続きは §14.3 の場合と同様である．カイ平方分布の数表の使い方は推定の場合と同様なので，以下は推定の例のみ挙げる．

**例** 19960117 スーパータイセーで買った地卵 10 個 pack は，5 個が茶色，5 個が白色であり，重さ（グラム）は茶: 61, 62, 64, 64, 68, 白: 58, 63, 64, 66, 67, であった．母集団が正規分布であり，茶色と白に区別がないとして，この卵の母分散 $\sigma^2$ を信頼係数 90% で求めよ．標本平均 63.7，不偏分散 $\times(n-1) = 2.7^2+1.7^2+0.3^2\times 3+4.3^2+5.7^2+0.7^2+2.3^2+3.3^2 = 78.1$．自由度 $n = 10-1 = 9$．両側に 5% ずつの危険域をとることにして，$\alpha = 95\%, 5\%$ に対応するカイ平方の値は §14.4 の表より，それぞれ 3.33, 16.92 だから，信頼区間は $3.33 < 78.1/\sigma^2 < 16.92$．即ち，$4.62 < \sigma^2 < 23.45$, あるいは，$2.15 < \sigma < 4.84$.

**注**. 平均は §14.4.5 で求める

#### 14.4.2 適合度の検定

**度数分布** 母集団が密度 $f$ を持つ連続分布のときを考え，サイズ $N$ の標本 $X_1, \cdots, X_N$, を考える．標本値（データ）は有限 $(N)$ 個の数値なので，標本値の分布（経験分布の標本値）は密度を持たない．密度の近似として，実数空間を離散化する，度数分布という現実的な処方がある．

実数空間を有限個 ($A$ 個) の区間 (bin) に分ける．通常両端を除き等間隔とする．両端については，データの少ないところをそれぞれまとめて，それぞれ一つの長い区間にすることが多い．各 bin に，その区間に値が入るデータの個数を対応させる写像を**度数分布** と呼ぶ．

$a \in \{1, 2, \cdots, A\}$ で bin を label する．Bin $a$ を区間 $(l_a, u_a]$ とすると，この bin にデータが入る確率 $p_a = \mathrm{P}(\ X_1 \in (l_a, u_a]\ ) = \displaystyle\int_{l_a}^{u_a} f(y)\,dy$．この bin に入るデータの個数を $Z_{N,a} = \sharp\{i \mid X_i \in (l_a, u_a]\}$ とおくと，$\{X_i\}$ の独立性から，$\{Z_{N,a} \mid a = 1, 2, \cdots, A\}$ は項数 $A$ の多項分布に従う．即ち，$\displaystyle\sum_{a=1}^{A} k_a = N$ に拘束された系であって，

$$\mathrm{P}(\ Z_{N,a} = k_a\ ,\ a = 1, 2, \cdots, A\ ) = \frac{N!}{\displaystyle\prod_{a=1}^{A} k_a!} \prod_{a=1}^{A} {p_a}^{k_a}\ .$$

**定理 65** $X = \displaystyle\sum_{a=1}^{A} \frac{(Z_{N,a} - Np_a)^2}{Np_a}$ とおくと，各 $Np_a$ が十分大きければ，$X$ の分布は自由度 $A-1$ のカイ平方分布に近づく．

**注**. Bin のうち一つの $Z$ は残りの bin の $Z$ で決まってしまうので，自由度が一つ減る，と覚えることができる．

**証明**. 既に述べたことより,

$$\mathrm{P}(\ X \le x\ ) = \sum_{\substack{\sum_a k_a = N, \\ \sum_a \frac{1}{Np_a}(k_a - Np_a)^2 \le x}} \frac{N!}{\prod_a k_a!} \prod_{a=1}^{A} {p_a}^{k_a} \,.$$

各 $Np_a$ が十分大きければ,そのまわりで,$z_a = \sqrt{\dfrac{N}{p_a}}(\dfrac{k_a}{N} - p_a)$ とおいて展開できる.拘束条件 $\displaystyle\sum_{a=1}^{A} k_a = N$ は $\displaystyle\sum_{a=1}^{A} z_a\sqrt{p_a} = 0$ になり,$\triangle k_a = 1$ に対応するのは $\triangle z = \frac{1}{\sqrt{Np_a}}$ である.$\log k_a! \approx k_a(\log k_a) - k_a + \dfrac{1}{2}\log(2\pi k_a)$(と $N!$ についての同様の式)を用いるいつもの手で,$f = \log N - \displaystyle\sum_a k_a/N \log(k_a/p_a)$ とおくと ($\displaystyle\sum_a p_a = 1$ に注意),

$$\mathrm{P}(\ X \le x\ ) \approx \int_{\sum_a {z_a}^2 \le x} N^{-1/2}\delta(\sum_a z_a\sqrt{p_a})\prod_a \sqrt{Np_a} d^A z e^{Nf} \frac{\sqrt{2\pi N}}{\prod_a \sqrt{2\pi k_a}} \,.$$

最後の因子については(これもいつもの手で) $k_a \approx Np_a$ を使うと,

$$\mathrm{P}(\ X \le x\ ) \approx \int_{\sum_a {z_a}^2 \le x} \delta(\sum_a z_a\sqrt{p_a}) d^A z e^{Nf} (2\pi)^{-A+1} \,.$$

ここで,

$$\begin{aligned} f &= \log N - \sum_a k_a/N \log(k_a/p_a) = -\sum_a k_a/N \log(k_a/(Np_a)) \\ &= -\sum_a (p_a + z_a\sqrt{\frac{p_a}{N}}) \log(1 + \frac{z_a}{\sqrt{Np_a}}) \approx -\sum_a (z_a\sqrt{\frac{p_a}{N}} + \frac{{z_a}^2}{2N}) \\ &= -\sum_a \frac{{z_a}^2}{2N} \,. \end{aligned}$$

最後の変形で拘束条件 $\displaystyle\sum_a z_a\sqrt{p_a} = 0$ を用いた.よって,

$$\begin{aligned} &\mathrm{P}(\ X \le x\ ) \approx \int_{\sum_a {z_a}^2 \le x} \delta(\sum_a z_a\sqrt{p_a}) d^A z e^{-\sum_a {z_a}^2/2} (2\pi)^{-A+1} \\ &= \int_{y \le a} e^{-y/2}\, dy \left( \int d^A z \delta(y - \sum_a {z_a}^2) \delta(\sum_a z_a\sqrt{p_a}) (2\pi)^{-A+1} \right) . \end{aligned}$$

次元を数えれば(デルタ関数は $-1$ 次元なので)右辺括弧内は $y^{(A-3)/2} \times (y$ によらない定数$)$ に等しい.よって

$$\mathrm{P}(\ X \le x\ ) \approx \mathrm{const} \times \int_{y \le a} y^{(A-1)/2-1} e^{-y/2}\, dy \,.$$

**Bin の選び方** ([林周二, 15.3] とは最後の項目を除き必ずしも一致しない).

(i) $Np_a$ 大 $(\ge 4)$ (bin 数 $A$ の制限)
(ii) $\displaystyle\inf_{a,a'} p_a/p_{a'} \approx 1$
(iii) カイ平方の値が小さすぎたら一応疑う.

**適合度** [林周二, 15.3]. Bin $a$ に $Z_{N,a}(\omega)=g_a$ 件のデータが入ったとする．この bin にデータが入る期待値は $g_a^*=Np_a=N\int_{l_a}^{u_a}f(y)\,dy$ であった．定理 65 から $\chi^2=\sum_{a=1}^{A}\frac{(g_a-g_a^*)^2}{g_a^*}$ が自由度 $A-1$ のカイ平方分布に従う $X=\sum_{a=1}^{A}\frac{(Z_{N,a}-g_a^*)^2}{g_a^*}$ の信頼区間にはいるかどうかで，適合度を検定できる．

**注**. (i) 理論分布は正規分布でなくても良い．データが bin に入る多項分布の分散和が極限でカイ平方分布に従うから．

(ii) 元々離散的な分布で各 bin に十分度数があれば，改めて度数分布を取り直すまでもなく，カイ平方分布が適用できる [小針, §8.3, 例題 5]．一般に実数空間でなくても $\Omega$ の可測な partition があれば，それを bin としてカイ平方分布が適用できる §14.4.4.

(iii) 理論分布のパラメータをいくつかデータから点推定した場合は，自由度がそれだけ減る，と考える．例えば，正規分布への当てはめの場合は，平均と分散の二つ減って，自由度 $n=A-3$ となる [林周二, 15.3].

### 14.4.3 因子 ，因子の独立性，集合族の生成する $\sigma$-加法族

[6]

**定義 9** *(i)* 確率空間 $(\Omega,\mathcal{F},P)$ における因子とは $\Omega$ の可測な有限 *partition*, 即ち，ある自然数 $m$ に対して $\Omega=\sum_{i=1}^{m}A_i$ を満たすような $\mathcal{F}$ の部分集合族 $\{A_1,\ A_2,\ \cdots,\ A_m\}$ のことである．

*(ii)* 因子 $\mathcal{A}=\{A_1,\ A_2,\ \cdots,\ A_m\}$ と因子 $\mathcal{B}=\{B_1,\ B_2,\ \cdots,\ B_n\}$ が独立とは，任意の $(i,j)\in\{1,2,\cdots,m\}\times\{1,2,\cdots,n\}$ に対して $A_i$ と $B_j$ が独立なことを言う．

命題 9 で定義した $\sigma(A)$ の拡張として，集合族の生成する $\sigma$-加法族を定義する．

**定義 10** $\Omega$ の部分集合よりなる集合族 $\mathcal{A}=\{A_i\mid i\in\Lambda\}$ に対して，$\mathcal{A}$ を含む最小の $\sigma$-加法族を，$\mathcal{A}$ が生成する $\sigma$-加法族と呼び，$\sigma(\mathcal{A})$ と書く．

**命題 66** *(i)* 因子 $\mathcal{A}=\{A_1,\ A_2,\ \cdots,\ A_m\}$ に対して，$\sigma(\mathcal{A})$ は $\mathcal{A}$ の中の元よりなる可能なあらゆる直和よりなる集合族である．

*(ii)* 因子 $\mathcal{A}$ と $\mathcal{B}$ が独立なことと，$\sigma$-加法族 $\sigma(\mathcal{A})$ と $\sigma(\mathcal{B})$ が独立なことは同値である．

証明は順にやればいずれも明らか.

**定義 11** 確率空間 $(\Omega,\mathcal{F},P)$ と $\sigma$-加法族 $\mathcal{G}\subset\mathcal{F}$ が与えられたとき，$P$ の $\mathcal{G}$ への制限は確率測度になる．これを $P$ の $\mathcal{G}$ における周辺分布と呼ぶ.

### 14.4.4 独立性の検定

[小針, §7.7, 命題 7.11; §8.3, 例題 6], [林周二, 15.4]. 二つの因子 $\mathcal{A}=\{A_i\mid i=1,2,\cdots,m\}$ と $\mathcal{B}=\{B_j\mid j=1,2,\cdots,n\}$ $(\Omega=\sum_{i=1}^{m}A_i=B=\sum_{j=1}^{n}B_j)$, が与えられている．通常，一方 $(\mathcal{B})$ が原因となるかもしれない因子で，他方 $(\mathcal{A})$ が結果の分類．$\mathcal{B}$ の違い（どの $j$ か）が本当に $\mathcal{A}$ の違いを生むかをデータから分析したい．例：味の素を食べると頭が良くなるか．$B_1=$ 飲んでる，$B_2=$ 飲まない（，$B_3=$ たまに飲む，など細かく分けられる）．$A$ は試験成績の上昇の度数分布.

$N$ 個のサンプル $X_k,\ k=1,2,\cdots,N,$ を因子で分類して度数分布表とする．即ち $(i,j)$ に $f_{ij}=\sharp\{k\mid X_k(\omega)\in A_i\cap B_j\}$ を対応させる表.

[6] 960621. 以下の数学的な定義は，既存の測度論や確率論には存在しないが，因子分析における統計学の用語の定式化を哲弥が私的に試みた.

| 現象 $\mathcal{A}\backslash$ 原因 $\mathcal{B}$ | $B_1$ | $B_2$ | $\cdots$ | $B_j$ | $\cdots$ | $B_n$ | 計 ($\mathcal{A}$ の分布) |
|---|---|---|---|---|---|---|---|
| $A_1$ | $f_{11}$ | $f_{12}$ | $\cdots$ | $f_{1j}$ | $\cdots$ | $f_{1n}$ | $a_1$ |
| $A_2$ | $f_{21}$ | $f_{22}$ | $\cdots$ | $f_{2j}$ | $\cdots$ | $f_{2n}$ | $a_2$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ |
| $A_i$ | $f_{i1}$ | $f_{i2}$ | $\cdots$ | $f_{ij}$ | $\cdots$ | $f_{in}$ | $a_i$ |
| $\vdots$ | $\vdots$ | $\vdots$ | | $\vdots$ | | $\vdots$ | $\vdots$ |
| $A_m$ | $f_{m1}$ | $f_{m2}$ | $\cdots$ | $f_{mj}$ | $\cdots$ | $f_{mn}$ | $a_m$ |
| 計 ($\mathcal{B}$ の分布) | $b_1$ | $b_2$ | $\cdots$ | $b_j$ | $\cdots$ | $b_n$ | $N$ |

**表** $f_{ij} = \sharp\{k \mid X_k(\omega) \in A_i \cap B_j\}$: $N$ 個のサンプル値 $X_k(\omega)$ のうち結果が $B_j$ で，結果との関係（独立性）が問題になっている要因が $A_i$ であるものの個数.

**表の右端** $a_i = \sharp A_i = \sum_j f_{ij}$: 事象 $A_i$ の起きた個数.

**表の下端** $b_i = \sharp B_j$: 要因 $B_i$ の起きた個数.

**$\mathcal{A}$ の例** (i) $a = 2$，$A_1 =$ 水道管が壊われた，$A_2 =$ 壊れなかった.
(ii) $a = 3$，$A_1 =$ 正解者，$A_2 =$ 不完全解答者，$A_3 =$ 無解答者.

**$\mathcal{B}$ の例** (i) $n = 2$，$B_1 =$ 水道管に保温用のテープを巻いた，$B_2 =$ 巻かなかった.
(ii) $B_1 =$ 年齢 10 代，$\cdots$，$B_n =$60 代以上.

さて，**仮に，$P$ の $\sigma(\mathcal{A})$ および $\sigma(\mathcal{B})$ における周辺分布が分かっているとする．**このとき，
帰無仮説 H:「因子 $\mathcal{A}$ と $\mathcal{B}$ は独立である.」
を検定する．仮定から，$\mathrm{P}(\ A_i \cap B_j\ ) = \mathrm{P}(\ A_i\ )\mathrm{P}(\ B_j\ )$ であり，かつ，右辺は分かっていることになっている．すると，§14.4.2 の最後の remark 群の中の一つより，$A_i \cap B_j$ を bin とする，度数分布の適合度検定を行うことができる.

**注**. 検定全てにわたって言えるが，これが唯一の検定方法というのではなく，これも検定の方法になっている，という意味である.

定理 65 から

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(f_{ij} - N\mathrm{P}(\ A_i\ )\mathrm{P}(\ B_j\ ))^2}{N\mathrm{P}(\ A_i\ )\mathrm{P}(\ B_j\ )} \tag{29}$$

が自由度 $mn - 1$ のカイ平方分布に従う.

ここで，実際には，$\sigma(\mathcal{A})$ **および** $\sigma(\mathcal{B})$ **における周辺分布は分かっていない**のが普通であることに注意．この場合，周辺分布は，表の右端，及び下端の度数分布（部分合計）を利用することになる．従って，実際には自由度がもっと減る．最後の行および列は周辺分布を作成するのに使うので，実質の自由度は $(m-1)(n-1)$ になると考えられる．よって，

**定理 67** *(29)* で $\mathrm{P}(\ A_i\ )$, $\mathrm{P}(\ B_j\ )$ をそれぞれ表の $a_i/N$, $b_j/N$ で置き換えた量

$$\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(f_{ij} - a_i\, b_j/N)^2}{a_i\, b_j/N}$$

は自由度 $(m-1)(n-1)$ のカイ平方分布に従う．$a_i$, $b_j$, $f_{ij}$ は全て個数で数えた値.

以後の検定は §14.4.2 と同様である．適合していないという結論が独立ではない（関係がある）という結論と対応する.

なお，適合度の検定と同様の注意が当てはまる．独に $a_i b_j/N$ が大きいとき（各 $f_{ij}$ が十分大きいとき）にはじめてよい近似になる.

### 14.4.5 $t$ 分布と $F$ 分布 16

[林周二, 16,17], [小針, §8.4, §7.5–6]
$Z$ が $N(0,1)$ に従い，$Y$ が $\{\chi^2\}_n$ に従う確率変数で，$Z$ と $Y$ が独立のとき，

$$T = \frac{Z}{\sqrt{\frac{Y}{n}}}$$

が従う分布を，自由度 $n$ の $T$ 分布といい，$T_n$ と書く．
$X$ が $\chi^2_m$ に従い，$Y$ が $\chi^2_n$ に従う確率変数で，$X$ と $Y$ が独立のとき，

$$F = \frac{\frac{X}{m}}{\frac{Y}{n}}$$

が従う分布を，自由度対 $(m,n)$ の $F$ 分布といい，$F^m_n$ と書く．

#### $F$ 分布の具体形

**補題 68 ([小針, §7.5, 補題１])** $X$ と $Y$ が独立で，分布の密度がそれぞれ $p$, $q$ のとき，$Z = \dfrac{aX}{bY}$ の分布の密度 $r$ は

$$r(z) = \int_{\mathbf{R}} p(bzy/a)\, q(y)\, \frac{b}{a}\, y\, dy$$

で与えられる．

**証明**.

$$\begin{aligned}
\mathrm{P}(\ X \le z\ ) &= \int \mathrm{P}(\ Z \le z,\ Y \in (y, y+dy]\ ) = \int \mathrm{P}(\ aX \le byz,\ Y \in (y, y+dy]\ ) \\
&= \int \mathrm{P}(\ aX \le byz\ )\mathrm{P}(\ Y \in (y, y+dy]\ ) \\
&= \int_{\mathbf{R}} q(y)\, dy \int_{x \le byz/a} p(x)\, dx = \int_{t \le z} \left( \int_{\mathbf{R}} p(byt/a) q(y)\, \frac{b}{a}\, y\, dy \right) dt \\
&= \int_{t \le z} r(t)\, dt
\end{aligned}$$

途中で変数変換 $ax = byt$ を行った．

補題 68 において，$a = 1/m$, $b = 1/n$, また，$p$, $q$ に (28) を代入．整理して，公式

$$\int_0^\infty t^{x-1} e^{-t} dt = \Gamma(x)$$

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

を用いる．ここで

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1} dx = 2\int_0^{\pi/2} \cos^{2a-1}\theta\, \sin^{2b-1}\theta\, d\theta$$

はベータ関数．計算を実行すると直ちに，$F^m_n$ の密度 $g^m_n$ は

$$(30) \quad g^m_n(z) = \chi(z \ge 0)\, \frac{m^{m/2} n^{n/2}}{B(m/2, n/2)}\, \frac{z^{m/2-1}}{(mz+n)^{(m+n)/2}}$$

と決まる．

**$T$ 分布の具体形**　$Z$ が $N(0,1)$ に従うならば，$Z^2$ は $\chi^2_1$ に従うので，$T$ が $T_n$ に従うならば，$T^2$ は $F^1_n$ に従う．よって，$T$ は $F^m_n$ の特別な場合（の平方根）として得られる．

**補題 69** $Z$ の分布の密度 $r$ が左右対称 ($\mathrm{P}(\ Z > z\ ) = \mathrm{P}(\ Z < -z\ )$) で，$X = Z^2$ の分布の密度が $p$ のとき，$r$ は

$$r(t) = p(t^2)|t|$$

で与えられる．

**証明**. 対称性の仮定より $t \geq 0$ の場合のみ証明すればよい．

$$\mathrm{P}(\ 0 \leq Z \leq z\ ) = \frac{1}{2}\mathrm{P}(\ -z \leq Z \leq z\ ) = \frac{1}{2}\mathrm{P}(\ X \leq z^2\ ) = \frac{1}{2}\int_0^{z^2} p(x)\,dx = \int_0^z p(t^2)t\,dt\,.$$

最後の変形で変数変換 $x = t^2$ を行った．

補題 69 と (30) より直ちに $T_n$ の密度 $f_n$ は

$$(31)\quad f_n(t) = \frac{1}{\sqrt{n}B(\frac{1}{2},\frac{n}{2})}\left(\frac{t^2}{n}+1\right)^{-(1+n)/2}$$

と求まる．

**分布関数の数表**　$T$ が $T_n$ に従うとき，$\alpha = \mathrm{P}(\ |T| > t\ ) = \left(\int_t^\infty + \int_{-\infty}^{-t}\right) f_n(x)\,dx$ となる $t = t(n,\alpha)$ の表を掲げる [小針, p.288], [林周二, 16.1]．$n \to \infty$ で正規分布の値に収束することに注意（表最下段を §14.1 の数表と比較せよ）．

| $n \setminus \alpha$ | 0.10 | 0.05 | 0.02 | 0.01 |
|---|---|---|---|---|
| 1 | 6.314 | 12.706 | 31.821 | 63.657 |
| 2 | 2.920 | 4.303 | 6.965 | 9.925 |
| 3 | 2.353 | 3.182 | 4.541 | 5.841 |
| 4 | 2.132 | 2.776 | 3.747 | 4.604 |
| 5 | 2.015 | 2.571 | 3.365 | 4.032 |
| 6 | 1.943 | 2.447 | 3.143 | 3.707 |
| 7 | 1.895 | 2.365 | 2.998 | 3.499 |
| 8 | 1.860 | 2.306 | 2.896 | 3.355 |
| 9 | 1.833 | 2.262 | 2.821 | 3.250 |
| 10 | 1.812 | 2.228 | 2.764 | 3.169 |
| 12 | 1.782 | 2.179 | 2.681 | 3.055 |
| 14 | 1.761 | 2.145 | 2.624 | 2.977 |
| 16 | 1.746 | 2.120 | 2.583 | 2.921 |
| 18 | 1.734 | 2.101 | 2.552 | 2.878 |
| 20 | 1.725 | 2.086 | 2.528 | 2.845 |
| $\infty$ | 1.6448 | 1.9600 | 2.326 | 2.5758 |

$F$ が $F^m_n$ に従うとき，$\alpha = \mathrm{P}(\ F > f\ ) = \int_f^\infty g^m_n(x)\,dx$ となる $f = f(m,n,\alpha)$ を $\alpha = 5\%$, $1\%$ について，それぞれ次の二つの表に掲げる [小針, pp.289–290], [林周二, 17.1]．$m = 1$（表第 1 桁）は $T_n$ の 2 乗，$m = 1$, $n = \infty$ は正規分布の 2 乗（$T$ の表と比較せよ）．

検定でよく用いられるのは $f \approx 4$ となる範囲．自由度が多いと無駄，少ないと帰無仮説を棄却できない．

$\alpha = 5\%$

| $n \setminus m$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | 161. | 200. | 216. | 225. | 230. | 234. | 237. |
| 2 | 18.5 | 19.0 | 19.2 | 19.2 | 19.3 | 19.3 | 19.4 |
| 3 | 10.1 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 |
| 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 |
| $\infty$ | 3.84 | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 2.01 |

$\alpha = 1\%$

| $n \setminus m$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1 | | | | | | | |
| 2 | 98.5 | 99.0 | 99.2 | 99.2 | 99.3 | 99.3 | 99.4 |
| 3 | 34.1 | 30.8 | 29.5 | 28.7 | 28.2 | 27.9 | 27.7 |
| 4 | 21.2 | 18.0 | 16.7 | 16.0 | 15.5 | 15.2 | 15.0 |
| 5 | 16.3 | 13.3 | 12.1 | 11.4 | 11.0 | 10.7 | 10.5 |
| 6 | 13.7 | 10.9 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 |
| 7 | 12.2 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 |
| 8 | 11.3 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 |
| 9 | 10.6 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 |
| 10 | 10.0 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 |
| $\infty$ | 6.63 | 4.61 | 3.78 | 3.32 | 3.02 | 2.80 | 2.64 |

**簡単な性質**

(32) $F_n^m((0,a]) = F_m^n([1/a,\infty))$

(33) $T_n([-a,a]) = F_n^1([0,a^2])$

**サンプルへの応用**　カイ平方分布の時と同様に，以下が成立する．

(i) 正規母集団 $N(\mu,\sigma^2)$ から大きさ $n$ のサンプルを選び，その標本平均 (20) を $\bar{X}_n$ とする．不偏分散を (21) $V_n$ とすると，

(34) $T = \sqrt{\dfrac{n}{V_n}}\,(\bar{X}_n - \mu)$

は $T_{n-1}$ に従う．また，$T^2$ は $F_{n-1}^1$ に従う．

なぜなら，$\dfrac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$ は $N(0,1)$ に従い (命題 53, 命題 47)，$\dfrac{(n-1)V_n}{\sigma^2}$ は $\chi_{n-1}$ に従う (§14.4 の不偏分散の分布の項) から．$V_n$ と $\bar{X}_n$ の独立性は 命題 64 [7]．

(ii) 二つの正規母集団 $N(\mu,\sigma^2)$, $N(\mu',\sigma'^2)$ からそれぞれ，大きさ $m$, $n$ のサンプルを選び，その不偏分散 (21) をそれぞれ $V_m$, $V_n'$ とすると，§14.4 より，$\frac{m-1}{\sigma^2}V_m$, $\frac{n-1}{\sigma'^2}V_n'$ の従う分布はそれぞれ $\chi_{m-1}^2$, $\chi_{n-1}^2$ で

[7] 正規分布における不偏分散と標本平均の独立性がここで重要であることは，[林周二] でも [小針] でも指摘されていないように思った．[林周二] は p.236 では $F$ に関しては言及している．

ある．よって，

$$(35)\quad \frac{\sigma'^2}{\sigma^2}\frac{V_m}{V_n}$$

の分布は $F_{n-1}^{m-1}$ である．

特に，**分散が等しい二つの正規母集団（平均は等しくなくてよい）からとった大きさ $m, n$ のサンプルの不偏分散の比，$\dfrac{V_m}{V_n}$ の分布は $F_{n-1}^{m-1}$ である．** 不偏分散の比を**分散比**と呼び，分散比に $F$ 分布を適用して得られる統計的推測を**分散分析**と呼ぶ．

具体的には 2 系統のデータ間の分散の相違の検定，及び，複数組のデータ間の平均値の差の検定のこと．

(iii) 平均値の差の検定 [林周二, 16.3]．**分散の等しい**二つの正規母集団 $N(\mu,\sigma^2)$, $N(\mu',\sigma^2)$ からそれぞれ，大きさ $m, n$ のサンプルを選び（サンプルというときは，常に全体で独立な確率変数列とする），その標本平均 (20) と不偏分散 (21) をそれぞれ $(\bar{X}_m, V_m)$, $(\bar{X}'_n, V'_n)$ とすると，

$$(36)\quad T = \frac{\bar{X}-\bar{X}'-\mu+\mu'}{\sqrt{(m-1)V_m+(n-1)V'_n}}\sqrt{\frac{mn(m+n-2)}{m+n}}$$

は $T_{m+n-2}$ に従う．

なぜなら，定理 46 より，$\dfrac{\bar{X}-\bar{X}'-\mu+\mu'}{\sigma\sqrt{\frac{m+n}{mn}}}$ は $N(0,1)$ に従い，§14.4 不偏分散の分布の項より，$\sigma^{-2}(m-1)V_m$, $\sigma^{-2}(n-1)V'_n$, はそれぞれ，$\chi^2_{m-1}$, $\chi^2_{n-1}$, に従うので，両者の和は $\chi^2_{m+n-2}$ に従うからである．

(iv) 平均値そのもの，及び，2 系統のデータの間の平均値の差の検定は $t$ 分布で行うので分散分析と呼ばないのが普通．しかし，$t$ は $F$ の特別な場合（の平方根）なので，分散分析 ($F$ 検定) でも同値な統計的推測ができる．

(v) §14.4 と同様，$t$ や $F$ による検定，推定，では正規母集団（または，正規母集団で近似できること）の仮定が必要である．

(vi) $F$ $(t)$ は §14.4 と同様，独立性や適合度の検定に使える [小針, p.218, 例題 11]．このときは母集団は問題にならない．Bin に入る確率が多（二）項分布であって，それが正規分布で近似できることを用いるからである．

### 14.4.6　例

[小針, pp.209–226] に基本的な練習問題が多数ある．ここでは §14.4 で母分散の推定を行った例を取り上げる．

19960117 スーパータイセーで買った地卵 10 個 pack は，5 個が茶色，5 個が白色であり，重さ（グラム）は茶: 61, 62, 64, 64, 68, 白: 58, 63, 64, 66, 67, であった．母集団は正規分布と仮定する．

**平均値の推定**　茶色と白に区別がないとして，この卵の母平均 $\mu$ を信頼係数 95% で求める．

既に見たように，標本平均 $\bar{X} = 63.7$，不偏分散 $V$ は $(n-1)V = 78.1$，データ数 $n = 10$，危険率 $\alpha = 5\%$．$T_{n-1}$ に従う量 (34) の付近の記述より，$T = \sqrt{\frac{n}{V}}(\bar{X}-\mu)$ は $T_{n-1}$ に従う．$t$ 分布の表より，$0.05 = \mathrm{P}(\ |T| > 2.262\ )$．即ち，95% 信頼区間は $|\sqrt{\dfrac{90}{78.1}}(63.7-\mu)| < 2.262$．$\mu = 63.7 \pm 2.1$ (95%)．§14.4 の標準偏差の推定 $2.15 < \sigma < 4.84$ (90%) と合わせて，小標本理論の例が完成した．

**$t$ 推定の言い替えとしての分散分析 ($F$ 推定)**　上記の例について，$t$ 分布の代わりに $F$ 分布を用いて母平均を推定する．$T_{n-1}$ に従う量 (34) の付近の記述より，$F = \frac{n}{V}(\bar{X}-\mu)^2$ は $F^1_{n-1}$ に従う．$F$ 分布の表で $m = 1$, $n = 9$ のとき（表の自由度の $n$ とデータ数の $n$ はずれているので注意），$0.05 = \mathrm{P}(\ F > 5.12\ )$．上記諸数値を代入すると 95% 信頼区間は $\dfrac{90}{78.1}(63.7-\mu)^2 < 5.12$．$2.262^2 = 5.12$ だから信頼区間は上記 $t$ 分布による推測と同じ結論になる．

**分散の相違の検定**　茶色の卵と白の分散の違いを信頼係数 90% で検定する．茶色と白の分布をそれぞれ，$N(\mu, \sigma^2)$, $N(\mu', {\sigma'}^2)$ とする．

帰無仮説 H: $\sigma = \sigma'$.

茶色と白の標本平均と不偏分散はそれぞれ $\bar{X} = 63.8$, $V = \frac{1}{4}(2.8^2 + 1.8^2 + 0.2^2 * 2 + 4.2^2) = 28.8/4 = 7.2$, $\bar{X}' = 63.6$, $V' = \frac{1}{4}(5.6^2 + 0.6^2 + 0.4^2 + 2.4^2 + 3.4^2) = 49.2/4 = 12.3$．分散比 $V/V' = 0.59$，データ数 $m = n = 5$，$\alpha = 5\%, 95\%$.

分散比 (35) に関する記述より，**仮説 H の下では**，$V/V'$ は $F^4_4$ に従う．$F$ 分布の表より $0.05 = \mathrm{P}(\ F^4_4 > 6.39\ )$．他方，(32) より，$\mathrm{P}(\ F^m_n > x\ ) = 1 - \mathrm{P}(\ F^m_n < x\ ) = 1 - \mathrm{P}(\ F^n_m > 1/x\ )$ よって，$0.95 = 1 - \mathrm{P}(\ F^4_4 > 6.39\ ) = \mathrm{P}(\ F^4_4 > 1/6.39\ ) = \mathrm{P}(\ F^4_4 > 0.156\ )$．よって，90% 信頼区間は $0.156 < V/V' < 6.39$．$V/V' = 0.59$ だから危険率 0.1 で H は棄却されない．

そこで，茶色と白では母分散は等しいと仮定することは，**データとは統計的に異常な判断ではない．**

**平均値の差の検定**　上記によって分散は白と茶色で等しいと仮定して，

H: $\mu = \mu'$

を 95% で検定する．

$T = \dfrac{\bar{X} - \bar{X}' - \mu + \mu'}{\sqrt{(m-1)V_m + (n-1)V'_n}} \sqrt{\dfrac{mn(m+n-2)}{m+n}}$ は (36) 付近の記述より $T_{m+n-2}$ に従う．

既に見たように，$\bar{X} = 63.8$, $\bar{X}' = 63.6$, $V = 7.2$, $V' = 12.3$, $m = n = 5$ なので，$T = 0.101$．$t$ 分布に関する表より，$0.05 = \mathrm{P}(\ |T_8| > 2.306\ )$．よって，H は棄却されない．

信頼率 90% でも $0.1 = \mathrm{P}(\ |T_8| > 1.860\ )$ なので棄却されない．即ち，平均に差があるというデータ上の根拠はない．

### 14.4.7　分散分析

２系統（組）のデータの平均の差の検定は上で，$t$ 分布を用いて行った．複数組のデータの平均の差の検定を分散分析で行うことができる．

２組ずつ対にして調べるとデータサイズが小さすぎるとき，全てのデータを有効利用するために複数組の分析を行う．

$m$ 組のデータ群がある．$1 \le i \le m$ に対して，$i$ 組のサンプルサイズを $n_i$, サンプルを $X_{ij}$, $1 \le j \le n_i$, 総サンプル数 $n = \displaystyle\sum_{i=1}^{m} n_i$，組内平均 $\bar{X}_i = \dfrac{1}{n_i}\displaystyle\sum_{j=1}^{n_i} X_{ij}$，全標本平均 $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n_i} X_{ij}$，とする．次式はいつものように証明できる．

$$S_{AR} \stackrel{\mathrm{def}}{=} \sum_i \sum_j (X_{ij} - \bar{X})^2 = \sum_i \sum_j (X_{ij} - \bar{X}_i)^2 + \sum_i n_i(\bar{X}_i - \bar{X})^2 \stackrel{\mathrm{def}}{=} S_{R(A)} + S_A,.$$

$S_{AR}$, $S_{R(A)}$, $S_A$ はそれぞれ全変動，組内変動の和，組間変動と呼ばれる．それぞれに対応する自由度は，標本平均の分だけ自由度が減ることから，$n-1$, $n-m$, $m-1$ である．従って，組内分散 $V_{R(A)}$ と組間分散 $V_A$ はそれぞれ，$V_{R(A)} = S_{R(A)}/(n-m)$, $V_A = S_A/(m-1)$ である．

組 $i$ の母集団は $N(\mu_i, \sigma^2)$ であるとする（**等分散の仮定**）．そして，帰無仮説

H: $\mu_1 = \mu_2 = \cdots = \mu_m$

をおく．すると，$S_{R(A)}/\sigma^2$, $S_A/\sigma^2$, はそれぞれ $\chi^2_{n-m}$, $\chi^2_{m-1}$ に従い，両者は独立である．よって，$F = V_A/V_{R(A)}$ は $F^{m-1}_{n-m}$ に従う．これを用いて，H，即ち，組による差があるかどうかを検定できる．

**例**　[林周二, 17.3] 199606 スーパータイセーで買った M 卵白 10 個 pack は，10 日ほど冷蔵して 19960623 測定重さは 57, 59, 59, 60, 60, 60, 61, 61, 61, 62, であった．19960117 買った地卵 10 個 pack は，茶: 61, 62, 64, 64, 68, 白: 58, 63, 64, 66, 67, であった．

$m = 3$, $n_1 = 10$, $n_2 = n_3 = 5$, $n = 20$, $\bar{X}_1 = 60$, $\bar{X}_2 = 63.8$, $\bar{X}_3 = 63.6$, $\bar{X} = \dfrac{1}{20}(60 \times 10 + 63.8 \times 5 + 63.6 \times 5) = 61.85$, $(n_1 - 1)V_1 = 18$, $(n_2 - 1)V_2 = 4 \times 7.2 = 28.8$, $(n_3 - 1)V_3 = 4 \times 12.3 = 49.2$,

$V_{R(A)} = S_{R(A)}/(n-m) = \sum_i \sum_j (X_{ij} - \bar{X}_i)^2/(n-m) = \sum_i (n_i - 1)V_i/(n-m) = 96/17 = 5.65$, $V_A = S_A/(m-1) = \sum_i n_i(\bar{X}_i - \bar{X})^2/(m-1) = 68.55/2 = 34.275$.

組 $i$ の母集団は $N(\mu_i, \sigma^2)$ （等分散の仮定）とし，帰無仮説

H: $\mu_1 = \mu_2 = \mu_3$

をおくと，$F = V_A/V_{R(A)} = 34.275/5.65 = 6.07$ は $F^2_{17}$ に従う．

[小針, pp.289–290] を見ると，$0.05 = \mathrm{P}(\ F^2_{17} > 3.59\ )$, $0.01 = \mathrm{P}(\ F^2_{17} > 6.11\ )$．よって，危険率 99% では H は棄却されないが， 95% では棄却されて，差がある，と言える．

**めのこでみても，平均で 4g 程度の差が見える．標準偏差が 1–2g 程度なので，M 卵と地卵は差がありそうな微妙なところにいる．よって，H は棄却と棄却不能の微妙なところにいるのはなんら不思議ではない．もちろん，差があるとしても，M 卵と地卵で計って差をつけて売っているのか，冷蔵庫に入れてしばらく放置したために重さが変わったのかは分からない．**

**問題** 19960117 買った地卵 10 個 pack は，茶: 61, 62, 64, 64, 68, 白: 58, 63, 64, 66, 67, 19960712 買った地卵 10 個 pack は，茶: 64, 64, 65, 67, 70, 白: 62, 62, 64, 66, 67, であった．同じ母集団から得られたサンプルと言えるか？

**問題** 199606 スーパータイセーで買った M 卵白 10 個 pack は，10 日ほど冷蔵して 19960623 測定重さは 57, 59, 59, 60, 60, 60, 61, 61, 61, 62, であった．

19960716 買った M 卵白 10 個 pack は，12 時間ほど冷蔵して重さは 57, 58, 58, 59, 61, 61, 62, 62, 62, 63, であった．

19960727 買った地卵 10 個 pack（茶）は，65, 65, 66, 66, 68, 68, 69, 69, 69, 69, であった．

19960705 買った地卵 10 個 pack は，茶: 64, 68, 68, 72, 73, 白: 65, 68, 70, 72, 72, であった．

### 14.4.8 相関係数の分析

[林周二, 19.1,2]. 母集団を 2 変数正規分布で相関係数 $\rho$ とし，大きさ $N$ の標本 $(X_i, Y_i), 1 \le i \le N$, を選んで，$\{X_i\}, \{Y_i\}$ の標本平均と不偏分散を，それぞれ $\bar{X}, \bar{Y}, V_X, V_Y$ とする．また，$C = \frac{1}{N-1}\sum_i (X_i - \bar{X})(Y_i - \bar{Y})$, $r = \dfrac{C}{V_X V_Y}$, とおく．

**命題 70** *(i)* $\rho = 0$ ならば $t = r\sqrt{\dfrac{N-2}{1-r^2}}$ は $T_{N-2}$ に従う．

*(ii)* $|\rho| \neq 1$ ならば， $z = \dfrac{1}{2}\log\dfrac{1+r}{1-r}$ は，$N(\dfrac{1}{2}\log\dfrac{1+\rho}{1-\rho}, \dfrac{1}{N-3})$ にほぼ従う．従って，

$$t = \sqrt{N-3}\,\frac{1}{2}\left(\log\frac{1+r}{1-r} - \log\frac{1+\rho}{1-\rho}\right)$$

は $N(0,1)$ にほぼ従う（$z$-変換）．

証明は [丸山, p.161], [丸山演, p.134], [ウィルクス].

前半は独立性の検定に，後半は相関係数の推定に用いることができる．具体例は [小針, §7.7 (命題 7.12, 7.13), §8.5], [林周二, 19.2].

### 14.4.9 二項分布と $F$ 分布

二項係数の途中和も $F$ 分布の分布関数も，不完全ベータ関数で表せる．統計的推測としては二項分布（硬貨投げ等）という限られた問題にしか適用できないが，面白い公式として，[小針, §7.6] を紹介する．

**命題 71** $F$ 分布 $F_n^m$ の分布関数 $G_n^m(x) = \mathrm{P}(\ F \leq x\ ) = \int_0^x g_n^m(z)\,dz$ と $N > r \geq 0$, $0 < p < 1$ に対して,

$$\sum_{k=0}^{r} {}_N C_k p^k (1-p)^{N-k} = G_{2(r+1)}^{2(N-r)}(\frac{n(1-p)}{mp}) = B(r+1, N-r)^{-1} \int_p^1 y^r (1-y)^{N-r-1}\,dy\,.$$

**証明**. (30) で変数変換 $mz + n = n/y$ を行えば，簡単な算術で

$$G_n^m(x) = B(\frac{n}{2}, \frac{m}{2})^{-1} \int_{\frac{n}{mx+n}}^1 y^{\frac{n}{2}-1}(1-y)^{\frac{m}{2}-1}\,dy$$

を得る．ここで,

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b}{\Gamma(a+b)}) = B(b,a)\,.$$

$m = 2(N-r)$, $n = 2(r+1)$, $mpx = n(1-p)$ とおくと,

$$G_{2(r+1)}^{2(N-r)}(\frac{n(1-p)}{mp}) = B(r+1, N-r)^{-1} \int_p^1 y^r (1-y)^{N-r-1}\,dy$$

他方で,

$$\sum_{k=0}^{r} {}_N C_k p^k (1-p)^{N-k} = B(r+1, N-r)^{-1} \int_p^1 y^r (1-y)^{N-r-1}\,dy\,.$$

この式は [小針, §7.6] に証明されている.（$r$ についての帰納法と部分積分で証明する.）

これを硬貨投げの検定に適用する例は [小針, §8.4, 例題 9,10].

### 14.4.10 大標本理論と小標本理論

§14.1 で正規分布を用いて推測（推定，検定）を行う際，例えば §14.2 で，$\sigma^2$ の値が分かっている，と仮定した．実際は $V_n$ で $\sigma^2$ の代用とすることで，正規分布をそのまま用いて推測を行った．この方法を大標本理論と呼ぶ.

大きなサンプルを取るために，経済的問題が生じたり独立性などの原則が崩れる恐れがあり，限界がある．そもそも，分散が分かっていて，平均が分かっていないという状況は考えにくい.

代用を必要とせず，従って，むやみに大きなサンプルを用意しないですむ方法が §14.4, §14.4.5, §14.4.5 の小標本理論である．$F$ 分布によって母集団が正規分布の場合の小標本理論が完成したとされる.

大標本理論は中心極限定理に頼って，データ数を重ねれば母集団は何でも良かった．これに対して，小標本理論は母集団の形を仮定する必要がある．特に，小標本理論の完成しているケースは母集団が正規分布と仮定する.

これは大標本理論の方がよい，という意味ではなく，母集団の形を詳しく仮定できればできるほど，より精密な結果を得る，という当然の事実を表現しているに過ぎない.

# 15 回帰分析

[林周二, 19.3], [渡辺浩]. $p+1$ 個の量 $X_1, X_2, \cdots, X_p, Y$ に関して，$n$ 組の測定データがあるとする.

| | $Y$ | $X_1$ | $X_2$ | $\cdots$ | $X_p$ |
|---|---|---|---|---|---|
| data 1 | $y_1$ | $x_{11}$ | $x_{12}$ | $\cdots$ | $x_{1p}$ |
| data 2 | $y_2$ | $x_{21}$ | $x_{22}$ | $\cdots$ | $x_{2p}$ |
| $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ | $\cdots$ |
| data $n$ | $y_n$ | $x_{n1}$ | $x_{n2}$ | $\cdots$ | $x_{np}$ |

これらのデータを 1 次式

$$Y = a_0 + \sum_{j=1}^{p} a_j X_j$$

で近似（回帰）したい．

言い換えると，$Z = (X_1, X_2, \cdots, X_p, Y) : \Omega \to \mathbf{R}^{\mathbf{p+1}}$ のサンプル（独立同分布確率変数列）$Z_k = (X_{k1}, X_{k2}, \cdots, X_{kp}, Y_k)$, $k = 1, 2, \cdots, n$, のサンプル値 $x_{kj} = X_{kj}(\omega)$ を元に，$Z$ の分布の集中する超平面を見つけたい．誤差

$$E = \sum_{k=1}^{n} (Y_k - a_0 - \sum_{j=1}^{p} a_j X_{kj})^2$$

を最小とする回帰係数 $a_j$, $j = 0, 1, 2, \cdots, p$, は，次式で与えられる．

$$\begin{aligned} a_j &= \sum_{k=1}^{p} C^{-1}_{jk} d_k \\ a_0 &= \bar{Y} - \sum_{k=1}^{p} a_k \bar{X}_k \end{aligned}$$

ただし共分散 $C(X, Y) = \mathrm{E}[\,(X - \mathrm{E}[\,X\,])(Y - \mathrm{E}[\,Y\,])\,]$,

$$\begin{aligned} C_{jk} &= C(\boldsymbol{X}_j, \boldsymbol{X}_k) = \sum_i (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k), \\ d_k &= C(\boldsymbol{Y}, \boldsymbol{X}_k) = \sum_i (Y_i - \bar{Y})(X_{ik} - \bar{X}_k). \end{aligned}$$

$Y_k$ には，$X_{kj}$ によって説明される部分
予測値: $\underline{Y}_k = a_0 + \sum_{j=1}^{p} a_j X_{kj}$
と，説明されない部分
残差: $Y_k - \underline{Y}_k$
がある．これに対応して，
$Y$ の変動: $S_{\mathrm{total}} = \sum_{k=1}^{n} (Y_k - \bar{Y})^2$
は，$X$ の変動に起因する部分
回帰変動: $S_r = \sum_{k=1}^{n} (\underline{Y}_k - \bar{Y})^2$
と，それ以外の部分
残差変動: $S_e = \sum_{k=1}^{n} (Y_k - \underline{Y}_k)^2$
に分かれる:

$$S_{\mathrm{total}} = S_r + S_e$$

そこで，仮定「 $Y_k$ は，$N(\alpha_0 + \sum_{j=1}^{p} \alpha_j X_{kj}, \sigma^2)$ に従う，独立な確率変数である」のもとで，$Y$ は，$X_1, X_2, \cdots, X_p$ の影響を受けるかどうかを検定する．

最小 2 乗法によって定めた値 $a_i$, $i = 0, 1, 2, \cdots, p$ は，母数 $\alpha_i$, $i = 0, 1, 2, \cdots, p$ の推測値であるが，

**定理 72 (回帰分析)** 帰無仮説 $H$: $\alpha_j = 0$ $(j = 1, 2, \cdots, p) \Longrightarrow F = \dfrac{\frac{1}{p} S_e}{\frac{1}{n-p-1} S_r}$ は $F^{p}_{n-p-1}$ に従う．

# 16 検定関数

[数学辞典, 284A] 検定を次のように一般化する．関数 $\phi: \mathbf{R} \to [0,1]$ が与えられたとき，サンプル $X(\omega) = x$ に対して，確率 $\phi(x)$ で仮説を棄却する統計的推測を考える．このときの $\phi$ を検定関数という．

古典的な検定は $\phi = \chi_A$ （$A$ は棄却域）とおいたと見ることができる．

# 17 ノンパラメトリック法 18

通常の統計的推論は母集団について，小数の未知母数以外は既知とする．応用上非現実的な仮定 [数学辞典, 445A].

ノンパラメトリック法（ロバスト法）は母集団の具体形になるべくよらない統計的推測．[数学辞典, 445B] に従って，符号検定と Wilcoxon の符号付き順位検定を紹介する．

## 17.1 符号検定

実確率変数 $X$ の分布関数を $F$，大きさ $n$ のサンプル (i.i.d.) $X_i$, $i = 1, 2, \cdots, n$．$0 < p < 1$ に対して $F(\xi_p) = p$ なる $\xi_p$ をとる．

帰無仮説 H: $\xi_p = \xi^*$ (または，H': $\xi_p \leq \xi^*$)

対立仮説 (§18.2) A: $\xi_p > \xi^*$

の検定．$i(\{X_j\}) = \sharp\{j \in \{1, 2, \cdots, n\} \mid X_j > \xi^*\}$ とおく．

$$\phi(\{X_j\}) = \begin{cases} 1, \text{ if } i(\{X_j\}) > c, \\ a, \text{ if } i(\{X_j\}) = c, \\ 0, \text{ if } i(\{X_j\}) < c \end{cases},$$

$c = np$[8]，を検定関数とする検定を符号検定と呼ぶ．

## 17.2 Wilcoxon の符号付き順位検定

帰無仮説 H: $\xi_{1/2} = \xi^*$

対立仮説 A: $\xi_{1/2} > \xi^*$

の場合．

$F$ が $x = \xi_{1/2}$ に関して対称で，密度 $f$ を持つ絶対連続分布関数とする．$|X_i - \xi^*|$ が $i$ を動かした中で $R_i^+$ 番目に大きいとし，

$$s_n = \sum_{i:\ X_i > \xi^*} R_i^+$$

とおく．

$$\phi(\{X_j\}) = \begin{cases} 1, \text{ if } s_n > c, \\ a, \text{ if } s_n = c, \\ 0, \text{ if } s_n < c \end{cases},$$

$c = n(n+1)/4$[9] を検定関数とする検定を，Wilcoxon の符号付き順位検定と呼ぶ．

---

[8] だと思うが，[数学辞典] には定義が欠落している

[9] だと思うが，[数学辞典] には定義が欠落

### 17.3 Kolmogorov–Smirnov 検定

母集団の分布関数 $F_0$，大きさ $n$ のサンプルの経験分布関数 $F_n$，とする．$d_n = \sup|F_n(x) - F_0(x)|$, $D_n = \sup(F_n(x) - F_0(x))$, 等に対して，

$$\lim_{n\to\infty} \mathrm{P}(\, d_n < z/\sqrt{n}\,) = L(z) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2z^2}$$
$$\lim_{n\to\infty} \mathrm{P}(\, D_n < z/\sqrt{n}\,) = 1 - e^{-2z^2}$$

となる．これらは適合度検定（H: $F = F_0$ の検定）に用いることができる．他にもいくつかある [数学辞典, 445F].

## 18 検定の定式化 17

無作為抽出されたサンプルをもとに，母集団に関する，ある程度信頼のおける命題を立てること．

母集団に関する適当な仮説（帰無仮説）を選び，その仮説下で適当な統計量（サンプルの関数）を選んで計算し，サンプル値（統計量の実現値）に基づいて，仮説の正否を判断する．帰無仮説が正しいとすると今得られているデータが起きる確率 $p$ は非常に小さい．あらかじめ定められた数値 $\alpha$ （危険率）より $p$ が小さいときに，帰無仮説を棄却する．

### 18.1 素朴な方法の直感的根拠

**検定，区間推定**　棄却域の選び方：なぜ片側危険域を採るか？途中を棄却してはいけないのか？どんな標本値でも（元の空間の大きさが大きければ）その生じる確率は一般に無限小である．それでも棄却したりはしない．

例：乱数列発生装置「153496478...」このような特定の列が出る確率は長さ $N$ の列では $10^{-N}$ であり，殆どあり得ない．よって，乱数とは言えない？

危険率はいくらに選ぶか？

**点推定**　なぜ $f_\theta(x)$ という特定のパラメータ空間を考察の対象とするのか？この parametrization はどうやって選択したのか？

### 18.2 対立仮説，第１種の過誤，第２種の過誤

[小針, §8.1], [楠岡 2], [数学辞典, 284A].

**対立仮説**　検定は，現実には帰無仮説を別の仮説（あるいは一群の仮説）と比較している．

例：既存の理論，常識（帰無仮説）と新説，新研究（対立仮説）の対決，正常状態と予想される異常事態．

帰無仮説の分布 $f_{\theta_0}$，対立仮説の分布 $f_\theta$ （素朴なケースでは $\theta \neq \theta_0$ から一つ選ぶ）．

**第１種の過誤**　棄却率 $\alpha$: 本当は正しいのに**うっかり**していて棄却してしまう確率 [渡辺浩]．データ $x = X(\omega)$ が危険域 $A$ $(\in \mathcal{B})$ に入ったら帰無仮説 $f_{\theta_0}$ を棄却する．

例：新製品試作品人気投票をやったところ，審査員の好みが偶然偏っていたため，没になった．

第 1 種の過誤 $\alpha = \displaystyle\int_A f_{\theta_0}(x)\,dx$. 普通は，先ず，$\alpha = 5\%, 1\%$ などに設定し，そうなるように $A$ を選ぶ．§14.2 も参照．$\alpha$ を小さくとれば，第 1 種の過誤は減らせる．

**第２種の過誤**　本当は間違っているのに**ぼんやり**していて正しいと言ってしまう [渡辺浩] (§C に渡辺さんとの電子メール問答).

例：仮説「星占いは当たる」がデータ不足で否定できなかったので，肯定した．

第２種の過誤 $\beta = \int_{A^c} f_\theta(x)\,dx$. $\alpha$ を決めたとき，これを小さく抑えたい．これが危険域の形 $A$ を決める．

例：対立仮説の分布が帰無仮説の分布より右にずれている場合は，右側危険域とするのが $\beta$ を最小に抑える．

言い換えれば対立仮説が危険域の形を決める (Neyman–Pearson) [数学辞典, 284B], [楠岡 2].

経験や直感など素朴な $\alpha$ や危険域の決め方は，頭の中に，対立仮説がある，と理解する．これはヒトが検定に何を求めているのかという行動原理の考察を必要とする．

**検定力**　同じ $\alpha$ に対して $\beta$ の小さい検定方法（今の場合は危険域の設定）を「検定力が強い」と言う．

§18.2 より，帰無仮説の分布 $f_{\theta_0}$，対立仮説の分布 $f_\theta$，第１種の過誤 $\alpha = \int_A f_{\theta_0}(x)\,dx$，第２種の過誤 $\beta(\theta) = \int_{A^c} f_\theta(x)\,dx$.

$P(\theta) = 1 - \beta(\theta)$ を（検定方式 $\alpha, A$ の）検定力，対立仮説 $\theta$ の関数とみて検定力関数と呼ぶ．最強力検定関数の特徴付けについては Neyman–Pearson の定理 [数学辞典, 284B], [楠岡 2], [林周二, 14.3] を参照.

**科学的禁欲主義，あるいは，損失の帰属**　帰無仮説が正しくないと結論するためには §14 などの素朴な危険率の設定で実用上の役に立つ.（簡単なので，簡単なチェックにはこのほうが役に立つ.）危険率 $\alpha =$ 帰無仮説が正しいのに捨てる誤りを犯す確率（第一種の過誤）．棄却できないとき帰無仮説が正しくないのに正しいとする誤りは第二義的な上に大きい（第二種の過誤）．このような経験的な危険率の設定の仕方の根拠：

(i) 科学的禁欲主義 [楠岡 2, §6.3]．第一種過誤は既存の理論が正しいのに新理論に飛びつくこと，第二種過誤は新しい理論が正しいのに既存の理論に執着すること．

(ii) 生産者の損得優先主義．抜き取り検査では第一種過誤は良品を不合格とする生産者の損失，第二種過誤は不良品を合格とする消費者の損失 [林周二, 13.3].

## 18.3　Bayes 統計学

[楠岡 2, §5]

### 18.3.1　古典的数理統計学の歴史

**18 世紀中頃** Bayes 統計学（Bayes の公式，事後確率，先験的確率）

**18 世紀末** 正規分布 (Laplace, Gauss)

**19 世紀後半** 大標本理論完成（$\chi^2$ 分布 (Karl Pearson)）

**20 世紀初頭** 小標本理論完成（$t$ 分布 (Gosset)，$F$ 分布，帰無仮説，客観的統計的推測（Bayes 統計学批判）(R. A. Fisher)）

**20 世紀前半** ２仮説検定完成（対立仮説，第１種過誤，第２種過誤，確率検定，検出力 (Neyman–E. S. Pearson)，Fisher–Neyman 論争）

**20 世紀中頃** Bayes 統計学と仮説検定の同値性（Wald の定理，期待効用関数 (Savage)，Allais の反論）

**残された問題** 期待効用と実際の主観の乖離

### 18.3.2　Bayes の公式

**定理 73** $A \in \mathcal{F}$, $C_i \in \mathcal{F}$, $i = 1, 2, 3, \cdots, n$, とし，$\displaystyle\sum_{i=1}^{n} C_i = \Omega$ とすると，

$$\mathrm{P}(\, C_i \mid A \,) \frac{\mathrm{P}(\, A \mid C_i \,)\mathrm{P}(\, C_i \,)}{\displaystyle\sum_{i=1}^{n} \mathrm{P}(\, A \mid C_i \,)\mathrm{P}(\, C_i \,)}$$

が成り立つ．ここで $\mathrm{P}(\, B \mid A \,) \stackrel{\mathrm{def}}{=} \mathrm{P}(\, B \cap A \,)/\mathrm{P}(\, A \,)$ を条件 $A$ の下での条件付き確率と呼ぶ.

**証明**. 定義から

$$\mathrm{P}(\, C_i \mid A \,) = \frac{\mathrm{P}(\, C_1 \cap A \,)}{\mathrm{P}(\, A \,)}\,.$$

与式の右辺がこの式の右辺に等しいことも定義からすぐ従う.

この式は数学的定理だから常に正しい．これをどう現実に応用するかが問題．Bayes 統計学では，条件付き確率 $\mathrm{P}(\, B \mid A \,)$ は $A$ が起こったときに，そのあとで，さらに $B$ が起こる確率，と理解．すると，Bayes の公式は,

$A$ という結果が起きたことが分かったときに，隠されていた過去の事実（原因）$\{C_i\}$ のうち $C_i$ が起きていた確率を，$C_j$ の条件付きで $A$ が起こる確率を用いて（遡って）表す式.

## 18.4　先験的確率

問題は**事前確率（先験的確率）** $\mathrm{P}(\, C_i \,)$ を，実用上のケースで求められるかどうか，という点．事前確率は主観的なものとして，排除する立場が Bayes 統計学批判の主要点.

**例**　$10^5$ 人が買い物に利用する A 商店街で福引きがあった．100 人に 1 人の割合で当たりが出る．ところが，実は $0 < p < 1$ を定数として，$10^5 p$ 人分の「お得意さま名簿」があって，福引き立会人が裏で，お得意さま，と確認すると，自動的に当たる細工があった．さて，福引きを行った Y 氏は当たった．氏はお得意さまと認められた確率はいくらか？

解：$C_1$ を Y がお得意さまである事象，$C_2 = {C_1}^c$ とする．$A$ を Y が当たる事象とする.

$$\mathrm{P}(\, C_1 \mid A \,) = \frac{\mathrm{P}(\, A \mid C_1 \,)\mathrm{P}(\, C_1 \,)}{\displaystyle\sum_{i=1}^{2} \mathrm{P}(\, A \mid C_i \,)\mathrm{P}(\, C_i \,)} = \frac{1 \times p}{1 \times p + 0.01 \times (1-p)} = \frac{p}{0.99p + 0.01}\,.$$

$p = 0.01$ ならほぼ 50%, $p = 0.0001$ なら 1%, $p = 0.5$ なら 99%.

この例は先験的確率 $\mathrm{P}(\, C_i \,)$ が分かっていれば解ける．しかし，裏名簿は普通，我々からは隠されているので，先験的確率は非常に知りにくい量であるのが普通.

**例 2**　人口 $10^5$ 人の A 市で事件が起こり，Y 氏が容疑者となった．犯人は血痕を残しており，その血液型は 100 人に 1 人の割合である．Y 氏は同じ血液型だった．氏が犯人の確率はいくらか？ [楠岡, §5.2]

解：例 1 で，お得意さま→犯人，当たり→同じ血液型，と置き換えればよい．$p$ に対応するのは「お得意さま名簿」→「市民の中の犯人の比率」になる．計算は例 1 と全く同じ.

$$\mathrm{P}(\, C_1 \mid A \,) = \frac{p}{0.99p + 0.01}\,.$$

$p = 0.01$ ならほぼ 50%, $p = 0.0001$ なら 1%, $p = 0.5$ なら 99%.

$p = \mathrm{P}(\ C_i\ )$ は犯人であることの事前確率「市民の中の犯人の比率」[10] である．$p$ が五分五分なら，Y が 99% 犯人．$10^5$ 人の中で犯罪者は 10 人くらいと思うなら，Y は 99% 犯人ではない.（犯人かどうかは確定していて，確率を論じられない，という考え方もある．ここでは，これは論じない.）

この例は「分からなければ五分五分」という考え方の不適切を示している，と考えるべきである．

**例 3**　3 人の囚人アラン，バーナード，チャールズが牢屋にいる．アランは翌日二人が処刑され一人が釈放されることを知ったが，本人の運命についてはそのときまで本人には秘密となっていた．そこで，アランは看守に「3 人のうち二人が処刑されるのだから，バーナードとチャールズのうち少なくとも一人は処刑される．どちらか一人が処刑されるとわかっても私のことについては情報を与えない．バーナードとチャールズのうち処刑される者の名前を一名だけ教えてほしい.」と言った．看守は納得して「バーナードは処刑される」と答えた．アランはこれを聞いて，釈放される可能性があるのは，自分とチャールズだけになったので釈放の確率が 1/3 から 1/2 に増えたと喜んだ．これは本当か？ [楠岡, §1.1, 問題 1.5; §5.2]

解：$C_A$ を A が釈放される事象，などとし，$A$ を，看守が「B 処刑」と答える事象とする．事前確率は $C_i = 1/3$ としても，この問題の場合条件付き確率 $\mathrm{P}(\ A \mid C_A\ )$ が，A には分からない．1/2 とすると釈放確率は変化しないが，$C_A$ のとき看守が公平な硬貨を投げて「B 処刑」と「C 処刑」の発言を選ぶかどうか A は知らない [11]．

## 18.5　非 Bayes 統計学にひそむ主観

危険域の形の決定や危険率の設定や母数空間の選択には経験や主観が入っている §18.1．実際ある意味で先験的 (主観的) 確率と見ることができる §18.6, §18.7.

**推定**　Parametrization を特定の形 $f_\theta$ に選んだのは先験的確率密度として $f_\theta$ に 1 それ以外を 0 に選ぶ主観が入っている．$\exp(-x^2)$ が正解のとき $\exp(-(x-\theta)^2)$ を parametrization に選ぶか $\exp(-x^2 - \theta x^4)$ を選ぶかで，データから推測される結果は当然変わる．

**検定**　対立仮説は無数にある．対立仮説の選びかた＝各仮説の起る確率という確率空間.

**注**. 通常素朴な方法では，対立仮説は決定論的に一つ固定するが，これは 1 つの可能性に確率 1 を与えたと見ればよい．(離散的だと危険率 $\alpha$ を固定できなくなるので，対立仮説は確率空間にする方が都合が良い．)

## 18.6　Wald の損失関数

[数学辞典, 285A, 287E], [楠岡 2, §7]

**検定理論の要約**　§14.3, §18, §16.

さいころは正しいか？

→ 仮説 $p = 1/6$ の検定 (帰無仮説 H)

→ 仮説が正しくないとき起こることの設定 (対立仮説 A) 複数でも良い

→ 検定関数 (§16)：サンプル値の関数として，説 $C \in \{H, A\}$ を採用する確率を $Q(C \mid X(\omega))$ とおく ($\phi(x) = Q(A \mid x)$)

→ 第 1，2 種の過誤 $\alpha, \beta$ を $Q$ で表す

[10] [楠岡] ではこの認識が欠けている
[11] [楠岡, §5.2 演習問題 5] では，看守の発言確率の問題が欠けている

➞ $\alpha = \alpha_0$ を固定（$Q$ の線型拘束条件）

➞ $\beta$ を最大にする $Q = Q_{\alpha_0}$ の決定

実際，$X$ をサンプル（独立確率変数列），仮説 $C$ に対応する母集団における期待値を $\mathrm{E}[\,\cdot \mid A\,]$（条件付き期待値と間違えるので良くない書き方）と書くと，

$$\begin{aligned}\alpha &= \mathrm{E}[\,Q(A \mid X) \mid A\,]\\ 1-\beta &= \mathrm{E}[\,Q(A \mid X) \mid A\,].\end{aligned}$$

検定理論は $Q$ の変分問題に帰着した．

この解は，次の性質を持つことは容易に分かる [楠岡 2, §6.3] ：

(37) $R_+(a) = \left\{x \mid \dfrac{\mathrm{P}(\,X = x \mid A\,)}{\mathrm{P}(\,X = x \mid H\,)} > a\right\}$， $R_-(a) = \left\{x \mid \dfrac{\mathrm{P}(\,X = x \mid A\,)}{\mathrm{P}(\,X = x \mid H\,)} < a\right\}$

とおくとき（条件付き確率と間違えるので良くない書き方），

(38) $\exists a;\ x \in R_+(a) \Rightarrow Q(A \mid x) = 1,\ x \in R_-(a) \Rightarrow Q(A \mid x) = 0\,.$

よって $Q_{\alpha_0}$ の決定は難しくはない．

特に，変分問題の解の全体 $\{Q_{\alpha_0} \mid 0 \le \alpha_0 \le 1\}$ は (38) が示す空間（確率測度の空間）$\mathcal{B} = \{Q \mid \exists a;\ x \in R_+(a) \Rightarrow Q(A \mid x) = 1,\ x \in R_-(a) \Rightarrow Q(A \mid x) = 0\}$ に含まれる．

**Bayes 統計学の立場**　$\mathcal{B}$ は Bayes 統計学に基づく解釈が可能である [楠岡 2, §7.1]．先験的確率，即ち，確率空間上に確率測度が定義されているとする．$\mathrm{P}(\,X = x \mid H\,)$ などは本当に条件付き確率の意味を持つようになる．このとき，Bayes の公式から，

$$\frac{\mathrm{P}(\,H \mid X = x\,)}{\mathrm{P}(\,A \mid X = x\,)} = \frac{\mathrm{P}(\,X = x \mid H\,)\mathrm{P}(\,H\,)}{\mathrm{P}(\,X = x \mid A\,)\mathrm{P}(\,A\,)}\,.$$

左辺から，この量が 1 より大ならば H を（確率 1 で）採用する ($Q(A \mid x) = 0$) し，小ならば H を棄却して A を（確率 1 で）採る ($Q(A \mid x) = 1$) べきである．ところが，$a = \dfrac{\mathrm{P}(\,H\,)}{\mathrm{P}(\,A\,)}$ とおくと，右辺と (37) の定義から，これは，$x \in R_-(a) \Rightarrow Q(A \mid x) = 0,\ x \in R_+(a) \Rightarrow Q(A \mid x) = 1,$ ということである．即ち，$\mathcal{B}$ は先験的確率を動かしたときの，Bayes 統計学の立場で選択される確率測度の空間と一致する．

このように，非ベイズ統計学の検定理論と Bayes 統計学の Bayes 公式に関係がついた．これを一般化したのが Wald の定理である．

**Wald の定理**　具体的な定義と定理は [楠岡 2, §7.2] を参照．アイディアは次の通り．

先ず，統計的推測を，期待利得を最大化する決定を選ぶ問題とみなす．即ち，母数 $\theta$，推測（推定または検定）$Q(C \mid x)$，データ $X$，に対して期待利得

$$U(Q,\theta) = \sum_C \sum_x Q(C \mid x)\mathrm{P}(\,X = x \mid \theta\,)u(C,\theta)$$

と書けるとする．任意の $\theta$ に対して $U(Q,\theta) \le U(Q',\theta)$, かつ，少なくとも一つの $\theta$ に対して $U(Q,\theta) < U(Q',\theta)$ ならば $Q'$ は $Q$ より真によい推測（決定）とよぶ．そして，admissible な決定とは，真により良い決定のない決定のこととする．

他方，Bayes 統計学における決定は，先験的確率 $\mathrm{P}(\,\theta\,)$ が与えられたときの事後相対期待利得，

$$V(C, x, \mathrm{P}(\,\theta\,)) = \sum_\theta u(C,\theta)\mathrm{P}(\,X = x \mid \theta\,)\mathrm{P}(\,\theta\,)$$

を最大にする $C$ である．即ち，最大にしない $C$ に対して $Q(C \mid x) = 0$ を満たす $Q$ である．これを Bayes 解と呼ぶ．

このとき，Wald の定理は，次のことを言う．任意の non-admissible な決定に対して，それよりも真によい admissible な決定があり，admissible な決定の全体は，$\mathrm{P}(\,\theta\,) > 0$ が全ての $\theta$ に対して成り立つときの対応する Bayes 解の，事前確率 $\mathrm{P}(\,\theta\,)$ を動かしたときの全体と一致する．

## 18.7 Savage の理論

[楠岡 2, §8], [数学辞典, 286F, 287F]．Wald の定理の結果，非ベイズ統計的推測は，どの $\theta$ で最大利得を得るか（検定では帰無仮説と棄却率のペアに対応）という，手で入れる変数がベイズ統計学の事前確率に相当することが分かった．

逆に言えば，ベイズ統計学を基本に置いて議論を展開しても，非ベイズ統計学に比べて新たな主観を導入したことにはならない．これを積極的にとらえて，ベイズ統計学に効用関数（期待利得）の概念を付加したものを，合理的な人間の意志決定原理という意味で統計学の出発点とし，統計学を合理的意志決定の学問とみるのが Savage の立場である．

Savage の期待効用理論．合理的意志決定を行う人間は以下を行動原理とする：

(i) 母数 $\theta$ と決定（推測） $C$ の関数として効用関数 $u(C,\theta)$ が定義されている
(ii) 母数に対する事前確率 $\mathrm{P}(\ \theta\ )$ が定義されている
(iii) 期待効用 $U(C) = \sum_\theta \mathrm{P}(\ \theta\ )u(C,\theta)$ を最大にする決定 $C$ を選ぶことが課題となる（データがあるときは事後確率に置き換わる）

## 18.8 Allais の反例 Edwards の例

[楠岡 2, §1.1 問題 1.10, 1.11; §8.4]．効用が $\mathrm{P}(\ \theta\ )$ の線形関数であるという仮定は，人間の行動原理としては不適当であるという指摘がある．

**問題 (Allais)**　次の２つの質問両方に答えよ．

**質問１**　次の２つのくじ A, B のどちらを選ぶか？

**A** 確実に 100 万円当たる

**B** 確率 0.1 で 500 万円，確率 0.89 で 100 万円，確率 0.01 で 0 万円当たる

**質問２**　次の２つのくじ C, D のどちらを選ぶか？

**C** 確率 0.1 で 500 万円，確率 0.90 で 0 万円当たる

**D** 確率 0.11 で 100 万円，確率 0.89 で 0 万円当たる

もし，期待効用理論が正しければ，各くじの期待効用は次のようになる．

**A** $U_A = U(100)$

**B** $U_B = 0.1U(500) + 0.89U(100) + 0.01U(0)$

**C** $U_C = 0.1U(500) + 0.9U(0)$

**D** $U_D = 0.11U(100) + 0.89U(0)$

$U_B - U_A = U_C - U_D$ が成り立つので，Savage の合理的人間は $A$ を選べば $D$ を，$B$ を選べば $C$ を選ぶ．実際は $A$, $C$ を選ぶ人が多い．

**注**. 960628 哲弥．くじが当たるまでの構え（未来への備え）のために負担があり，はずれると損失が余分に起きる．即ち事前確率最大の母数に対する効用は相対的に高くなる．$Q$ dependence が non-linear であるということ．これで説明ができると思う．

**問題 (Edwards)**　赤玉，白玉がそれぞれ 7:3, 3:7 で入っている 2 つのつぼ A, B のいずれか一方から，玉を 12 個復元抽出したとき 8 個の赤玉． 4 個の白玉を得た．つぼが A である確率を計算せずに直感で答えよ．

[楠岡 2, §8.4] によれば，正解 0.967365 に対して，0.9 以上の答えが少なく， 0.95 以上は殆どいなかったとのこと．

**注**. 960628 哲弥．私は第 1 感 95% だった．

# A 合同法一様擬似乱数生成検査プログラム

§11.3 で実際に用いたプログラム．

```
c 合同法一様疑似乱数自己相関と分布関数検査 19950518;19
      real*8 r,r32
      real*8 listorg
      real*8 ave1,ave2,var1,var2,cov,cor,supf
      integer rrnd,irnd,i
      integer max,inextchk,ichkdelta,nchk
      parameter (max=2**14, ichkdelta=4)
      parameter (r32=2d0**32, rrnd=48828125)
      common /acom/listorg
      dimension listorg(max)
      data irnd/1000001/
      inextchk=ichkdelta
      nchk=1
      write(6,'(1x,a,\)') 'iteration 回数 (≦ 2^14) = '
      read(5,'(i10)') maxiter
      open(1,file='rnd2.dat')
      write(1,'(a6,2x,a12,1x,a12,2x,4x,a13,1x,3x,4x,a12)')
     & ' data#','E[Xk;1 ≦ k<N]','V[Xk;1 ≦ k<N]',' ρ [Xk,X{k+1}]',
     & 'sup|x-Fn(x)|'
      write(1,'(8x,a12,1x,a12,2x,4x,a14)')
     & 'E[Xk;1<k ≦ N]','V[Xk;1<k ≦ N]','cov[Xk,X{k+1}]'
      do 1 i=1,maxiter
      r=irnd/r32+0.5d0
      irnd=irnd*rrnd
      listorg(i)=r
      if (i .eq. inextchk) then
      if (i .gt. 2) then
      call average(i,ave1,ave2,var1,var2)
      call covarichk(i,ave1,ave2,cov)
      cor=cov/dsqrt(var1*var2)
      call sort(i)
      call districhk(i,supf)
      write(1,'(i3,a1,i2,2x,f12.10,1x,f12.10,2x,d18.10,3x,d18.10)')
     & ichkdelta,'^',nchk,ave1,var1,cor,supf
      write(1,'(8x,f12.10,1x,f12.10,2x,d18.10)')
     & ave2,var2,cov
      endif
      inextchk=inextchk*ichkdelta
      nchk=nchk+1
      endif
 1    continue
      write(1,'(/a)')
     & '理想一様乱数：E=0.5, V=1/12=0.0833, ρ=cov=0, sup=0'
      end
```

```
      subroutine average(n,ave1,ave2,var1,var2)
* ave1=average[list(i);i=1,2,..,n-1]==E^{1,n-1}[list]
* var1=variance[list(i);i=1,2,..,n-1]=E^{1,n-1}[(list-ave1)^2]
* ave2=E^{2,n}[list]
* var2=E^{2,n}[(list-ave2)^2]
      integer max,n,i
      real*8 list,ave1,ave2,var1,var2,avecom,varcom
      parameter (max=2**14)
      dimension list(max)
      common /acom/list
      avecom=0d0
      do 1 i=2,n-1
      avecom=avecom+(list(i)-avecom)/dfloat(i-1)
 1    continue
      ave1=avecom+(list(1)-avecom)/dfloat(n-1)
      ave2=avecom+(list(n)-avecom)/dfloat(n-1)
* EN=\sum_{k=1}^N ak/N, VN=\sum_{k=1}^N (ak-EN)^2/N
* -> V{N+1} - VN * N/(N+1) = N (EN-E{N+1})^2 = (a{N+1}-EN)^2 * N/(N+1)^2
* !!! N==n-2 !!!
      varcom=0d0
      do 2 i=2,n-1
      varcom=varcom+((list(i)-avecom)**2-varcom)/dfloat(i-1)
 2    continue
      var1=(varcom+(list(1)-avecom)**2/dfloat(n-1))
     &     *dfloat(n-2)/dfloat(n-1)
      var2=(varcom+(list(n)-avecom)**2/dfloat(n-1))
     &     *dfloat(n-2)/dfloat(n-1)
      return
      end

      subroutine covarichk(n,ave1,ave2,cov)
      integer max,n,i
      real*8 list,ave1,ave2,var1,var2,cov
      parameter (max=2**14)
      dimension list(max)
      common /acom/list
* cov=average[(list(i)-ave1)(list(i+1)-ave2);i=1,2,..,n-1]
* ; ave1=average[list(i);i=1,2,..,n-1]==E^{1,n-1}[list]
*   ave2=E^{2,n}[list]
      cov=0d0
      do 1 i=1,n-1
      cov=cov+((list(i)-ave1)*(list(i+1)-ave2)-cov)/dfloat(i)
 1    continue
      return
      end
```

```
      subroutine sort(n)
      integer max,n,i,j,pointer,sloc
      real*8 listorg,listsort,dummy,infty,infone,small
      parameter (max=2**14,infty=2d0,infone=infty+1d0)
      dimension listorg(max),listsort(max),dummy(max)
      common /acom/listorg, /bcom/listsort
      do 1 i=1,n
      dummy(i)=listorg(i)
 1    continue
      pointer=1
      do 2 i=1,n
      small=infty
      sloc=0
      do 3 j=1,n
      if (dummy(j) .lt. small) then
      sloc=j
      small=dummy(j)
      endif
 3    continue
      if (sloc .eq. 0) stop 'Abnormal stop in subroutine sort.'
      listsort(pointer)=small
      dummy(sloc)=infone
      pointer=pointer+1
 2    continue
      return
      end

      subroutine districhk(n,supf)
      integer max,n,i
      real*8 list,supf,rn,ri,dummy
      parameter (max=2**14)
      dimension list(max)
      common /bcom/list
* list(i);i=1,n (increasing 0<list<1)
* sup_{0 ≦ x ≦ 1} |x-sup{i|list(i) ≦ x}/n|
* = max_{i=1,2,...,n} [ |list(i)-i/n| v |list(i)-(i-1)/n| ]
      rn=dfloat(n)
      supf=0d0
      do 1 i=1,n
      ri=dfloat(i)
      dummy=dabs(list(i)-ri/rn)
      if (dummy .gt. supf) supf=dummy
      dummy=dabs(list(i)-(ri-1d0)/rn)
      if (dummy .gt. supf) supf=dummy
 1    continue
      return
      end
```

# B vonMises 乱数生成速度計測プログラム

[服部中島] で実際に用いたプログラム．§11.4.2 参照.

```
c random U(1) variable (von Mises distribution) generation timing: 19950220
      implicit real*8(a-h,o-z)
      integer*4 irr,ir30,irm
      logical ein,estart,estop,edat
      parameter (pi=3.14159265358979323846d0, pi2=pi/2d0)
      parameter (ir30=1073741824, irm=48828125,r31=2147483648d0)
      parameter (as=0.79895368608398d0)
      parameter (p1=0.4162d0, p2=1.5056d0)
      parameter (q0=p2-2d0*p1*as, q1=as*(p1*as-p2))
      parameter (eps=1d-3, eps2=2d0-eps)
      parameter (ass=0.7994774d0)
      data irr/1000001/
      qb=(dcosh(dsqrt(as*(1d0+eps)*eps)*pi)-1d0)/as/(1d0+eps)/eps/2d0
      inquire (file='cosh.in',exist=ein)
      inquire (file='start',exist=estart)
      inquire (file='stop',exist=estop)
      inquire (file='cosh.dat',exist=edat)
      if ( (.not. ein) .or. estart .or. estop)
     &stop 'file not ready'
      open (1, file='cosh.in')
      read (1, '(i10,2d15.7)') nm,a0,da
      close (1)

      open (1, file='start')
      close (1)
      ntry=0
      do 10 n=1,nm

      irr=irr*irm
      if (irr .lt. 0) irr=(irr+ir30)+ir30
      r=dfloat(irr)/r31

      a=a0+(r-5d-1)*2d0*da
      alph2a=dmin1(eps2, dmax1(eps, p1*a+q0+q1/a))
*       b1=dsqrt(dmin1(2d0, (dexp(2d0*a)-1d0)/a/qb)/alph2a-1d0)
      if (a .ge. ass) then
      b1=dsqrt(2d0/alph2a-1d0)
      else
      b1=dsqrt((dexp(2d0*a)-1d0)/a/qb/alph2a-1d0)
      endif
      alpha=dsqrt(alph2a*a)
      t1=datan(dtanh(pi2*alpha)*b1)
 1    continue
```

```
      irr=irr*irm
      if (irr .lt. 0) irr=(irr+ir30)+ir30
      r=dfloat(irr)/r31

      h2=dtan((2d0*r-1d0)*t1)
      h1=h2/b1
      hp=1d0+h1
      hm=1d0-h1
      h=dlog(hp/hm)/alpha
      g=dexp(a*(dcos(h)-1d0))*(1d0+h2**2)/hp/hm
      ntry=ntry+1

      irr=irr*irm
      if (irr .lt. 0) irr=(irr+ir30)+ir30
      r=dfloat(irr)/r31

      if (g .lt. r) goto 1
10    continue
      open (1, file='stop')
      close (1)
      ratio=dfloat(nm)/dfloat(ntry)
      open (1, file='cosh.dat')
      if ( .not. edat) then
      write (1, '(a,a)') 'cosh(1995)   ',
     &' α^2=min(a(2-ε),max(εa,(p1*δa+p2)*δa));  δa=max(a-a*,0)'
      write (1, '(a,1p,e10.4,0p,2(3x,a,f8.6),3x,a,f20.18)')
     &' ε=',eps,'p1=',p1,'p2=',p2,'a*=',as
      else
501   continue
      read (1,'()',end=500)
      goto 501
500   continue
      backspace (1)
      endif
      write (1,
     &'(1p,a2,e9.3,a2,e7.1,0p,4x,a5,i9,3x,a5,i9,3x,a5,f10.8)')
     &'a=',a0,'±',da,'生成=',nm,'試行=',ntry,'効率=',ratio
      end
```

## C 第２種の過誤に関する問答

渡辺さんとの電子メール問答[12].

Q. 「第２種の過誤」は「実在するとは言えない」という結論で,「実在しない」とは
言わないので,「過誤」ではないと思います. はっきり間違っている「第１種の過誤」
との微妙な logic の差を,「第２種の過誤」という用語は意識させにくいので,
初学者には望ましくないと感じましたが, どうでしょうか？

A 「過誤」ではないと言われれば、確かにそのとおりですね。
御指摘のとおり、この問題は大変重要だと思います。ただし僕としては、「過誤」と呼ぶべきでないという意見に同意するとともに、呼んでよいという考えも持っています。

その前にまず：第１種の過誤も過誤ではないのではないかという疑問があります。なぜなら「危険率」という逃げ道を作っているからです。そもそも統計学が過ちを犯しうるものだとすれば、統計学は自然科学の手法たり得ません。僕は統計学の試験問題に「統計解析手法によって検定結果が異なるとき、それは統計学が論理的に矛盾することを意味しているのではないか。論じなさい。」というのを出しました。初学者が学ぶべき最も基本的で重要な統計学の論理を、ここに学ぶことができると思ったからです。

上記の意味で、「過誤」と言う言葉は適切ではないと思います。が、小針さんの言わんとする所は、上記の論点とは異なるようですね。彼は、「第１種の過誤」と「第二種の過誤」の質的な相違を問題にしているように見受けられます。それは
> 「第２種の過誤」は「実在するとは言えない」という結論で,
> 「実在しない」とは言わない
によって端的に表現されています。初学者はこの点をくれぐれも誤解しないように、ということでしょうね。しかしながら、第１種の過誤に比べて第２種の過誤を過誤らしくないと感じさせる理由にはもう一つ、「何々効果が実在する」と言うに際して、客観性を失ってはならないという近代科学の基本的な姿勢があります。慎重過ぎただけである「第二種の過誤」は過誤でないというのが小針さんと服部君の主張であると思います。この主張に僕も異論はありません。

さてそれでもなお「過誤」と呼んでよいと僕が思うのは：検定結果に基づく誤った決断は人間の罪であって、統計学の罪ではないのですが、
少数の data をもとに母集団の性質を決定する完璧な方法は存在し得ないと言う意味で、
統計学は不完全であり、その意味でやはり人間に対して責任があると思うからです。
いかに客観的で整合的で正しい言明であっても、真理ではないということがあります。
%% 決して過ちを犯さない人が、理想の人ではない。

[12] nabe35: 19960127(2), Sun, 28 Jan 96 00:05:05 JST

# 参考文献


[広辞苑] 新村出編, 広辞苑第 4 版, 1991, 岩波書店.

[数学辞典] 日本数学会編, 岩波数学辞典第 3 版, 1985, 岩波書店.

[Billingsley] P. Billingsley, *Convergence of Probability Measures,* John Wiley & Sons, New York, 1968.

[Devroye] L. Devroye, *Non-Uniform Random Variate Generation,* Springer–Verlag.

[Knuth] D. E. Knuth, *The art of computer programming,* 2nd ed., vol. 2, (Addison-Wesley, Reading, MA, 1981).

[丸山] 丸山儀四郎, 確率および統計, 基礎数学講座, 共立出版.

[丸山演] 丸山儀四郎, 確率および統計, 数学演習講座, 共立出版.

[Ripley] B. D. Ripley, *Computer generation of random variables: a tutorial,* International Statistical Review **51** (1983) 301–319.

[服部] 服部哲弥, 宇都宮大学大学院博士前期課程数理解析特論（情報工学科）講義録 (1995).

[服部中島] T. Hattori, H. Nakajima, *Improvement of efficiency in generating random $U(1)$ variables with Boltzmann distribution,* Journal of Computational Physics **121** (1995) 238–245.

[林周二] 林周二, 統計学講義, 丸善.

[小針] 小針あき宏, 確率・統計入門, 岩波書店. 日見

[楠岡] 楠岡成雄, 確率と確率過程, 岩波書店（岩波講座応用数学 基礎 13）.

[楠岡 2] 楠岡成雄, 確率・統計, 森北出版.

[宮沢光一] よい本を書いているようだが, 未読. [楠岡 2], [林周二] が異なる本ながらそれぞれこの著者を引用している.

[森村木島] 森村英典・木島正明, ファイナンスのための確率過程, 日科技連.

[西尾] 西尾真喜子, 確率論, 実教出版.

[渡辺浩] 渡辺浩, 日本医科大学応用数学講義録 (1995).

[ウィルクス] ウィルクス著, 田中・岩本訳, 数理統計学, 増訂新版 1,2, 東京図書 (1972).

## 講義

統計数学 (CA108) 数学科２年　選択　通年　４単位　教職課程科目．担当：服部哲弥（６号館２階研究室）．成績追指示．講義の誤りを即座に指摘してくれた場合は加点する．複数参考書適宜詳講義.