数理統計学講義ノート

stat05.tex

服部哲弥

20040607–13;15;20–30;0701–02;04–06;0710;13;0911;28;20050815(誤植);

数理統計学

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.

目 次

1 不公平な硬貨 — 離散分布の確率論. 4

1.1 2 項分布. . . . 4
1.2 離散分布の平均と分散. . . . 5
1.3 2 項分布の中心極限定理. . . . 6

2 なぜ一列並びか — 連続分布と確率変数の分散. 8

2.1 JR 東日本 Y 駅みどりの窓口 1990 年. . . . 8
2.2 連続分布の平均と分散. . . . 8
2.3 確率変数の分布，期待値，分散，標準偏差. . . . 9
2.4 窓口の並び方と分散. . . . 10
エピローグ：Y 駅，その後. . . . 12
補足：確率測度の近代的定義と確率測度の定義域 σ 加法族. . . . 12

3 神のみぞ知る視聴率 — 確率変数の独立性と標本. 13

3.1 標本と確率変数. . . . 13
3.2 独立と相関. . . . 15
3.3 乱数. . . . 17
再考：標本と確率変数. . . . 18
再考：乱数. . . . 19
補足：意志決定支援手段としての統計学. . . . 20

4 宮城沖地震を逃れる確率 — 点推定. 20

4.1 標本平均と不偏分散. . . . 20
4.2 概収束と大数の強法則. . . . 22
4.3 正規分布と分布の収束と中心極限定理. . . . 23
4.4 例：宮城沖地震の発生間隔分布. . . . 24
補足：LLN, CLT，および特性関数. . . . 24

5 さいころの目は不公平か？ — 検定の原理. 26

5.1 （続）正規分布. . . . 26
5.2 検定の原理. . . . 28
5.3 さいころの公平性の検定. . . . 29
補足：命題 13 の証明. . . . 29
補足：第 2 種の過誤. . . . 31

6 視聴率調査，何人分調べれば十分か？ — 区間推定の原理と大標本理論. 32

6.1 区間推定の原理. . . . 32
6.2 例題 — 視聴率調査. . . . 33
6.3 例題 2 — 硬貨投げ，推定と検定. . . . 34
6.4 例題 3 — 事故，母集団がポワッソン分布の場合. . . . 35

7 パッケージ 1：1 つの正規母集団の推定・検定— χ^2 分布（分散）と t 分布（平均）. 37

7.1 χ^2 分布. . . . 37
7.2 分散の推定 39
7.3 t 分布 40
7.4 平均値の推定. . . . 41
7.5 例題. . . . 42
補足：定理 28 の証明. . . . 43

補足：命題 32 の証明. 44

8 パッケージ 2：正規母集団の比較の検定— F 分布（等分散の検定）と t 分布（等平均の検定）. 44
8.1 F 分布. 44
8.2 等分散の検定. 46
8.3 等平均の検定. 47
8.4 例題. 47
8.5 例題 2. 48

9 メンデルの法則に隠された真実 — 統計学その他の話題. 49
9.1 その他の話題. 49
9.2 χ^2 分布や F 分布を用いた種々の統計的推測. 51
9.3 その他の方法. 54

10 バスはなぜ来ないか — ポワッソン確率過程. 55
10.1 ばらつきか間引きか？ 55
10.2 ポワッソン分布. 56
10.3 練習問題. 57
10.4 ポワッソン確率過程と指数分布. 57
10.5 練習問題. 59
10.6 極端な仮想のケース — でたらめに到着するバス. 60
10.7 練習問題. 61

11 宮城沖地震発生期間の分布 — ブラウン運動の脱出時刻. 62

参考文献. 64

1 不公平な硬貨 — 離散分布の確率論.

硬貨投げ（2 項分布）を例として，離散分布について，確率測度，平均，分散の定義を復習する.

1.1 2 項分布.

硬貨を n 回投げる．表裏は公平には出ないが，それ以外の点ではいかさまではないとする．どの回も表の出る確率が p であるとする．確率が負というのはあり得ないので $0 < p < 1$．$p = 0$ は必ず裏の出る硬貨，$p = 1$ は必ず表の出る硬貨.（入れておいても最初のうちは問題ないが，確率論で扱う必要もないので最初から除いておく.）

注 1 「それ以外の点でいかさまでない」という言葉の一般的定義も問題だが，ここではひとまず「硬貨の場合はこうする」という言い方にしておく．念頭にあるのは，各回の試行の「独立性」ということ（具体的には *(1)* の式が成り立つこと）.

一方，その「独立性」を現実の賭の場面でどうやって統計学的に検証するかは大問題．検証方法（そもそも何を検証しうるのか）を考えることは統計学の目的の一つ．ここでは，その議論は置いておいて，単に「以下の議論や公式が成り立つようになっている」とする. ◇

全体集合 Ω

高校で習ったように，確率を定義するにはまず全体集合 Ω，起こりうることの全て，を念頭に置く．実は,「正しい全体集合の取り方」はとてつもなく自由であって，現代的な確率論ではそのことが議論を簡潔にする面があるが，その議論は省略して，今の場合は一つの自然な取り方として n 回の表裏の出方の列全体，をとる．スペースの節約のため表を 1 裏を 0 と書くことにすると，0, 1 の長さ n の列の集合 $\Omega = \{(s_1, s_2, \cdots, s_n) \in \{0,1\}^n\}$ を全体集合にとれる.

1 回あたり，表の出る確率を p として，異なる回では積で計算できるとする（これが，各回が独立ということであり，表の出る確率以外の意味では公平，ということの意味とする）．表 1 枚当たり p，裏 1 枚当たり $1-p$ なので，Ω 上の確率測度は

$$\mathrm{P}[\ (s_1, s_2, \cdots, s_n)\ ] = p^{s_1+\cdots+s_n}(1-p)^{n-(s_1+\cdots+s_n)}. \tag{1}$$

確率変数とは Ω 上の関数

さて，n 回の表裏の出る順番もたいへん興味のある問題（確率連鎖，特にランダムウォークの問題）だが，やさしい問題から始めることにして，投げた n 回のうち表の回数 k の分布に注目する．数学的に言うと，確率 P の定義されている集合 Ω を定義域とする関数を確率変数という．$w \in \Omega$ ごとに表の枚数はもちろん決まるから「表の枚数」は Ω 上の関数である．これを $N_n : \Omega \to \{0, 1, \cdots, n\}$ と書くと，P の下で N_n の値がどう分布するかに興味がある．このことを P における N の分布などと呼び，たとえば $\mathrm{P} \circ N_n^{-1}$ と書く．$Q_n = \mathrm{P} \circ N_n^{-1}$ とおくと k 回表が出る確率は

$$Q_n(\{k\}) = \mathrm{P}[\ N_n^{-1}(k)\ ] = \mathrm{P}[\ N_n = k\ ]. \tag{2}$$

ここで P は Ω 上の確率だから Ω の部分集合を引数に持つはずだが，$N_n = k$（表が k 枚）であるような要素の集合（事象）を

$$\{w = (s_1, \cdots, s_n) \in \Omega \mid N_n(w) = k\} = \{N_n = k\}$$

のように Ω の要素を省略して書くことが多い．当然その事象の確率を $\mathrm{P}[\ N_n = k\ ]$ と書くことになる.（「表の枚数 N が k となる確率」と素朴に素直に読めるので，この書き方は確率論らしくて良い，と個人的には思う.）

確率変数の分布

関数（確率変数）N_n は具体的に

$$N_n((s_1, \cdots, s_n)) = s_1 + \cdots + s_n$$

と書けることはちょっと考えると分かるので，(1) を使って，

$$\begin{aligned} Q_n(\{k\}) = \mathrm{P}[\ N_n = k\ ] &= \sum_{\substack{s_1,\cdots,s_n\in\{0,1\};\\ s_1+\cdots+s_n=k}} \mathrm{P}[\ \{(s_1, \cdots, s_n)\}\ ] = \sum_{\substack{s_1,\cdots,s_n\in\{0,1\};\\ s_1+\cdots+s_n=k}} p^k(1-p)^{n-k} \\ &= p^k(1-p)^{n-k}\sharp\{s_1,\cdots,s_n\in\{0,1\}|s_1+\cdots+s_n=k\}. \end{aligned} \tag{3}$$

最後の集合の要素の数は n 枚のうちちょうど k 枚が表に出る場合の数だから ${}_n\mathrm{C}_k$ に等しいので，結局

$$Q_n(\{k\}) = {}_n\mathrm{C}_k p^k (1-p)^{n-k},\ k = 0, 1, \cdots, n-1, n, \tag{4}$$

となる．この確率測度を 2 項分布と呼び $B_{n,p}$ と書く．

添え字は n と p を決めるごとに一つの確率測度が決まることから，n と p をパラメータとする確率測度の族である（似たような確率測度を集めている）ことを表す．

1.2 離散分布の平均と分散．

（不）公平な硬貨を n 回投げたとき k 回表が出る確率（2 項分布）を (4) で定義した．2 項分布は k のとりうる $n+1$ 個の値それぞれに確率 $Q(\{k\}) = Q_n(\{k\})$ が与えられている．そして一般の事象（すなわち，集合 $A \subset \{0, 1, \cdots, n\} = \Omega$）の確率は

$$Q(A) = \sum_{k \in A} Q(\{k\}) \tag{5}$$

で与えられる．この定義は，排反事象の和集合の確率は各事象の確率の和に等しい（確率の加法性），という直感的な確率のイメージに合っている．なお，確率はこのように集合に対して非負の数字を与えるのが良い定義であることが分かっている．確率の定義されている集合のことを事象と言う．(今のところ Ω の全ての部分集合に対して確率を定義しているので事象という言葉と集合という言葉は同一である.）数学的には (2) と (5) が確率測度（分布）$Q = Q_n$ を定義している，ということになる．

確率の定義
事象の集まりで
定義された
非負加法的関数
$Q(\Omega) = 1$

一般に，Ω が有限集合で，仮にそれを（2 項分布のときのように）$\Omega = \{0, 1, \cdots, n\}$ と記号化して書くとき，非負の値 $Q(\{k\}), k \in \Omega$, が定義されていて，

$$\sum_{k=0}^{n} Q(\{k\}) = 1 \tag{6}$$

であり，任意の $A \subset \Omega$ に対して $Q(A)$ が (5) で定義されているとき，Q を Ω 上の離散分布という．

高校の教科書ではこのときの $\{k\}$ たちを根源事象と呼んでいた．

2 項分布は離散分布の例である（正しく言うと，n と p を決めるごとに一つの離散分布を与えるので，(4) は無数の離散分布の例を与えている）．2 項分布以外の例としてはポワッソン分布がよく知られている (§10.2)．

さて，確率は（離散分布で言えば $Q(\{k\})$ のように）たくさんの数字（確率）を与えないと決まらないが，日常生活でことあるごとに多数の数字を列挙するのは不便である．通常最初に必要になるのはだいたい何枚くらい表が出るかという「目安」だが，理論上最も重要で応用上も最も頻繁に用いられる目安は（ご承知の通り）確率測度（分布）の平均である．

平均と分散
（離散分布）

離散分布の平均は，高校時代に習ったとおり，

$$\mu = \sum_{k \in \Omega} k\, Q(\{k\}) \tag{7}$$

で定義される．

分布と言うことは平均という一つの数字だけでは表せない，ということである．n 枚の硬貨の内の表の枚数は常に平均という一つの値が出るのではなく，ときによってばらつく．ばらつきがあってこそ確率や統計らしい，と言える．最も頻繁に用いられる理論上も重要なばらつきの目安は分散であり，(これも高校時代に習ったとおり)

離散分布の分散の定義は

$$v = \sum_{k\in\Omega} (k-\mu)^2\, Q(\{k\}). \tag{8}$$

(8) の 2 次式を展開して変形すると，

$$v = \sum_{k\in\Omega} k^2\, Q(\{k\}) - \mu^2 \tag{9}$$

もすぐ分かる．

以上の離散分布の一般論を 2 項分布 (3) の場合に具体的に計算する．個別の例では個別の計算方法が必要になる．2 項分布の場合は以下のように計算するのが一番楽である．

いったん 2 項分布のことは忘れて，公式をいくつか用意する．x, y を実数，n を自然数とすると，高校時代に習ったように 2 項定理

$$(x+y)^n = \sum_{k=0}^{n} {}_n\mathrm{C}_k x^k y^{n-k} \tag{10}$$

が成り立つ．x, y は任意なので特に $x=p$, $y=1-p$ とおくと

$$\sum_{k=0}^{n} {}_n\mathrm{C}_k p^k (1-p)^{n-k} = 1$$

となる．これは 2 項分布 $B_{n,p}$ の定義 (3) が正しく確率を定義していること（具体的には $Q_n(\Omega)=1$）を示す．次に (10) を（y を固定して x だけの関数と思って）x で微分して x をかけると

$$nx(x+y)^{n-1} = \sum_{k=0}^{n} {}_n\mathrm{C}_k k x^k y^{n-k}. \tag{11}$$

x, y は任意なので特に $x=p$, $y=1-p$ とおくと

$$\sum_{k=0}^{n} {}_n\mathrm{C}_k k\, p^k (1-p)^{n-k} = np.$$

(7) と比べると $\mu = np$ が分かる．最後に (11) を再度 x で微分して x をかけると

$$nx(x+y)^{n-1} + n(n-1)x^2(x+y)^{n-2} = \sum_{k=0}^{n} {}_n\mathrm{C}_k k^2 x^k y^{n-k}. \tag{12}$$

x, y は任意なので特に $x=p$, $y=1-p$ とおくと

$$\sum_{k=0}^{n} {}_n\mathrm{C}_k k^2\, p^k (1-p)^{n-k} = np + n(n-1)p^2.$$

上で導いた $\mu = np$ と (9) を用いると $v = np + n(n-1)p^2 - (np)^2 = np(1-p)$ を得る．まとめると，

2 項分布 $B_{n,p}$ の平均と分散は $\mu = np$ と $v = np(1-p)$ で与えられる．

1.3 2 項分布の中心極限定理．

ところで §1.1 で 2 項分布の度数分布表をいくつかの n について（適当に縮尺を変えながら）掲げたが，n が大きくなるほど，正規分布の密度関数のグラフに似ていくように見える．この観察は正しくて，次の事実が成り立つ．

定理 2 (2 項分布の局所中心極限定理) $0 < p < 1$ とし，Q_n を $B_{n,p}$ とする．実数 y に対して $[y]$ を y 以下の最大の整数として $k_n(x) = [np + x\sqrt{np(1-p)}]$ とおくとき

$$\begin{aligned}&(\lim_{n\to\infty} \sqrt{np(1-p)}\, Q_n(\{k_n(x)\}) =)\\ &\lim_{n\to\infty} \sqrt{np(1-p)}\, {}_n\mathrm{C}_{k_n(x)}\, p^{k_n(x)}(1-p)^{n-k_n(x)} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}.\end{aligned} \tag{13}$$

◇

($[y]$ を y 以下の最大の整数としたが，この定理では y の隣の整数でありさえすれば極限を取ると違いは見えなくなる．また，$Q_n(\{k\})$ は $|k| \leqq n$ でしか定義されていないが，左辺の数列は大きい $|x|$ に対しては対応して十分大きい n から始まる数列を考えていることにする．または，$k < 0$ や $k > n$ のとき $Q_n(\{k\}) = 0$ として $\Omega = \mathbb{Z}$ の上の確率測度と思っていると考えても差し支えない.)

定理 2 の証明は，Stirling の公式

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n} n^n e^{-n}} = 1, \tag{14}$$

および，よく知られた e の定義

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

と，その拡張

$$\lim_{n\to\infty} e^{-x\sqrt{n}} \left(1 + \frac{x}{\sqrt{n}}\right)^n = e^{-x^2/2}$$

を知っていれば，通常の数列の極限の計算なので，練習問題として省略する．なお，Stirling の公式は，分母は $n = 1$ でも誤差は 8.5% 以内であり，$n!$ の近似式として理論上も実用上も頻繁に用いられる．(14) の証明は，たとえば [小針, 命題 5.3].

(13) の右辺の関数は標準正規分布 $N(0,1)$ と呼ばれる連続分布の密度関数である．この文の意味は次節 §2 の主題の一つである．

$Q_n(\{k_n(x)\})$ が（適当なスケール変換で）収束するという局所中心極限定理は 2 項分布のように比較的性質の良い分布に対して成り立つもので，証明も一般に込み入っているが，x の区間（一般に事象）の確率の収束（これを精密な数学用語では分布の弱収束という）は非常に広い範囲の確率測度に対して成りたつ．これを中心極限定理と言って，正規分布（secunormaldistri）の理論上及び実用上の重要性の根拠である． 中心極限定理

定理 3 (2 項分布の中心極限定理) 確率 p で表が出る硬貨を n 回投げて表の回数 N_n とすると $\dfrac{(N_n - np)}{\sqrt{np(1-p)}}$ の分布は $n \to \infty$ のとき標準正規分布 $N(0,1)$ に弱収束する．すなわち（正規分布のように極限分布が密度を持つ連続分布の場合の同値な言い換えとして）任意の実数の区間 $[a,b]$ に対して

$$\lim_{n\to\infty} Q_n(a \leqq \frac{N_n - np}{\sqrt{np(1-p)}} \leqq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

が成り立つ． ◇

定理の結論の左辺は (2) から

$$\lim_{n\to\infty} \frac{1}{\sqrt{np(1-p)}} \sum_{k;\ a \leqq \frac{k-np}{\sqrt{np(1-p)}} \leqq b} \sqrt{np(1-p)}\, Q_n(\{k\})$$

となるので，(13) から（ほぼ）定理 3 が得られそうだと分かる．実際は定理 3 は分布の特性関数を用いることで，局所中心極限定理を経由せず，また，2 項分布に限らずもっと一般の分布に対して統一的に証明できる．

2 なぜ一列並びか — 連続分布と確率変数の分散.

平均値は日常普通に用いられる量であるが，これだけですむなら分布を考える必要はなく平均値を確定値として話をすればよい．この節では平均の次に重要な分布の目安である分散を窓口における一列並びという例について取り上げ，なぜ平均だけでは不十分で値が散らばる量，分布，を考えないといけないかを説明する．また，次節 §3 以降で調査や観測や実験結果のデータを確率変数と見ることで確率論を適用するのに備えて確率変数を，前節 §1 の離散分布の対義語である連続分布とともに，定義して，その平均や分散を定義する．

2.1 JR 東日本 Y 駅みどりの窓口 1990 年.

窓口に待ち行列はつきものである．JR みどりの窓口や銀行などの ATM，飛行機のチェックインカウンター，公衆トイレやスーパーのレジ，など[1]，窓が本当にあるかどうかが重要ではなくてここで注目したい特徴は，処理のための複数の地点があり，それを目指して順番に並ぶ行列がある，という 2 点である．窓口が複数ある時，各窓にそれぞれ待つ方法（以下では仮に並列並び，と名付けておく）と，待ち行列は一列で順次開いた窓へ進む一列並びがある．昔（1980 年代頃までの日本）は並列並びが普通だった．1990 年頃,「欧米に習え」といって一列並びが宣伝されるようになり，最近はスーパーのレジを除けば一列並びが普通である．ではなぜ一列並びが有利だろうか？

現在でこそ当たり前に見られる一列並びも，宣伝され始めた 1990 年代は，いったん広まって，スペースのゆったり取れる場所（飛行場や銀行）ではそのまま定着したが，それ以外では元の並列並びに戻すなどの混乱も見られた．実際，JR 東日本の Y 駅のみどりの窓口（指定券前売り窓口）は，昔は並列並びだったのを，一列並びが話題になってしばらく後 1990 年始めに一列並びに変えた．ところが何ヶ月もたたないうちに一列並びをやめて並列並びに戻した．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 点は平均値という統計学の問題である.「平均が変わらないから一列並びには意味がない」という主張にみえるが，これは適切な理由とは言えない．**待ち時間の平均は変わらなくても，一列並びには統計学的に意味がある**ことをこの節 §2 で論じたい.

2.2 連続分布の平均と分散.

一列並びの問題を論じるには待ち時間の分布の議論が必要である．離散分布とは対照的に時間は取りうる値が連続的なので，§1.2 で紹介した離散分布の定式化が使えないことはすぐ分かるだろう．取りうる値が連続的な母集団の上の確率測度（分布）は連続分布が必要である[2].

> この講義で「連続分布」と書いた場合は全て「ルベーグ測度に対して絶対連続な分布」の意味と了解していただきたい.

そこで一列並びの本題に入る前に密度を持つ連続分布を紹介する.

全体集合 Ω

離散分布の場合に §1.1 で指摘したように，確率を定義するにはまず全体集合 Ω を与える．密度を持つ連続分布は，典型的には実数の集合 $\mathbb{R}$ が Ω である．時間は負にならないので非負実数の集合 $\mathbb{R}_+$ が Ω になる．取りうる値が区間 $[a,b]$ に限られている場合もある．いくつかの実数の組の分布が問題になるときは $\Omega=\mathbb{R}^n$

[1] 1995 年頃に用意した講義ノートには，公衆電話ボックス，という例も挙げていたが，携帯電話の爆発的普及によって公衆電話に並ぶ待ち行列の場面はなくなってしまった.
[2] 近代確率論の言葉で言えば，これから紹介しようとしているのは，ルベーグ測度に対して絶対連続な確率測度，である．ただ，そこまで精密な数学の話はとてもできない.

も未 s つどを持つ連続分布が定義される全体集合になり得て，n に応じて n 次元分布と呼ぶ．たとえば英語と数学の試験の成績を組にして考える，身長と体重の関係を調べるなど．

密度を持つ連続分布は §1.2 の (5) とは異なり根源事象の和で表すことはできない．$\Omega = \mathbb{R}$, $\mathbb{R}_+$, $[a,b]$ などの上の分布である 1 次元分布の場合，代わりに密度関数 $\rho\colon\ \Omega \to \mathbb{R}_+$ があって，

$$Q(A) = \int_A \rho(x)\,dx \tag{15}$$

と書けるものを言う．ただし，ρ は非負値関数

$$\rho(x) \geqq 0,\ x \in \Omega, \tag{16}$$

で

$$\int_\Omega \rho(x)\,dx = 1 \tag{17}$$

を満たす区分的に連続な関数とする．(ルベーグ積分論によれば ρ はより複雑な関数でもかまわないが，この講義はそこまで踏み込まない.)

一般に n 次元分布は n 変数関数で同様の表示ができるものを言う．たとえば 2 次元分布は

$$\rho(x,y) \geqq 0,\ (x,y) \in \mathbb{R}, \tag{18}$$

および

$$\int_\Omega \rho(x,y)\,dx\,dy = 1 \tag{19}$$

を満たす区分的に連続な 2 変数関数 $\rho\colon\ \mathbb{R}^2 \to \mathbb{R}_+$ があって，

$$Q(A) = \int_A \rho(x,y)\,dx\,dy \tag{20}$$

と書けるものを言う．

(15) から特に $A = \{x\}$ のような

1 点だけからなる集合は連続分布では確率 0 となる：

$$Q(\{x\}) = \int_x^x \rho(y)\,dy = 0. \tag{21}$$

これは離散分布と著しい違いである．

平均と分散（連続分布）

離散分布（§1.2）と同様に連続分布も平均 μ と分散 v が分布の目安として重要である．密度を持つ連続分布の場合は (7) と (8)，(9) に対応するのは

$$\mu = \int_\Omega x\,\rho(x)\,dx \tag{22}$$

および

$$v = \int_\Omega (x-\mu)^2\,\rho(x)\,dx = \int_\Omega x^2\,\rho(x)\,dx - \mu^2. \tag{23}$$

また $\sigma = \sqrt{v}$ を標準偏差という．

定義から平均は分布の「中ほど」，標準偏差は $|x-\mu|$，つまり平均からの分布の「広がり」を表す目安になる．

2.3 確率変数の分布，期待値，分散，標準偏差．

E[X], V[X]

(2) で導入したように，一般に確率変数（関数） $X\colon\ \Omega \to \mathbb{R}$ に対して，

$$Q(A) = (\mathrm{P} \circ X^{-1})(A) = \mathrm{P}[\ X^{-1}(A)\ ] = \mathrm{P}[\ X \in A\ ] \tag{24}$$

を X の分布と呼ぶ．分布 Q に従う確率変数 X などという言い方もする．これは P が離散分布でも連続分布でもかまわないし，関数 X の取りうる値が整数のように離散的でも実数のように連続的でもかまわない．X の取りうる値が離散的ならば X の分布 Q は離散分布に，X が実数値をとるなら X の分布 Q は連続散分布に，なる．

X の期待値と分散をそれぞれ $\mathrm{E}[\,X\,]$, $\mathrm{V}[\,X\,]$ と書く．X の取りうる値が有限個，つまり X の分布 $Q = \mathrm{P}\circ X^{-1}$ が離散分布の場合は，たとえばその値が $\{0,1,2,\cdots,n\}$ で尽くされるとすると

$$\mathrm{E}[\,X\,] = \sum_{k=0}^{n} k\mathrm{P}[\,X=k\,] = \sum_{k=0}^{n} kQ(\{k\}) \tag{25}$$

が定義である．取りうる値が可算無限個でも右辺の和を級数に置き換えるだけで同様の式が定義である．X が実数値確率変数でその分布 $Q = \mathrm{P}\circ X^{-1}$ が密度 ρ を持つ連続分布ならば

$$\mathrm{E}[\,X\,] = \int_{\mathbb{R}} x\rho(x)\,dx \tag{26}$$

となる．(離散近似して (25) の極限をとることで得られるが，積分論に立ち入ることになるので省略する．) (7), (22) と突き合わせると，$\mathrm{E}[\,X\,]$ とは X の分布の平均のことである．

近代確率論を避けて通ったため期待値の定義は場合によって見かけが違うが，ひとたび期待値を定義すると分散 $\mathrm{V}[\,X\,]$ は X が離散値か連続値などにいっさい関係なく

$$\mathrm{V}[\,X\,] = \mathrm{E}[\,(X-\mathrm{E}[\,X\,])^2\,] \tag{27}$$

で定義される．X の分布 $Q = \mathrm{P}\circ X^{-1}$ が離散分布のときは，(25) から Q の平均は $\mu = \mathrm{E}[\,X\,]$ なので，(25) の関数 X を関数 $(X-\mu)^2$ に置き換えると

$$\mathrm{V}[\,X\,] = \sum_{k=0}^{n} (k-\mu)^2 Q(\{k\})$$

となって，(8) から，$\mathrm{V}[\,X\,]$ は X の分布 Q の分散 v のことである．X の分布 $Q = \mathrm{P}\circ X^{-1}$ が離散分布のときも，(26), (22), (23) からやはり $\mathrm{V}[\,X\,]$ は X の分布 Q の分散 v のことである．

期待値と分散の定義から，確率変数ではない定数 a に対して

$$\mathrm{E}[\,aX\,] = a\mathrm{E}[\,X\,],\ \mathrm{V}[\,aX\,] = a^2\mathrm{V}[\,X\,]. \tag{28}$$

また，確率変数 X, Y に対して

$$\mathrm{E}[\,X+Y\,] = \mathrm{E}[\,X\,] + \mathrm{E}[\,Y\,]. \tag{29}$$

分散は a^2 になっていることに注意．期待値の線型性と合わせるためにしばしば分散の代わりに標準偏差 $\sigma_x = \sqrt{\mathrm{V}[\,X\,]}$ が用いられる．期待値は加法性 (29) を持つが分散では条件が必要である（§3.2）．また，定義から当たり前だが，特に

$$\mathrm{E}[\,1\,] = 1,\ \mathrm{V}[\,1\,] = 0. \tag{30}$$

2.4 窓口の並び方と分散．

平均と分散が確率分布の目安だ，と言ってきた．平均は，たとえば，保険料を決めるとき，一人当たり平均いくら年金や入院費用を払わなければならないかを決める場合など，その実用上の役割は直感的に多くの人が把握していると思う．では分散の実用上の意味は何か？その一例が §2.1 で提示した一列並びの問題である．

最初に，平均値に関する Y 駅の主張は正しいこと，つまり，平均だけを考えるならば，一列並びにする必要はないこと，を確認する．

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（本質を損なわない範囲で簡単化したモデルを考える）．

(i) 窓口の数を M，自分が待ち行列の N 人目，すなわち，自分が到着する直前に $N-1$ 人が待っていたとする．M に比べて N は十分大きいとする．特に，窓口に空き時間はなく，全ての窓口がいつも仕事をしているとする．

(ii) 「一列並び」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，その空いた窓口へ順に客が行って処理を受ける.「並列並び」ではどの窓口も均等に並ぶとする．すなわち M 個の窓口それぞれに N/M 人ずつ並び，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る[3]．

(iii) 客 i $(i=1,2,\cdots,N)$ の処理時間 S_i は確率変数である[4]．

(iv) $\{S_i\}$ は独立とし，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とする．各々の客を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いうこと．特に平均処理時間 $\tau=\mathrm{E}[\,S_i\,]$ は客によらず一定とする．我々は処理時間の分布を熟知してはいないが，少なくとも平均処理時間 τ は経験的に知っているとする[5]．

(v) 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．この他一列並びと並列並びは上で書いた窓口の選び方以外の条件は変わらないとする．

一人の客の処理をするのに平均 τ の処理時間を要するとしたので自分が 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．一列並びでも並列並びでも常時 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 $\overline{T}$ は $\overline{T}=\dfrac{N}{M}\tau$ となり，並び方によらない．すなわち平均値に関しては Y 駅の主張は正しい．** なお，τ を経験的に知っていると仮定したので，我々は平均待ち時間 $\overline{T}$ も経験的に分かる[6]．$\overline{T}$ は等しくても一列並びと並列並びは実際に体験してみると何かが違う．違いのありかを示すために次の点に注目する．

前売り指定券を買う人は，次の約束まで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おうと考えている．サラリーマンにとって，前売り券を買うために休暇をとって一日並ぶ覚悟，というのは盆や正月の帰省など特殊な場合であろう．そこで，電車の発車や次の約束・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 $\overline{T}$ に比べて長ければ $(t_0>\overline{T})$ 間に合うと判断して前売り券を買うために並ぶ．実際の待ち時間 T は平均 $\overline{T}$ の周りにばらつく確率変数である．買えることもあるが，たまたま窓口が前の客に手間取って買えない場合もある．$T<t_0$ が実現すれば券を買えるが，$T>t_0$ 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ければいけない．

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である．平均待ち時間 $\overline{T}$ が等しくても待ち時間 T の分布によって「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，$t_0>\overline{T}$ のとき，分布が広がっているほど $\mathrm{P}[\,T>t_0\,]$ は大きい．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

[3]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．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が，スーパーの常連とおぼしき人たちは驚くほど正しく「でこぼこ」に並ぶ．彼らは極めて複雑なモデルを瞬時に解いているのだ！

[4]確率変数という言葉は (2) で導入した．実際に測定すれば分かるが，客によって指定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．詳しいことは §2.3 で説明するが，S_i が確率変数であるとは，適当な確率空間 (Ω,P) があって S_i は関数 $S_i:\ \Omega\to\mathbb{R}$ であることを意味する．Ω 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 の全体を意味し，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．

[5]§3.2 の紹介を先取りすると，$S_1,\,S_2,\,\cdots,\,S_n$ が独立とは，$\mathrm{P}[\,S_1\in A_1,\ S_2\in A_2,\ \cdots,\ S_n\in A_n\,]=\prod_{i=1}^{n}\mathrm{P}[\,S_i\in A_i\,]$ が任意の事象列 A_i, $i=1,\cdots,n$, に対して成り立つことをいう．この定義から，$i\neq j$ ならば $\mathrm{E}[\,S_iS_j\,]=\mathrm{E}[\,S_i\,]\mathrm{E}[\,S_j\,]$ のように独立な確率変数の積の期待値は期待値の積になる．

[6]スーパーのレジの脚注に書いたように,（筆者以外の）人々は驚くほど頭を使っている．

れる[7]．平均ではなく，ある日の待ち時間そのものを表す確率変数は，モデルの定義から

$$\begin{aligned} &\text{一列並びの場合は}\quad T^{(1)} = \frac{1}{M}\sum_{i=1}^{N} S_i\,, \\ &\text{並列並びの場合は}\quad T^{(2)} = \sum_{i=1}^{N/M} S_{j_i}\,. \end{aligned} \tag{31}$$

ここで 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，あるいは $\mathrm{E}[\,T^{(1)}\,] = \mathrm{E}[\,T^{(2)}\,] = \overline{T}$ ，すなわち待ち時間の期待値（平均値）は等しい．

処理時間の分布は客によらないという仮定から，客一人当たり窓口当たりの処理時間の分散 $\mathrm{V}[\,S_i\,]$ は i によらないので，$\sigma^2 = \mathrm{V}[\,S_i\,]$ とおく．このとき，

$$\begin{aligned} \mathrm{V}[\,T^{(1)}\,] &= \frac{N}{M^2}\,\sigma^2\,,\ \text{一列並びの場合,} \\ \mathrm{V}[\,T^{(2)}\,] &= \frac{N}{M}\,\sigma^2\,,\ \text{並列並びの場合,} \end{aligned} \tag{32}$$

となる．この証明には §3.2 の (43) が必要である．ここでは (43) を認めて $\mathrm{V}[\,T^{(1)}\,]$ について (32) を証明しておこう．

$\mathrm{V}[\,T^{(1)}\,] = \dfrac{N}{M^2}\,\sigma^2$ **の証明．** (43) にあるように，確率変数列 $\{S_i \mid i = 1, \cdots, N\}$ が独立ならば $\mathrm{V}[\,S_1 + S_2 + \cdots + S_n\,] = \mathrm{V}[\,S_1\,] + \mathrm{V}[\,S_2\,] + \cdots + \mathrm{V}[\,S_n\,]$ が成り立つ．この事実と (28) と (31) から，

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

$M > 1$ だと $\mathrm{V}[\,T^{(1)}\,] < \mathrm{V}[\,T^{(2)}\,]$ となるから，一列並びが並列並びに比べて，待ち時間の分散，すなわち散らばり具合が小さい．それゆえ一列並びのほうが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.）**並んでいられる時間）が限られているときに，一列並びのほうが並列並びに比べて買い損なう恐れが小さい．**

JR 東日本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点で優れている．しかし，当時は，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．

エピローグ：Y 駅，その後．

Y 駅では 2000 年頃から一列並びを再開して，現在では定着している．

補足：確率測度の近代的定義と確率測度の定義域 σ 加法族．

(16) （2 次元なら (18)，以下この種の注を省略して 1 次元分布の場合だけを書く）から確率が非負：

$$Q(A) \geqq 0,\quad \text{任意の事象 } A \text{ に対して,} \tag{33}$$

を得る．(17) から全確率は 1：

$$Q(\Omega) = 1. \tag{34}$$

（むしろ，(17) を要求するのはこのため，というのが正しい順序.）最後に積分の定義から

$$A \cap B = \emptyset \text{ ならば}\quad Q(A \cup B) = Q(A) + Q(B), \tag{35}$$

[7] 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．それは §5.1 で正規分布について講義した後の各自の自習に任せよう．

すなわち有限加法性が成り立つ．もちろんこれを繰り返せば任意の自然数 n に対して

$$A_i \cap A_j = \emptyset,\ i \neq j,\ 1 \leqq i,j \leqq n, \quad \text{ならば} \quad Q(\bigcup_{k=1}^{n} A_k) = \sum_{k=1}^{n} Q(A_i). \tag{36}$$

実は，自然数で番号づけられる無限個（可算無限個）の事象列の場合も，右辺の和を極限の意味にとることで，同様の式（可算加法性，σ 加法性）が成り立つ（正確に言うと，成り立つ，と要請して矛盾が起きないことが知られている）：

$$A_i \cap A_j = \emptyset,\ i \neq j,\ i,j \in \mathbb{N}, \quad \text{ならば} \quad Q(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} Q(A_i). \tag{37}$$

右辺は級数の和の意味で，(33) と (34) から単調非減少上の有界だから必ず和がある（収束する）ことに注意．左辺の「無限和」はどれかの A_k に入っている要素を全て集めた集合，の意味．たとえば，全ての有理数を集めた集合 $\mathbb{Q}$ は（$\mathbb{Q}$ は自然数で番号づけられる，すなわち，一列に並べられる，ことが分かっているので）(21) と (37) から

$$Q(\mathbb{Q}) = 0. \tag{38}$$

確率の定義
正値性
全確率 1
σ 加法性

実は，ここまでの話の順序を逆にして，以上の (33)，(34)，(37) の 3 条件を満たす集合関数 Q を確率（確率測度）と呼ぶ，というのが近代確率論における確率測度の定義である．(33)，(34)，(37) は離散分布でも成り立つので，この 3 条件から始めれば離散分布と連続分布を統一的に扱える．しかも，これらの，一見当たり前すぎて何も出なさそうな定義だけで，許される確率は案外素直なものに限られることも分かっている．もっと大事なことは，このように一般的な定義にすることで $\mathbb{R}^n$ 上の確率だけでなく，もっと巨大かつ抽象的な，たとえば関数の集合上の確率，といったものを数学的に扱えるようになる．議論を統一的に無駄なく展開できる上に，高度な分析が可能になるのでお薦めだが，おもしろい具体例に行き着くまでに勉強するべきことが多いのが難点である．

確率 Q の定義域（事象，つまり，確率が定義できる集合，の集まり）を後回しにしていた．(15) や (20) の積分範囲 A が事象だが，Ω の任意の部分集合を積分範囲として積分ができるとは限らない．そこで「積分が定義できる集合」のみを集めた集合族，Q の定義域，をとりあえず $\mathcal{F}$ とおき，$\mathcal{F}$ に入っている（確率が定義できている）集合を事象と呼ぶ．どんな集合が事象なのか（確率が存在しうる）分からないと言っても，最低限次のような集合は明らかに定義域に入る：

事象とは
確率が定義
できる集合

$\Omega \in \mathcal{F}$ （全体集合の確率は 1）．
$\mathbb{R},\ [\alpha,\infty),\ (-\infty,\beta],\ [\alpha,\beta] \in \mathcal{F}, \quad \alpha,\beta \in \mathbb{R}$ （区間上で積分できる）．
$A_k \in \mathcal{F},\ k \in \mathbb{N}, \quad \text{ならば} \quad \bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ （事象であることが分かっている有限個または可算無限個の集合があれば，(37) からその和集合も事象になる）． (39)
$A \in \mathcal{F}$ ならば $A^c \in \mathcal{F}$（事象 A があれば補集合 A^c は，確率が足して 1 になるように決まるから，事象になるべきである）．

実は (15) のような積分で定義される $\mathbb{R}$ 上の確率測度の定義域 $\mathcal{F}$ は (39) を満たす集合族にとればよい，というのが近代確率論の定式化である．そのうち一番小さい集合族を Borel σ 加法族と呼んで，$\mathcal{B}$ 等とも書く．なぜこのような当たり前に見えることできちんとした数学になるか説明するのは時間がかかるのでここではこれ以上は立ち入らない．（近代確率論という言葉で全てをごまかしているわけではないことだけ分かってもらえればよい．）

3 神のみぞ知る視聴率 — 確率変数の独立性と標本．

この節では，母集団から無作為抽出した標本（データ）から確率論に基づいて母集団についての統計的推測を行うという，この講義で紹介する統計学の目的を説明する．複数回の無作為抽出，再調査，追加観測，実験の追試などは独立確率変数列（の標本）と理解するので，確率変数の独立性を定義する．併せて対義語である相関も便宜上この機会に定義しておく．無作為抽出の典型例かつ補助手段としての乱数も紹介する．

3.1 標本と確率変数．

§1 は「不公平な硬貨の確率論」という表題から期待するものと比べて抽象的かもしれない．（不）公平な硬貨を n 回投げたとき k 回表が出る確率を (4) で定義したが，現実に目に見えるのは，ある特定の硬貨の裏表の列（そこでの記号で言えば 0 と 1 の列）であって，たとえば表の枚数 k は硬貨を投げたときの実際の結果として一つの値に決まる．k は平均 μ に等しいと限らず，ばらつく．そのばらつきから計算される分散と v も一致しない．表の枚数が 2 項分布をすると言っても分布や平均や分散は実際に見ることはできな

い．当然「表が出る確率」p も正確に知る方法はない．これらをあたかも既知のように説明した §1 は現実離れしている．

値がばらつく現象の分布や平均などの重要な特徴を調べるとき，人は**調査や観測や測定や実験**を行い，俗にデータと呼ばれる数値の列を得る．国勢調査のように全数調査を行う場合もあるが，殆どの場合，起こりうるばらつきの全てを集めず，ごく一部を取り出してそこから分布の全体像や主要な性質を推測する．取り出された部分は特別な代表というよりは偶然そのときその順序で出ただけ，という理解の下に処理される点が特徴である．人為を加えない偶然の値のばらつきであることを生かして少ないデータから本来の分布を予測しようというデータの集めかたを，調査の場合には**無作為抽出 (random sampling)** と呼ぶ．自然科学の実験では実験を繰り返してデータ列を得るが，このとき現れるばらつきは（背後の自然法則そのものが統計力学的現象や量子力学的現象のように確率的でない限り）実験装置の精度などの制御不能な攪乱による**実験誤差**である[8]．データという言葉は広い意味に使われるが，このように，**人為的でない偶然によるばらつきを含むデータ**だけをこの講義の考察の対象とし，確率論の適用を考えて以下**標本**と呼ぶ．これに対して我々が全貌を見ることのない**神のみの知る全体像**を**母集団**と呼ぶ．つまり，母集団が確率空間の全事象 Ω，実験誤差や無作為抽出で得られたデータを標本 $w \in \Omega$ とみなして確率論を適用する．

母集団は Ω 上の確率 P をこめて指すのが普通である．P が名前の付いている分布の場合，その名前をつけて，たとえば**正規母集団**などと呼ぶ．§5.1 で紹介する正規分布は非常に頻繁に用いられるので，統計学の一般論を展開する場合は正規母集団に基づく場合をまず調べるのが基本であり，この講義でも §7 と §8 で正規母集団の統計的推測の理論（**小標本理論**）を紹介する．個別の例では §1.1 の硬貨投げの 2 項分布のように，種々の分布が理論的理由から用いられる．

統計学の対象になる標本は複数の人に調査をしたり実験を繰り返したり，というように一つの標本の中に「ばらつきを測るために繰り返す」操作が含まれることが多い．例で説明する．国民全員のうちあるテレビ番組を見ている人の割合を，その番組の**視聴率**という[9]．日時とチャンネルと必要なら地域を決めれば決まる一つの数である．しかし，全員を調査するのはその効果に比べて費用がかかりすぎるので厳密な意味での正確な視聴率は現実的には知ることができない．**視聴率調査**とは何人かを抜き出して番組を見た人の割合を算出することである．視聴率調査の結果は番組制作に関わる人たちの人生を左右すると言われるが，調査対象は人口に比べれば無視できるほど少ない．調査に費用がかかっては調査会社が成り立たない．全員を調べていないので調査結果は正しい視聴率からずれる．視聴率調査会社が 2 つ以上あってそれぞれ独立に調査をすれば結果はばらつく．

データは標本 w だと書いたが，複数の人に調査する，という点を明確にするためには確率変数のほうが扱いやすい．ばらつきを引き起こす全ての要因を含む膨大な記録が一つの標本 w だが，たとえば視聴率調査では誰に調査したかという記録は重要ではなく個々人について見た (1) 見ない (0) の情報だけを抽出する．これは Ω 上の関数だから**確率変数**である．複数の人にわたるデータの列をすぐ次に §3.2 で説明する独立な確率変数列 X_k, $k = 1, 2, \cdots, n$, と考える．統計学の基本は人による違いを問題にせず，個々人に p の割合でテレビ番組を見たいという傾向があると考える（つまり一人ずつの調査が母集団からの標本になっている）．よって X_k の分布は母集団の分布で決まると考えるので，k によらない同じ分布になる．まとめて，$\{X_k\}$ は**独立同分布確率変数列**をなすという．

> 統計学の目的は，ある分布の標本，特に独立同分布確率変数列の標本，が与えられたとき，元の分布について主張できることの研究，ということになる．

確率変数は単に関数という意味なので，一つの問題でもどの確率変数に注目して問題を立てるか自由度がある．視聴率の場合は n 人中見た人数 $m = m(w)$ を確率変数と思うこともできるし，一人一人の見た (1) 見ない (0) の列 X_k, $k = 1, 2, \cdots, n$, に注目してもよい．前者では m は 2 項分布 $B_{n,p}$ に従うのに対して，後

[8] 制御できるものは全て制御するのが実験だから，ばらつきが制御不能というのは同義反復だが…

[9] オンエアされていない地域は除く，番組というより各時刻ごとに各チャンネルごとに定義される，世帯単位か個人単位か，広告主の観点からは世代や収入別，など細かいことは無数に考え得るが，ここでは全て省略して単純な 2 項分布の例として考える．

者では X_k たちが独立同分布確率変数列になる．真の視聴率を p とすると，X_k たちの分布 $\mathrm{P}\circ X_k^{-1}$ は，

$$\begin{cases} \mathrm{P}[\, X_k = 1\,] = p \\ \mathrm{P}[\, X_k = 0\,] = 1-p \end{cases} \tag{40}$$

で与えられる．独立同分布確率変数列の長さ n を**データの大きさ**と呼ぶ．二つの見方の関係は $m = X_1 + \cdots + X_n$ で数学的に翻訳可能である．後者の見方から前者の 2 項分布が従うことは，§3.2 で独立という言葉を説明すれば証明できる．2 項分布の節 §1.1 では話の順序を逆にして，次節の結果を先取りして用いて (1) を導いていた．

視聴率調査の続きは §6 にゆだねることにして，ここではじっとがまんして独立確率変数列の説明に戻る．

3.2 独立と相関．

3.2.1 独立な確率変数列．

§2.3 で確率変数 X とその分布 $\mathrm{P}\circ X^{-1}$，期待値 $\mathrm{E}[\, X\,]$，分散 $\mathrm{V}[\, X\,]$ を定義した．

§1.1 や §2.4 や前節 §3.1 で確率変数列の独立という概念を先取りして利用した．確率変数の列 $X_k,\ k = 1,\cdots,n,$ が独立とは，

確率変数列が独立

$$\mathrm{P}[\, X_k \in A_k,\ k = 1,\cdots,n\,] = \prod_{k=1}^{n} \mathrm{P}[\, X_k \in A_k\,] \tag{41}$$

が全ての事象 A_k に対して成り立つことを言う，というのが一般的定義である．X_k たちの取りうる値が離散的な場合（X_k の分布が離散分布の場合）(41) は

$$\mathrm{P}[\, X_k = a_k,\ k = 1,\cdots,n\,] = \prod_{k=1}^{n} \mathrm{P}[\, X_k = a_k\,]$$

があらゆる取りうる値の組 $\{a_k \mid k = 1,\cdots,n\}$ に対して成り立つことと同値になることは，定義から分かる．この場合には近代確率論を習っていなくても (42) を実際に確かめることができる．証明はこまかい技術的なことなので略すが，独立性は次の形でよく用いられる．

命題 4 $X_k,\ k = 1,\cdots,n,$ が独立であることと，任意の非負連続関数達 $f_k :\ \mathbb{R} \to \mathbb{R},\ k = 1,\cdots,n,$ に対して

$$\mathrm{E}[\, \prod_{k=1}^{n} f_k \circ X_k\,] = \prod_{k=1}^{n} \mathrm{E}[\, f_k \circ X_k\,] \tag{42}$$

が成り立つことは同値である． ◇

$\circ$ は $(f\circ g)(w) = f(g(w))$ で定義される関数（確率変数）の合成を表し，関数（確率変数）の積は $(fg)(w) = f(w)\times g(w)$ のように値の積で定義される．特に，

> 独立な確率変数の積の期待値は期待値の積に等しい．

このことから X と Y が独立ならば

$$\mathrm{V}[\, X+Y\,] = \mathrm{V}[\, X\,] + \mathrm{V}[\, Y\,] + 2\,\mathrm{E}[\, X - \mathrm{E}[\, X\,]\,]\,\mathrm{E}[\, Y - \mathrm{E}[\, Y\,]\,] = \mathrm{V}[\, X\,] + \mathrm{V}[\, Y\,]. \tag{43}$$

> **独立な**確率変数の和の分散は分散の和に等しい．

(29) で注意したが，無条件で加法的になる期待値との違いに再度注意してほしい．

3.2.2 確率変数の密度と独立性.

確率変数 X_1 と X_2 の組 (X_1, X_2) の分布が密度を持つ連続分布のとき，X_1 と X_2 の独立性を密度を使って表現する．まず，変数の組の分布というのは 1 変数の場合 (24) と同様で,

$$Q(A) = (\mathrm{P} \circ (X_1, X_2)^{-1})(A) = \mathrm{P}[\ (X_1, X_2)^{-1}(A)\ ] = \mathrm{P}[\ (X_1, X_2) \in A\ ] \tag{44}$$

で定義される $\Omega = \mathbb{R}^2$ 上の確率測度である．違いは当然ながら $A \subset \mathbb{R}^2$，つまり，2 次元集合の上の分布だという点だけである．A は 2 次元事象ということだが，典型的には長方形

$$A = [a, b] \times [c, d] = \{(x, y) \mid a \leqq x \leqq b,\ c \leqq y \leqq d\}$$

は全て 2 次元事象である．このとき (44) から

$$Q([a, b] \times [c, d]) = \mathrm{P}[\ (X_1, X_2) \in A\ ] = \mathrm{P}[\ a \leqq X_1 \leqq b,\ c \leqq X_2 \leqq d\ ].$$

次に，この 2 次元分布が密度を持つ連続分布ということは (18) で触れたとおり,

$$Q(A) = \mathrm{P}[\ (X_1, X_2) \in A\ ] = \int_A \rho(x, y)\, dx\, dy \tag{45}$$

が常に成り立つ区分的に連続な非負値 2 変数関数 ρ があるということである.（本当はもっと緩い条件でいいが，ここでは ρ が滑らかな場合しか扱わないので，この範囲内にとどめておく.）(X_1, X_2) の 2 次元分布とは，X_1 をある学校のクラスの学生の英語の試験成績，X_2 を数学の試験成績，とするとき，英語も数学もできる人，英語はできるけど数学ができない人，数学はできるけど英語ができない人，など，それぞれの割合を与える分布である．これに対して英語の成績は不問にして数学の成績だけで分布を書けば X_2 の分布となる．X_1 の分布も同様である．試験成績は離散分布だが，親玉の 2 次元分布が連続分布ならば X_i それぞれの分布 $Q_i = \mathrm{P} \circ X_i^{-1}$ も連続分布になる．実際たとえば X_1 の場合は (45) から

$$\mathrm{P}[\ X_1 \in A_1\ ] = \mathrm{P}[\ (X_1, X_2) \in A_1 \times \mathbb{R}\ ] = \int_{A_1} \left(\int_{\mathbb{R}} \rho(x, y)\, dy \right) dx$$

なので X_1 の分布は密度関数

$$\rho_1(x) = \int_{-\infty}^{\infty} \rho(x, y)\, dy \tag{46}$$

を持つ連続分布である．同様に X_2 の分布の密度関数は

$$\rho_2(x) = \int_{-\infty}^{\infty} \rho(x, y)\, dx. \tag{47}$$

本題の独立性に戻って，この，連続分布を持つ X_1 と X_2 が独立ということを密度で表す.

命題 5 (X_1, X_2) の分布が密度 ρ の連続分布のとき，X_1 と X_2 が独立であることと

$$\rho(x, y) = \rho_1(x)\, \rho_2(y), \quad x, y \in \mathbb{R}, \tag{48}$$

が同値である．ここで ρ_1 と ρ_2 は *(46)* と *(47)* の X_1, X_2 それぞれの分布の密度関数である． ◇

証明． X_1, X_2 が独立ならば (41) と (45) から

$$\int_{A_1\times A_2}\rho_1(x)\,\rho_2(y)\,dx\,dy=\int_{A_1}\rho_1(x)\,dx\times\int_{A_2}\rho_2(y)\,dy=\mathrm{P}[\,X_1\in A_1\,]\,\mathrm{P}[\,X_2\in A_2\,]$$
$$=\mathrm{P}[\,X_1\in A_1,\ X_2\in A_2\,]=\mathrm{P}[\,(X_1,X_2)\in A_1\times A_2\,]=\int_{A_1\times A_2}\rho(x,y)\,dx\,dy.$$

A_1, A_2 を小さな区間にとって極限を考えるいつもの方法から (48) を得る．

次に (48) が成り立っているとすると，今の計算を中ほどで分割することで，(41)（$n=2$ の場合）が成り立つから X_1, X_2 は独立になる． □

3.2.3 確率変数の相関．

独立確率変数は分散の加法性 (43) を始め数多くの扱いやすい性質があり，無作為抽出や追試等で得られた複数のデータを独立確率変数とみなすのが統計学の出発点である．逆に変数の間に独立でない兆候が見られれば未知の自然法則や社会法則や場合によってはデータ採取上の不備の可能性がある．2 つの確率変数がどれくらい独立か独立でないかを表す目安として相関は統計学でよく用いる．

$C(X,Y)=\mathrm{E}[\,(X-\mathrm{E}[\,X\,])(Y-\mathrm{E}[\,Y\,])\,]$ を確率変数 X, Y の共分散と言う．特に $\mathrm{V}[\,X\,]\mathrm{V}[\,Y\,]\neq 0$ のとき，$r(X,Y)=\dfrac{C(X,Y)}{\sqrt{\mathrm{V}[\,X\,]\,\mathrm{V}[\,Y\,]}}=\dfrac{\mathrm{E}[\,(X-\mathrm{E}[\,X\,])(Y-\mathrm{E}[\,Y\,])\,]}{\sqrt{\mathrm{V}[\,X\,]\,\mathrm{V}[\,Y\,]}}$ を X, Y の相関係数という．

命題 6 $\mathrm{V}[\,X\,]\mathrm{V}[\,Y\,]\neq 0$ のとき $-1\leqq r(X,Y)\leqq 1$．特に X, Y が独立ならば $C(X,Y)=0$. ◇

証明． 最初の主張は $f(t)=\mathrm{E}[\,(t(X-\mathrm{E}[\,X\,])-(Y-\mathrm{E}[\,Y\,]))^2\,]\geqq 0$ の判別式から，後の主張は命題 4 から． □

独立ならば相関係数 $r(X,Y)$ は 0（逆は言えないので応用上は注意）．一方 $r(X,Y)=\pm 1$ を変形すると $\mathrm{E}[\,(X\mp\frac{\mathrm{E}[\,XY\,]}{\mathrm{V}[\,Y\,]}\,Y)^2\,]=0$ を得るが，期待値は和または積分だから非負関数の期待値が 0 になるのは関数そのものが事実上恒等的に 0 の場合だけである[10]．よって，$X=\pm\dfrac{\mathrm{E}[\,XY\,]}{\mathrm{V}[\,Y\,]}\,Y$ となる．言い換えると相関係数が ± 1 になるのは,（事実上）Y と X が比例関係にある場合に限り，正負は X と Y が同符号か逆符号かを表す．比例関係というのは x キロメートルが $1000x$ メートルに等しい，というように一つの量を別の単位で見ているという意味である．こうして相関係数は独立な確率変数の組と本質的に一つの確率変数のどちらに近いかの目安になる．

3.3 乱数．

乱数は無作為抽出の典型例でもあり無作為抽出を実行するための補助手段でもある．乱数とは並び方に規則性のない数列というのが本来の「定義」だが，ことは簡単ではなく，乱数とは何かについての決着はついていないようにみえる．それどころか，本来の意図とは逆に，疑似乱数と呼ばれる規則的に発生させた数列が計算機の発達によって実用上は主流である．原理的問題はともかく，統計学の応用上は重要なので，ここではもっとも簡単でもっとも広く流布している**合同法による一様分布する疑似乱数**（一様乱数）を紹介する．

[10]「事実上」というのは測度論の言葉で言えば「確率 0 を除いて」というのが正しい言い方で，被積分関数が 0 でない点を全て集めた集合（事象）を N と置くとき $\mathrm{P}[\,N\,]=0$ となることをいう．つまり確率的に無視できる事象を除けば恒等的に 0 と思ってよい．特に X, Y が区分的に連続な関数ならば恒等的に 0 ということを意味するので，以下そのような意味と考えて読み進んで差し支えない．

昔，といっても，コンピュータ時代以前というだけだからそう遠い昔では無さそうだが，は正 20 面体さいころを本当にふって乱数表を作っていたそうである．現代ではプログラムを組んで計算機で計算させる．これは決まったアルゴリズムだから，初期値を与えるとあとは**規則的**に数列が作られる．不規則，という乱数の根本的定義と矛盾しているので，規則的に作られた（従って容易に再現可能な）一見不規則に見える数列は乱数と区別して**疑似乱数**と呼ばれる．もっとも実用上は今日では乱数とは疑似乱数のことである．一見原理的矛盾に見える何度でも再現できる数列ということを積極的に利用して，プログラムの誤りの検出やプログラムのバージョンアップの効果の測定に用いるのがこんにちの疑似乱数に対する肯定的姿勢の理由の一端である．

擬似乱数は，建前としては初期値の集合を全事象 Ω とし，各初期値を等確率でとる確率 P が定義された確率空間の上の確率変数列 $\{X_n\}$ が n の漸化式で与えられるものである．もっとも簡単な漸化式として $X_{n+1} = f(X_n)$ という形のみとりあげる．ω が初期値だから $X_0(\omega) = \omega$．このとき X_n と X_{n+1} は（どちらか少なくとも一方が常に決まった数，というナンセンスな場合を除いて）必ず**独立でない**ことが証明できる．無作為抽出，すなわち独立同分布確率変数列の標本を得ることが乱数の目的だとすると原理的な矛盾である．しかし，アルゴリズムが簡単でかつ高速なので，原理的問題にもかかわらず普通は使われる．

一様分布

区間 $[0,1)$ 上の**一様分布**とは密度関数が $[0,1)$ 上恒等的に 1 の連続分布のことである[11]．言い換えると $\mathrm{P}[\,[a,b]\,] = b - a$ が任意の $0 \leqq a \leqq b < 1$ に対して成り立つ分布である．

一様分布を元の分布とする疑似乱数（一様乱数）の生成アルゴリズムで合同法と呼ばれるものは

$$U_{n+1} = aU_n \mathrm{mod M},\quad \mathrm{n} = 0, 1, 2, \cdots,$$

で定義される．標準的には $M = 2^{32}$ と $a = 48828125$ が採用され，初期値は $U_0(w) = w = 1000001$ などを選ぶ．一般には $M = 2^b$ と $a = 4 \pm 1 \mathrm{mod} 8$ で初期値を奇数にとったときそのときに限り最大周期 $M/4$ が実現することが知られている．[Knuth, pp.19–21].（全部で M 個の数字しかとれないので，必ず周期的に繰り返すことに注意．周期が長いほど長く乱数として使えるので望ましい.）整数で用いるときはこのままでいいが，U_n/M をとれば $[0,1)$ 上の一様乱数となる．

自己相関

本来の意味での独立性はないが，独立性らしさとして**自己相関係数**

$$r(X_n, X_{n+1}) = \frac{\mathrm{E}[\,(X_n - \mathrm{E}[\,X_n\,])(X_{n+1} - \mathrm{E}[\,X_{n+1}\,])\,]}{\sqrt{\mathrm{V}[\,X_n\,]\mathrm{V}[\,X_{n+1}\,]}}$$

が n が大きいときに 0 に近いことなどを検査するのが普通である（相関係数については §3.2.3 参照）．より精密に議論するときは (X_n, X_{n+1}) を 2 次元に図示したときに見える格子状の構造（相関の存在）が均一になる（小さな正方形に近い）ことを検査する．

経験分布

独立同分布確率変数列 $\{X_k \mid k = 1, \cdots, n\}$ の標本（つまり数列）$a_k = X_k(w)$, $k = 1, \cdots, n$, を集めた集合を母集団 $\Omega_n = \{a_1, \cdots, a_n\}$ とし全ての点が等確率 $\mathrm{P}[\,\{a_k\}\,] = \dfrac{1}{n}$ とした離散分布を元の分布との対比で**経験分布**と呼び，確率論では $\dfrac{1}{n}\displaystyle\sum_{k=1}^{n} \delta_{X_k(w)}$ と書くことがある．

自己相関の他に**疑似乱数に通常求められる条件**は，高速に生成できること，および経験分布が n が大きいとき元の母集団の分布に近いことである．前者は大規模な調査や実験が念頭になるのだから意味のある結果を得るとすれば当然，後者は経験分布に関する大数の強法則から，乱数が元の分布のシミュレーションであるために必要．特に標本の平均と分散はそれぞれ母集団の平均と分散に，データの大きさが大きいとき近くあるべきである．

一様乱数さえあれば，他の分布に従う疑似乱数もそれに基づいて生成できる．関心の中心は個別の分布に対して生成速度が速くなる工夫である．必要になったときに [Devroye] などを参照して頂きたい．

再考：標本と確率変数．

§3.1 では，統計学的アプローチでは，現実に起きたことを見えない母集団の標本と考え，そのなかの注目している数値データを確率変数としてとらえると書いた．

[11] 厳格主義的理由から 0 は含めて 1 を除いているが，連続分布では 1 点の確率は 0 なので閉区間 $[0,1]$ としても本質的な差はない．

極端な立場をとると，我々が認識するの宇宙はただ一つだから，Ω とは「あり得たかもしれないが実際には存在しなかったあらゆる宇宙」を集めたものであり，我々の現実はその中のただ一つの標本 $w \in \Omega$ である．1 等を夢見て宝くじを買うが実際ははずれる，1 等からはずれまでの全ての可能性を集めたのが Ω であり，はずれをのぞく大部分の標本 $w \in \Omega$ は現実には起こらない．Ω 全体は神のみぞ知る，人は知ることがない．宇宙はただ一つ現にあったものだけ，という立場からは，異なる w とは仮想的な「パラレルワールド」があって，分布から取り出された標本だけが異なり，他の条件が全く同じ世界，ということになる．たとえば，一等の宝くじが当たった自分がどこかにあって，その世界と現実の世界で平均を取るのが w について平均をとることになる．

そこまで誇大妄想にならないで，日本人全体を母集団 Ω とみて，その中から無作為に選んだ人たちにアンケートや調査を行う場合でも，Ω 全体を知ることはまれである．むしろ全体を厳密に知ることの費用に比べて効果が著しく低い場合に統計学的アプローチが意味を持つ．もちろん，自然科学の実験はどれほど繰り返しても全ての制御できない攪乱の結果を集めることはできない．この意味で統計学で考える母集団に関する確率論的計算は我々が直接見ることのできない神の世界の確率論ということになる．

ただし，慎重に計画された調査，観測，実験では，**追試**が可能であるという立場に立つ．つまり，執り行う人や日時が異なるけれども，同じ母集団に対する異なるデータ w とみなせると考える．実験における追試や視聴率で言えば他社の調査結果などがこれに属する．これが統計学の立場からみた**実験の再現可能性**である．実験の再現可能性は自然科学が世の中の動きを表す自然法則を発見するにあたって，他の知的活動と根本的に区別される．宇宙を唯一の歴史と考えるのではなく，その中の限られた一部分は繰り返すことができる，と考えるからこそ自然法則という考え方が成り立つ．それが自然に対するのによい見方かどうかはそうやって得られた自然法則を元に予測を行うことで，新しいものを生み出したり新しいことを実行してそれが期待通りの成果を生むかどうかで判断する．月にロケットをとばして帰って来れたり，電磁波を使ってテレビや電子レンジが使えるなど，こんにち自然科学的方法を疑うことは不可能である．この実験等の繰り返しにおいて誤差だけは再現不可能である．再現不可能な部分を誤差と呼んでいるとも言える．この誤差部分を異なる w と見立てることができる．この場合は分布にも若干現実性がある．それでも全ての $w \in \Omega$ を尽くすことはない．

繰り返すと,「パラレルワールド」を見ることができる「神」にとってはデータは確率変数である．大きさ n の標本 X_k, $k = 1, \cdots, n$, は独立確率変数であるとして統計学の理論は展開する．たとえば，視聴率調査で無作為に n 人を選んで順に番組を見たか見ないかを $1, 0$ で記録したものを $\{X_k\}$ としたとき，それが独立同分布確率変数で分布が (40) で与えられる，というのは神の立場である．その建前は，人は皆確率 p くらいでその番組を見たいと思っていて，心の中でそれぞれが独立に不公平な硬貨を投げて表裏に合わせて実際に見るか見ないかを決めている，という考えである．独立としたときに見た人数 m が $B_{n,p}$ に従う，というのが (1) で 2 項分布を導いた原理に他ならない．これに対して人が見ることのできるのは現実という唯一の w であり，そこでは標本は $X_k(w)$, $k = 1, 2, \cdots, n$, という，確定した数値の列である．$\{X_k\}$ が独立確率変数になるように正しく標本を選んでいるかどうかは神にしか証明できない．人の立場から近づきたいという思いが「無作為抽出」や「人為による偏りのない」実験という言葉の意味である．数理統計学は無作為抽出の方法は与えない．無作為抽出できたとして話を進める．

ここで書いたことはこのように考えねばならない，という意味ではない．ただ，統計学の応用が有効な場面では多くの場合，このように考えるのが確率論を応用する上で分かりやすいというのが私の経験である．

再考：乱数．

何をしたいか？ 統計学的調査では「無作為抽出」が必要である．標本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，が定義だが，ある選択がそうなっているかどうか証明できるのは神だけである（§3.1 および前節の補足）．原理的な問題を横に置いても，たとえば調査会社やアルバイトの人に調査を委託するときに選択手順がなければ伝えようがない．そこで規則性のない数の列を記した表を用意して，総リストからその数の番号の対象だけを選ぶ，ということが行われてきた．これを**乱数表**と呼ぶ．統計学の基礎教科書は巻末に乱数表が載っているのが普通であった．

乱数には不規則な現象の**シミュレーション**というもう一つの用途がある．統計学との関連に限って言うと，高額の大規模調査や大規模実験を行う前に，小規模な萌芽的実験や机上の理論的研究によって成果の予測を慎重に検討して計画を立てるのは当然である．計算機の発達で，計算機で模擬実験を行うシミュレーションが強力な手法になっている．

たとえば，何らかの理論によって母集団の分布を推測したとき，そこからデータを取出すとどんな結果が期待されるかという標本を得たい．それをたとえば本番のデータと比べること（§5 統計学的検定）で理論の正しさを統計学的に「証明」したい．その際,「無作為抽出」や「制御不能な攪乱としての誤差」のような人為の入らない数値が必要である．独立同分布確率変数列の標本 $\{X_k(\omega)\}$ がほしいことになるので，分布は与えられているが，並び方は不規則な数列，が実用上必要な乱数である．

何が難しいか？ 統計学的世界観の特徴として一つの標本に意味はない．一方乱数列を一つ与えたらそれは一つの標本を特別扱いしたことになる．他方で，どの標本も平等といいながら人は明らかに差別している[12]：

(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 $X_1(\omega) = 3, \cdots,$ なる標本 $\omega \in \Omega$ が実現したということである．独立同分布で 1 回ごとに公平なさいころだとするとこれが起こる確率はわずか $2^{-10} \approx 0.1\%$ だが，いちおう元の分布に従った乱数のように見える．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，乱数には見えない．なぜか？

[12]なにやら現実の社会の人の行動とよく似ているが，統計学は人の意志決定を支援する学問なので人の可能性も限界も直接反映する．もっとも，学問はすべからく人のやることなので，その可能性も限界も人のそれで決まっていると言われればそれまでだ．

(iii) どれほど長い列にしても 1 が 10 回並ぶことがないように工夫したら,「そんな規則性を入れたら乱数でない」と怒られた．上記の列はたまたま最初の 10 回に 1 が並んだだけなのに採用されなかった．どうすればいいのか？

以上の例は原理的問題もさることながら，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，と筆者は考える．つまり，乱数の定義から「不規則」や「無作為」という条件ははずすべきときが来ていると思う．

補足：意志決定支援手段としての統計学．

現実と確率論をつなぐものとして統計学を考えたとき，橋渡しは入口と出口の 2 箇所ある：

(i) 標本をどのように見れば数学的対象となりうるか（数学的議論が容易になるか）？
(ii) 数学的結論をどのような形で表現すれば使いやすいか（意志決定の支援に役立つか）？

この節 §3 では前者について議論してきたが，値がばらつく現象における標本の偶然性は調査や実験という入口の問題だけではない．たとえ母集団の分布が完全に分かっても，分布すなわちあることが起きる割合が分かっただけなので，その時間的順序，つまり次に何が起きるか正確には予言できない．何が起こるか分かったら決定論的であって統計的現象の範疇ではない．もちろん月の軌道が正確に予言できるからロケットを打ち上げて月に着陸できるのであって，行ってみたら月の軌道が偶然大きくそれていてロケットは大宇宙の暗黒の中に取り残されていた，というのでは話にならない．ニュートンの運動方程式と重力いう決定論的な法則がなければ月めがけてロケットを打ち上げる計画はあり得なかった．他方，さいころの次の目が分かったら双六はゲームにならない．つまり，予言できないこと，ばらつきの偶然性が本質である現象もたしかにある．統計学が扱うのは後者である．

統計的現象は確定的な予言が原理的に存在しないので，ある予言をしたとき当たったかどうか自体を自分で必要に応じて定義しないといけない．それを決めるのは個々人の個々の場面での決断（意志決定）である．その意味で統計学は事実の予測をする学問ではなく意志決定を支援する学問である．データに基づいて決断（たとえば会社の活動方針や国の政策決定）がなされるとき，その「情報処理」の過程を数学的な議論に基づく部分と個人の主観（意志）に任されている部分に分ける技術が統計学だと言ってもいいように思う．結果として，決断する立場（経営者や政治家）からは主観の揺らぎにごまかされずに真に有利な選択肢を選べるし，外部から監視する立場（株主，公正取引委員会，オンブズマン，有権者）からは表面上の説明の裏にあるかもしれない嘘を見抜くことができる．大学や研究所などで実験や観測のデータから法則を発見・検証する場面では，誤差（ノイズ）の大きいデータから真実を見抜く手がかりを与え，他方で希望的観測に惑わされることなく冷静に必要な実験を重ねる決心を与える．

確率論に基づく数学的結論をどのような形で表現すれば使いやすいか（意志決定の支援に役立つか）について，正規母集団に関する推測という基礎的具体的な例について後半の講義で詳しく説明する．

4　宮城沖地震を逃れる確率 — 点推定．

手元にあるデータ（標本）から神のみの知る元の分布（母集団）について何かの結論を統計学的に得ることを統計的推測という．そのもっとも単純な形式として標本の関数（データで決まる確率変数）で母集団の性質の推定量とする点推定がある．そのもっとも基礎的でよく用いられる標本平均と不偏分散を説明する．

それらが推定量として適当であることの根拠との関連で，大数の強法則を紹介する．後での都合上，大数の強法則と並んで基礎的な極限定理である中心極限定理も紹介する．

4.1　標本平均と不偏分散．

§3 に説明したとおり，この講義における統計学の目的は，ある分布の標本または各確率変数がその分布に従う独立同分布確率変数列の標本が与えられたとき元の分布について主張できることを調べることである．

データの大きさ（独立同分布確率変数列の長さ）が大きければ経験分布 $\frac{1}{n}\sum_{k=1}^{n}\delta_{X_k(w)}$ （§3.3 参照）が事実上全ての w に対して元の分布に近づくことが知られている．つまりただ一つの現実（標本）しか知らなくても元の分布を推測できる．しかし，一般には地震の発生間隔のように小さなデータしか得られない場合や視聴率調査のように経済的理由から小さなデータで大きな母集団を推測したい場合まで，n は経験分布が元の分布に近いと言えないほど小さなところで結論を出したい．

たとえば，2004 年 6 月現在の地震調査研究推進本部の web page「宮城県沖地震の長期評価」

`http://www.jishin.go.jp/main/chousa/00nov4/miyagi.htm`

の表 2 によれば宮城県沖地震の過去の発生記録は次のようになっている．

発生年/月/日	間隔（年）
1793/02/17	
1835/07/20	42.4
1861/10/21	26.3
1897/02/20	35.3
1936/11/03	39.7
1978/06/12	41.6

発生間隔は一定ではなく不規則にばらついているようにみえる．この 5 つの数字を大きさ 5 のデータ，すなわち背後にある真の分布に従う長さ 5 の独立確率変数列の標本と見て背後の分布を見いだしたい．

一見全く異なる経験分布から元の分布について何が言えそうか？母集団の分布に従う独立同分布確率変数列 X_k, $k=1,\cdots,n$, の標本 $\{X_k(w) \mid k=1,\cdots,n\}$ が与えられたとする．同分布確率変数列としたので従う分布は k によらないからその分布に従う確率変数を X とおく（単に以下で毎回添え字を付けたり k によらないという注釈を付けたりしないですますため）．

小さなデータで母集団を完全に決めるのは不可能である．母集団の分布から計算される代表的性質，目安となる量，最もほしい性質，であって比較的小さいデータでも近い値になる（収束が早い）ものを**母集団特性値**と呼び，それを推測することを考える．母集団の平均 $\mu = \mathrm{E}[\,X\,]$ や分散 $v = \mathrm{V}[\,X\,]$，場合によっては高次のモーメント $\mathrm{E}[\,X^n\,]$ 等が母集団特性値である．母集団特性値の推定量としての確率変数（**統計量**とも呼ぶ）の標本 w での値で特性値を推定することを**点推定**と呼ぶ（対義語として §6 区間推定を参照）．必ずしも一つに決まらないが，通常

母平均 μ の推定量として**標本平均**

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \tag{49}$$

が用いられ，母分散 v の推定量として**不偏分散**

$$V_n = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \overline{X}_n)^2 \tag{50}$$

が用いられる．

別の言い方として，母集団の分布の形を平均と分散など小数のパラメータを除いて何らかの理論的・経験的考察に基づいて決めておき，パラメータだけをデータから推定した，と言うこともできる．講義の後半ではこの立場で話を進める．

(49) や (50) が推定量として適切である主な根拠，**一致性，不偏性，有効性**をまとめておく（[林周二, 14.1], [辞典, 287C]）．

一致性． データ数 n を大きくした極限で母集団の統計量に収束する性質．標本平均，不偏分散とも一致性を持つ：

$$\lim_{n\to\infty} \overline{X}_n = \mathrm{E}[\,X\,],\ a.e., \tag{51}$$

$$\lim_{n\to\infty} \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2 = \mathrm{V}[\,X\,],\ a.e. \tag{52}$$

右辺の X は上で導入したのと同じ意味で，X_k たちと同じ分布を持つ確率変数．a.e. と書いたのは，両辺とも $w \in \Omega$ の関数だが，左辺が収束して右辺に等しい w の集合の確率が 1，つまり等号の成り立たない w の集合の確率が 0，という意味である．「事実上全ての w」ということになる．我々の世界が一つの標本 w だという立場でも，「神の目から見て確率 0 という異常事態」がない限り，$\lim_{n\to\infty} \overline{X}_n(w) = \mathrm{E}[\,X\,]$，つまり，十分データを重ねれば標本平均はいつかは母平均に近づくことを意味する．Ω 上の関数列の収束としては各点収束（無視できる点を除いて）を意味する．この収束を確率変数列の概収束と呼ぶ．

(51) は大数の強法則と呼ばれる有名な定理で，期待値の存在する独立同分布確率変数列ならば必ず成り立つことが知られている．(52) は，左辺の lim の中の式を変形すると，任意の定数 μ に対して

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

$\mu=\mathrm{E}[\,X\,]$ ととれば第 1 項は大数の強法則 (51) において X_j を $(X_j-\mathrm{E}[\,X\,])^2$ に置き換えれば $\mathrm{E}[\,(X-\mathrm{E}[\,X\,])^2\,]=\mathrm{V}[\,X\,]<\infty$ のとき $\mathrm{V}[\,X\,]$ に概収束することが分かる．第 2 項は大数の強法則から 2 乗の中身が 0 に概収束する．よって (52) が $\mathrm{V}[\,X\,]<\infty$ のとき成立する[13]．

不偏性． 一致性だけでは，たとえば不偏分散の定義 (50) の分母は $n-1$ でなくても n でもかまわない（$\lim_{n\to\infty}\dfrac{n}{n-1}=1$ だから）．わざわざ $n-1$ にする理由は，そのとき

$$\mathrm{E}[\,\overline{X}_n\,]=\mathrm{E}[\,X\,],\quad \mathrm{E}[\,V_n\,]=\mathrm{V}[\,X\,], \tag{53}$$

が成り立つからである．これらは定義から直接確かめられるので，演習問題としておく．一致性は概収束なので，現実の世界でデータをとり続ければ母集団特性量に収束することが事実上約束されている．これに対して不偏性の左辺の期待値は神のみの知る「パラレルワールド」に関する平均，あるいはもう少し現実主義的には追試などにわたる平均（の極限），だから現実のデータではその価値はあまりなさそうである．準理論上の扱いやすい性質，というほうが妥当であろう．不偏分散という用語は当然不偏性から来ている．

有効性． 以上の一致性と不偏性だけで決まらない場合は候補となる推定量（確率変数）Y のうち分散 $\mathrm{V}[\,Y\,]$ が最小の量が望ましい．ちらばりが小さければ n を大きくしたとき早く母集団特性量に収束するから．これを有効性という．適当な条件があると分散が最小の不偏推定量が一つ決まる．

4.2 概収束と大数の強法則．

(51) の説明で用意した確率論用語，概収束と大数の強法則を復習しておく．

確率空間 $(\Omega,\mathcal{F},\mathrm{P})$ 上の確率変数 $X_k:\ \Omega\to\mathbb{R}$ の列 $(k=1,2,\cdots,)$ と確率変数 $X:\ \Omega\to\mathbb{R}$ に対して，確率変数列 X_k, $k=1,2,\cdots,$ が確率変数 X に概収束するとはある確率 0 の事象 $N\subset\Omega(\mathrm{P}[\,N\,]=0)$ がとれて，$\Omega\setminus N$ 上での各点収束

$$\lim_{n\to\infty}X_n(w)=X(w),\quad w\in\Omega\setminus N,$$

が成り立つことを言い，$\lim_{n\to\infty}X_n=X$, a.e., と書く．

定理 7 (大数の強法則) X_k, $k=1,2,\cdots,$ が独立同分布確率変数列とし，その分布は $\mathrm{E}[\,|X_1|\,]<\infty$ を満たすとする．このとき

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}X_k=\mathrm{E}[\,X_1\,],\ a.e.,$$

すなわち単純平均は分布の平均（定数関数）に概収束する． ◇

この定理の証明には確率論の基礎知識・基礎技術が必要なので割愛する．確率論の基礎教科書を参照して頂きたい．

[13]記録によると 1996 年 5 月 27 日に楠岡成雄先生に教わった．なぜこれだけが記録されていたか記憶にない．

4.3 正規分布と分布の収束と中心極限定理.

この機会にもう一つ確率論から用語と定理を紹介しておく.

§1.3 で 2 項分布 $B_{n,p}$ が適当なパラメータの調節の下で $n \to \infty$ で正規分布に弱収束することを紹介し, 中心極限定理と呼んだ. その時点では定理をきちんと書くための用語が足りなかったので曖昧に書いたが, §6 での都合もあるのでここでもう少し正確に書いておく.

平均 μ 分散 v の正規分布とは

$$\frac{1}{\sqrt{2\pi v}} e^{-(x-m)^2/(2v)} \tag{54}$$

を密度関数とする 1 次元連続分布であり, $N(m,v)$ と略記することが多い. 特に $m=0$, $v=1$ のとき, つまり $N(0,1)$, を標準正規分布と呼ぶ.(その密度関数は (13) で先取りしていた.)

(54) の定数は全事象の確率が 1 になるように決まっている. これは, ガウス積分

$$\int_{-\infty}^{\infty} e^{-x^2/2}\, dx = \sqrt{2\pi} \tag{55}$$

に変数変換 $x = \dfrac{z-m}{\sqrt{v}}$ 容易に分かる.

1 次元分布(離散分布でも連続分布でもかまわない)の列 P_k, $k=1,2,\cdots$, が 1 次元連続分布 P に弱収束するとは任意の区間 $[a,b]$ に対して

$$\lim_{k\to\infty} \mathrm{P}_k[\,[a,b]\,] = \mathrm{P}[\,[a,b]\,]$$

が成り立つことを言う.
2 次元以上の分布でも同様なので繰り返さない.(離散分布への弱収束も統一的に定義できるし, そのほうが普通の書き方だが, 煩雑になるので省略する.)

概収束と違って, 分布の収束である.

定理 8 (中心極限定理) X_k, $k=1,2,\cdots$, が独立同分布確率変数列とし, その分布は $0 < \mathrm{V}[\,X_1\,] < \infty$ を満たすとする. このとき $\dfrac{1}{\sqrt{n\,\mathrm{V}[\,X_1\,]}} \displaystyle\sum_{k=1}^{n} (X_k - \mathrm{E}[\,X_1\,])$ の分布は $n \to \infty$ で標準正規分布 $N(0,1)$ に弱収束する. ◇

この定理の証明には確率論の基礎知識・基礎技術が必要なので割愛する. 確率論の基礎教科書を参照して頂きたい.

2 項分布で先取りした中心極限定理(定理 3)は定理 8 で $\{X_k\}$ が (40) なる分布を持つ独立同分布確率変数列の場合に相当することは計算してみると分かる.

大数の強法則が概収束, すなわち我々の唯一の現実(標本)でもほぼ間違いなく観測できるはずの収束なのに対して, 分布は神のみぞ知る, という立場からは分布の収束の現実的価値は分かりにくいように見えるが, 古典的な統計学(大標本理論, 小標本理論)の根拠となる. 大数の強法則から, $\overline{X}$ は $n \to \infty$ で母集団の平均に概収束するのでたくさん標本を集めれば母集団の平均値の推測の精度が上がる, という期待が持てる. 実は標本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. 基本的にはこれが点推定の本来の根拠であろう. 現実にはそこまでデータを集める経済的価値がない場合が殆どである. n が大きいけれども ∞ とは言えない状況(大標本理論)は §6, 正規母集団について小さい n で行う統計的推測(小標本理論)は §7, §8 を参照.

4.4 例：宮城沖地震の発生間隔分布．

前節 §4.1 の点推定量，(49) の標本平均 $\overline{X}_n$ と (50) の不偏分散 V，の具体例として前節冒頭に挙げた宮城沖地震発生間隔一覧に適用すると，

$$\text{大きさ } n=5,\ \ \overline{X}_5 = 37.06,\ \ V = 43.74\ \ (\sqrt{V} = 6.61).$$

周期が約 37 年でばらつきの目安（標準偏差）が 7 年弱ということになる．最後の地震発生が 1978 年だったから 2008 年頃から 2022 年頃の間ならいつおきても当然ということになる．

もう少し確率らしい予想をするために，§11 で地震発生間隔分布の簡単なモデルとして (80) で定義される連続分布（平均と分散をパラメータとする分布の族）を提案する．そのパラメータを上記で点推定すると母分布を推定したことになる．詳しくは §11 に譲るが，この母分布を元に 2004/06/17 現在地震が起きていないという条件付き確率を計算できる[14]．結果は 2004 年 6 月 17 日現在で次表のようになった．

期限	前回からの経過時間 t	期限までの発生確率 $q(t)$
2007/03/31	28.8	6.5%
2009/03/31	30.8	15%
2012/03/31	33.8	32%

これから各自が宮城沖地震を逃れる確率を読み取ることは各自の自習に任せよう．

発生時期予想以上に重要なのは被害予測だが，

`http://www.city.sendai.jp/syoubou/bousai/sairai/`

によれば，1978 年の地震では，当時の仙台市域で死者 13 人，そのうち 9 人はブロック塀の倒壊によるものであった．現在市はブロック塀の生け垣などへの転換を補助金等によって推奨しており，この点では前回と同条件ならば死の覚悟を要する地震ではないと考えられる．

もっとも，1978 年地震のときは火災発生がわずか 8 件で，これは，夏場の食事準備と無関係な時刻で，しかも直前に大きめの地震があって多くの人がそこでガスを止めていた，という幸運が重なったためと考えられているので，次回の地震では火災のために死者がかなり出る可能性を考慮しないといけない．

また，前回の地震でも死者こそ極めて限られていたものの，けが人は 1 万人弱，住家の全半壊が 4000 戸強なので，小さな地震ではない．電気と電話は 1,2 日ですぐにほぼ全面復旧したが，ほぼ全面復旧まで約 1 週間かかった水道には備えが必要である．ガスは 1ヶ月近くかかるので全くあてにならない．

補足：LLN, CLT，および特性関数．

大数の法則の証明の概略（4 次モーメント有限の場合）．

定理 7 の大数の強法則 (**L**aw of **L**arge **N**umbers) を証明するためにはマルチンゲール的な事象の場合分けの議論が必要だが，仮定を強くして,「性質の良い」分布を持つ確率変数列に限ればより初等的に証明できる．詳しいことは確率論の教科書に任せて，証明のあらすじを紹介する．

定理 9 (大数の強法則（4 次モーメント有限な場合）) 独立同分布確率変数列 $X_n,\ n\in\mathbb{N},$ が $\mathrm{E}[\,X_1\,]=0,\ \mathrm{E}[\,X_1^4\,]=\delta^2<\infty$ を満たせば $\dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ は 0 に概収束する． ◇

注 10 分布の平均 μ が 0 でない場合は $Y_n = X_n - \mu$ に定理 *9* を適用すればよい． ◇

証明． $\mathrm{E}[\,X_1^2\,]=\sigma^2$，および，$W_n=\displaystyle\sum_{k=1}^{n} X_k$ とおく．高次のモーメントが有限ならば非負低次のモーメントも有限になることは承知としよう（たとえば 0 次のモーメントは $\mathrm{E}[\,|X|^0\,]=\mathrm{E}[\,1\,]=1$ だからいつでも有限）．このことから

[14] もちろん，この文章を書いていたのが 2004 年 6 月だから．書いている時点では内容を更新するわけにいかないが，これを読んでいる時点でまだ地震が起きていなければ §11 に説明してある手順で内容を更新できる．

W_n^4 を X_k たちについて展開したとき $\mathrm{E}[\ X_i^3 X_j\ ]$ の形の項は独立性から $\mathrm{E}[\ X_i^3\ ]\mathrm{E}[\ X_j\ ]$ と，期待値の積に分けられて，同分布性から $\mathrm{E}[\ X_j\ ] = \mathrm{E}[\ X_1\ ] = 0$ となる．よって，

$$\mathrm{E}[\ \left(\frac{1}{n}W_n\right)^4\ ] = \frac{1}{n^4}\left(\sum_{k=1}^{n}\mathrm{E}[\ X_k^4\ ] + 3\sum_{k=1}^{n}\sum_{j\neq k, 1\leqq j\leqq n}\mathrm{E}[\ X_k^2\ ]\mathrm{E}[\ X_j^2\ ]\right) = \frac{\delta^2}{n^3} + 3(n-1)\frac{\sigma^4}{n^3}$$

となるので，

$$\mathrm{E}[\ \sum_{n=1}^{\infty}\left(\frac{1}{n}W_n\right)^4\ ] = \sum_{n=1}^{\infty}\mathrm{E}[\ \left(\frac{1}{n}W_n\right)^4\ ] < \infty.$$

（確率変数の級数の期待値で，確率変数が非負関数なら期待値と級数の順序を交換できるのは単調収束定理と呼ばれる性質.）期待値が有限値ということは確率変数が無限大になる点の集合が確率 0 でないといけない.「確率 0 を除いて成立」というのを a.e. と書くから

$$\sum_{n=1}^{\infty}\left(\frac{1}{n}W_n\right)^4 < \infty,\ a.e.,$$

となるがこれは，$\lim_{n\to\infty}\left(\frac{1}{n}W_n\right)^4 = 0,\ a.e.$, 即ち，$\lim_{n\to\infty}\frac{1}{n}W_n = 0,\ a.e.$, を意味する． □

特性関数と中心極限定理の証明の概略．

中心極限定理 (**C**entral **L**imit **T**heorem) 定理 8 の証明には特性関数が便利である．一般に分布の弱収束を証明するのに都合がよいが，正規分布には特性関数がとりわけ有効である．

確率変数 X の特性関数 ϕ_X は

$$\phi_X(\xi) = \mathrm{E}[\ e^{\sqrt{-1}\xi X}\ ],\quad \xi\in\mathbb{R},$$

で定義される．この定義は X の分布が連続分布か離散分布かによらない一般的な定義である．

ρ を密度関数とする連続分布の特性関数 ϕ_ρ は

$$\phi_\rho(\xi) = \int_{-\infty}^{\infty}\rho(x)\,e^{\sqrt{-1}\xi x}\,dx,\quad \xi\in\mathbb{R},$$

で定義される．連続分布かどうかに無関係な一般的な定義が可能だが省略する．どんな実数値確率変数 X でもどんな分布でも特性関数は全ての ξ で存在することもルベーグ積分論を知っていれば容易に分かる．分布を与えると特性関数が決まるが，逆に特性関数を与えると分布が決まる．密度関数を持つ場合だけ書いておこう．

定理 11 (P. Lévy の反転公式) ϕ_ρ を密度関数 ρ を持つ分布の特性関数とすると，$a < b$ に対して，

$$\rho(x) = \lim_{\epsilon\to 0}\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{(e^{-\sqrt{-1}\epsilon t}-1)}{-\sqrt{-1}\epsilon t}e^{-\sqrt{-1}at}\phi_\rho(t)\,dt.$$

◇

単に特性関数と分布が 1 : 1 に対応するだけではなく，分布の弱収束と特性関数の各点収束が対応する．

定理 12 (Glivenko の定理) $\phi,\ \phi_n,\ n\in\mathbb{N},$ をそれぞれ分布 $\mu,\ \mu_n,\ n\in\mathbb{N},$ の特性関数とする．このとき特性関数の各点収束

$$\lim_{n\to\infty}\phi_n(t) = \phi(t),\quad t\in\mathbb{R},$$

と μ_n が μ に弱収束することは同値である． ◇

定理 11，定理 12 とも証明は確率論の教科書にゆだねる．

正規分布 $N(\mu, v)$ の特性関数 $\phi_{\mu,v}$ は

$$\phi_{\mu,v}(\xi) = \exp(\sqrt{-1}\mu\xi - v\xi^2/2),\ \xi\in\mathbb{R},\tag{56}$$

となる．これは

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

の指数部を平方完成

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

して，正則関数の複素積分の積分路変更の議論を経て，ガウス積分 (55) に持ち込めば分かる．

定理 8 の証明． 特性関数の使いかたの概略のみ説明する．詳しくは確率論の教科書を参照して頂きたい．

分布の平均 μ が 0 でない場合は $Y_n = X_n - \mu$ を考えればよいので，以下 $\mathrm{E}[\,X_1\,] = 0$ とする．$\dfrac{1}{\sqrt{n\,\mathrm{V}[\,X_1\,]}}\displaystyle\sum_{k=1}^{n} X_k$ の分布が $N(0,1)$ に収束することを示したいので，定理 12 から，その特性関数 ϕ_n が (56) に各点収束することを示せばよい．$\{X_n\}$ が独立同分布だから，X を同じ分布を持つ確率変数とすると

$$\log \phi_n(\xi) = n \log \mathrm{E}[\, e^{\sqrt{-1}\frac{\xi}{\sqrt{n\,\mathrm{V}[\,X\,]}}X}\,].$$

$\mathrm{E}[\,\cdot\,]$ の中の指数関数を実部と虚部に分けて，3 次以上の剰余項のテーラーの定理を用いると $|\sin x| \leqq 1$ などで剰余項を一部処理することで，剰余項項からの寄与は $n \to \infty$ で消えることが分かる．右辺を ξ^2 まで書くと $\mathrm{E}[\,X\,] = 0$ に，したがって $\mathrm{V}[\,X\,] = \mathrm{E}[\,X^2\,]$ にも，注意して

$$\log \phi_n(\xi) \sim n \log \mathrm{E}[\,1 - \frac{\xi^2}{2n\,\mathrm{V}[\,X\,]}X^2\,] \sim -\frac{\xi^2}{2\mathrm{V}[\,X\,]}\mathrm{E}[\,X^2\,] = -\frac{\xi^2}{2}\,.$$

すなわち，$\lim_{n\to\infty} \phi_n(\xi) = e^{-\xi^2/2}$ となって，$N(0,1)$ に収束することが分かる． □

5 さいころの目は不公平か？ — 検定の原理．

さいころは目の数によって材料を削り取る量が異なるため目の出る確率が正確に $p = \dfrac{1}{6}$ ずつではない，という某テレビ番組があったらしい[15]．この真偽を統計学的に検証するために，さいころを $n = 10000$ 回投げて 6 の出た回数 m を調べたところ $m = 1569$ だった．このとき 6 の出る確率は $1/6$ より少ないと結論していいだろうか？たしかに $\dfrac{m}{n} = \dfrac{1569}{10000} = 0.1569$ は $\dfrac{1}{6} = 0.1667$[16]に比べて相対的に $\dfrac{0.1667 - 0.1569}{0.1667} = 0.059 = 6\%$ くらい小さい．6% が大きなばらつきか小さなばらつきかというのが問題である．

ばらつきの範囲内で無視できると答えるにしてもさいころが不公平と答えるにしても根拠がほしい．しかし，公平なさいころでもばらつきがあるから偶然少なめに出ることもあるだろう．平均より 6% 少ないという結果は公平なさいころからも不公平なさいころからも起こりうる．「確率いくら以上で○○」という言い方以上の結論は出せない．ではどういう確率が計算できるのか？基礎的な考え方を説明する．

その前に，具体例についての具体的計算を行う準備のために理論上も実用上ももっとも重要な連続分布である正規分布についていくつかの知識を追加しておく．

5.1 （続）正規分布．

§1.3 と §4.3 で紹介した正規分布は (54) に定義したように平均 μ と分散 v という二つのパラメータを持つ 1 次元連続分布の族（集まり）で，$\mu \in \mathbb{R}$ と $v > 0$ が与えられたとき $\dfrac{1}{\sqrt{2\pi v}} e^{-(x-\mu)^2/(2v)}$ を密度とする分布のことを言う．ガウス分布とも言い，$N(\mu, v)$ と略記する．特に $N(0,1)$ を**標準正規分布**と呼ぶ．

単なる言い換えだが，$N(\mu, v)$ に従う確率変数 X に対して

$$\mathrm{P}[\, a \leqq X \leqq b\,] = \frac{1}{\sqrt{2\pi v}} \int_a^b \exp(-(x-\mu)^2/(2v))\, dx,$$
$$\mathrm{E}[\, f(X)\,] = \frac{1}{\sqrt{2\pi v}} \int f(x) \exp(-(x-\mu)^2/(2v))\, dx,$$

などとなる．

[15] この情報を提供して下さった東北大学数学教室事務の須藤砂織さんに感謝します．

[16] 5 小数点以下桁目を四捨五入して有効数字 4 桁とした．元々誤差を伴う状況なので長い桁数は意味がないので，この講義では有効桁数で切った近似値もことわりなく等号で結ぶ．

命題 13 *(i)* $N(\mu, v)$ の平均は μ 分散は v. (そこで最初から $N(\mu, v)$ のことを平均 μ 分散 v の正規分布と呼んでいた.)

(ii) $N(\mu, v)$ は $x = \mu$ に関して対称な分布である. 従って特に X が標準正規分布に従うとき, $\mathrm{P}[\, X \in A\,] = \mathrm{P}[\, -X \in A\,]$ が任意の区間 A に対して成り立つ. たとえば $\mathrm{P}[\, X \leqq -a\,] = \mathrm{P}[\, X \geqq a\,]$, $a \in \mathbb{R}$.

(iii) X の分布が $N(\mu, v)$ で c が定数のとき $Y = cX$ の分布は $N(c\,\mu, c^2 v)$, $Z = X + c$ の分布は $N(\mu + c, v)$[17]. すなわち, 定数倍とずらしに対して正規分布 (の族) は閉じていて, c 倍に対して平均は c 倍, 分散は c^2 倍になり, ずらしに対して平均はそのままずれるが分散は影響されない. 特に X の分布が $N(\mu, v)$ のとき $Z = \dfrac{X - \mu}{\sqrt{v}}$ の分布は標準正規分布.

(iv) X と Y がそれぞれ $N(\mu_x, v_x)$ と $N(\mu_y, v_y)$ に従う**独立な**確率変数ならば $X + Y$ の分布は $N(\mu_x + \mu_y, v_x + v_y)$ になる. (繰り返せば, 正規分布に従う 3 つ以上の独立確率変数の和も正規分布に従う.)

(v) 平均 μ で分散 v が有限な母集団から無作為抽出で選ばれたデータ (独立同分布確率変数列) $X_1, \cdots, X_n$ について $\sqrt{\dfrac{n}{v}}\,(\overline{X}_n - \mu)$ は n が十分大きいときほぼ標準正規分布 $N(0, 1)$ に従う. $\overline{X}_n$ は *(49)* で定義した標本平均. ◇

証明はここでは省略する (この節 §5 末の補足を参照).

命題 13 にも見られるように, 正規分布は期待値や分散やモーメント $\mathrm{E}[\, X^n\,]$ の計算が (従って, テーラー展開を考えれば一般の $\mathrm{E}[\, f(X)\,]$ の計算も) 具体的な数式で表せる. しかし, 確率そのもの, たとえば $\mathrm{P}[\, X \leqq a\,]$, $\mathrm{P}[\, X \geqq a\,]$, $\mathrm{P}[\, a \leqq X \leqq b\,]$, $\mathrm{P}[\, |X| \geqq a\,]$ などは a, b を含む初等関数では表せないことが分かっているので, 実際の確率を計算するときは適当な桁数の数値で近似的に表す. §5.1 から標準正規分布でしかも $x \geqq 0$ の場合について $\mathrm{P}[\, X \geqq x\,] = \displaystyle\int_x^\infty e^{-x^2/2} \frac{dx}{2\pi}$ が分かっていれば任意の区間の確率が計算できる. 左右対称性から $\mathrm{P}[\, X \geqq 0\,] = 0.5$ であることは分かっている.

x	$\displaystyle\int_x^\infty e^{-y^2/2} \frac{dy}{\sqrt{2\pi}}$
0	0.5
1	0.1587
1.6448	0.0500
1.9600	0.0250
2	0.0228
2.326	0.0100
2.5758	0.0050
3	0.00135

X が標準正規分布に従うとき, たとえば 99 の確率で $X \leqq 2.326$, 99% の確率で $|X| \leqq 2.578$ すなわち $X \geqq 2.578$ または $X \leqq -2.578$, 95% の確率で $|X| \leqq 1.960$, などとなる. 一般の $N(\mu, v)$ に従うときは命題 13 から標準偏差 $\sigma = \sqrt{v}$ とおくと, 99% の確率で $|X - \mu| \leqq 2.576\sigma$, 95% の確率で $|X - \mu| \leqq 1.960\sigma$, などとなる. 標準正規分布では $\mathrm{P}[\, |X| \geqq 3\,] = 0.0027$ なので, $N(\mu, v)$ に従う X が平均から 3σ 以上離れるのは 1000 回中 3 回に満たない. うそつきのことをせんみつ (千三) という古語があるそうだが, 統計学では「3σ はずれたデータだ」(信じられない) というのがそれに対応することになる. もちろん, 平均からはずれているから悪いわけではない. せんみつの別の意味に不動産業がある. 商談の成立するのが千口のうちで三口ほどという意味で, たいへんな職業だ.

[17] 大部分の統計学の教科書は X の従う分布や $Y = cX$ 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 Ω に取り直して $X(x) = x$ などと考えていると思えばよい.

5.2 検定の原理.

検定とは無作為抽出された標本（制御できない誤差を伴う実験等のデータを含む）をもとに，母集団に関する命題を何らかの形で成立確率に言及した上で立てることを言う．母集団に関する仮説を選び，その仮説下で適当な統計量である確率変数を選び，実現した標本 w（確率変数の w での値，統計量の実現値）に基づいて仮説の正否を判断する．

仮説が正しくないことを主張する形をとることが多いため，立てた仮説を**帰無仮説**と呼ぶ．常識または既存の理論を仮説に立てて実測によって常識を覆す（新しい自然法則の発見），あるいはまた，問題のない正常状態を仮説として実測によって否定されれば異常が起こったと判断して対策を立てる（抜き取り検査）．

例 14 母集団についての大前提：日本人の身長の分布は正規分布である．
帰無仮説 H：*50* 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：*50* 年前と現在それぞれの *50* 例（たとえば学級 1 クラス分）の身長の標本平均と不偏分散． ◇

このときの統計的検定の方法は確立している（正規母集団に関する小標本理論）．§8 でその方法を具体的に説明する．概要を先取りして説明すると，まず，母分散は変化している証拠はないことを検定する（母分散の変化の可能性までこめて平均の差を検定する Welch の検定というのもあるが §8 では省略する）．次に標本平均の今昔の差と全データの不偏分散の比をとる．正規母集団の前提があるのでこの比（確率変数）の従う分布は具体的に求まる（t 分布）から標本から得られる比の値が起きる確率を議論できる．大ざっぱに言うと，この比が非常に大きいというデータが得られたとすると，帰無仮説 H が正しいならそんな大きな比が起きる確率 p はたいへん小さい．あらかじめ意志決定の基準として定めておいた数値 α（危険率）より p が小さいときに H を棄却する．

定義 15 以下の手続きを検定と呼ぶ．小さいと自分が判断する $\alpha > 0$ を固定し，仮説 *H:*「母集団の分布は P である」を立てる．

$$\alpha = P(A) \tag{57}$$

を満たす事象 A を決めておき，**棄却域**（危険域，*critical region*）と呼ぶ．通常は P が実数上の分布で A は中央付近の区間の外側に選ぶ（$\mathbb{R}^n$ 上の分布でも同様に決められる）が，この選び方には議論がある．
サンプル x を得たとき，$x \in A$ ならば H を**棄却** *(reject)* する．つすなわち「母集団の分布は P でない」と結論する．$x \notin A$ ならば H を**採択** *(accept)* する．すなわち「母集団の分布は P 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．
α を**有意水準** *(significance level,* 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*], [*林周二*, 13.3].* ◇

注 16 全数調査をすれば平均身長は正確に決まる．このとき $1mm$ でも平均身長に差があれば確定的に差がある．しかし，身長はいつの時代も何 $10cm$ もの幅で分布するので，平均身長に $1mm$ の差があっても意味を持つことはほとんどない．母平均の差をなぜ知る必要があるのか検定する前に考える必要がある．統計学では意味を問うことなく標本から母集団について論じることのできる確率を算出する． ◇

第 1 種の過誤と第 2 種の過誤.

H が本当は正しいのに**うっかり**して棄却することを**第 1 種の過誤**と言う．危険率 α は第 1 種の過誤が起きる確率である．

例 17 新製品の試作品の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． ◇

α は通常小さく 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，それは根拠がある．(と，安心していては間違うこともあるので実用上は注意しないといけない．危険域をどう取るかは意志決定の問題，最近流行の言葉で言えば，統計学の利用者の自己責任である．統計学は 'as is' で意志決定支援情報を提供するだけである．)

> H が本当は間違っているのに**ぼんやり**して採択することを**第 2 種の過誤**と言う．ここまでの議論では第 2 種の過誤の確率は決まらない．

例 18 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． ◇

H が採択された場合，H が本当に正しい確率論的根拠はここまで説明した検定手続に内在していない．場合によってはその確率は高いかもしれない．H が帰無仮説と呼ばれ，棄却するつもりで立てる理由はここにある．

5.3 さいころの公平性の検定．

§5 節始めの問題に戻って，さいころを $n = 10000$ 回投げて 6 の出た回数が $m = 1569$ だったとき 6 の出る確率は 1/6 より少ないと結論していいか，という問題を考えよう．

統計学の問題とするために，帰無仮説 H: $p = 1/6$ を危険率 1% で検定することにしよう．

統計量は 6 の出た割合 $X = \dfrac{m}{n}$ で標本は $x = X(w) = \dfrac{1569}{10000}$.

定理 3 で中心極限定理（定理 8）を 2 項分布の場合に先取りした．そのときは硬貨と書いたが，p が同じなら硬貨でもさいころでも同じことだから，そこでの N_n が今回の $m = n\,X$ である．従って定理 3 から，帰無仮説 H の下で $Y = \sqrt{\dfrac{n}{p(1-p)}}(X - p)$ の分布は $n \to \infty$ のとき標準正規分布 $N(0,1)$ に弱収束するので，n が十分大きいと考えて Y の分布は $N(0,1)$ と近似する．

危険率 $\alpha = 0.01$ だから $\mathrm{P}[\,|Y| > a\,] = 2 * \mathrm{P}[\,Y > a\,] = 0.01$ となる a を正規分布の表から求めると，$a = 2.5758$．よって H を $|Y| > 2.576$ のとき棄却し $|Y| < 2.576$ のとき採択する．すなわち

$$|X - p| > 2.576\sqrt{\frac{p(1-p)}{n}} = 2.576\sqrt{\frac{5}{6^2 \times 10000}} = 0.0096,$$

より，$|m - np| > 0.0096n = 96$ のとき棄却する．

$np = 1667$, $m = 1569$ ならば $|m - np| = 98$ だから H は棄却される．$n = 10000$ 回だと約 100 回平均 np よりよぶんにまたは少なく出れば H を危険率 1% で棄却できるが，そうでなければ目の出方が不均一とは言えない．

n を変えるとどうなるかは演習問題とする．答は n とともに平均 np からの採択域の広がりが $\sqrt{n}$ に比例する（相対的には $\sqrt{n}/n$ で狭まる）．

n が小さいときは正規分布近似は悪いので 2 項分布に戻って計算し直さないといけないが，仮にさいころの目の出方が均一でないとしても，$n = 10000$ くらいでは統計的に検出できないというのが筆者の直感である．(そんな簡単に差が見えれば賭博師の間で有名になったはずで，確率論の起源にその逸話が残っていないのは不自然である！)

補足：命題 13 の証明．

命題 13 の証明． (i) 記号の便利のため $N(\mu, v)$ に従う確率変数を X とおく．

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

よって，n が奇数ならば $\mathrm{E}[\,(X-\mu)^n\,] = 0$．特に $n = 1$ のとき $\mathrm{E}[\,X\,] = \mu$．すなわち，$N(\mu, v)$ の平均は μ．

ガウス積分 (55) で $x = y\sqrt{a}$ と変数変換すると $\displaystyle\int_{-\infty}^{\infty} e^{-ay^2/2}\,dy = \sqrt{\frac{2\pi}{a}}$. a で微分して $a = 1/v$ とおくと

$$-\frac{\sqrt{2\pi v}}{2}M_2 = -\frac{1}{2}\int_{-\infty}^{\infty} y^2 e^{-y^2/(2v)}\,dy = -\frac{v}{2}\sqrt{2\pi v}\,.$$

よって $\mathrm{V}[\,X\,] = \mathrm{E}[\,(X-\mu)^2\,] = M_2 = v$ すなわち，$N(\mu, v)$ の分散は v.

(ii) $N(\mu, v)$ の密度を見れば明らか.

(iii) 変数変換 $y = cx$ によって

$$\mathrm{P}[\,Y \leqq a\,] = \mathrm{P}[\,X \leqq a/c\,] = \int_{-\infty}^{a/c} \frac{1}{\sqrt{2\pi v}} \exp(-(x-\mu)^2/(2v))\,dx$$
$$= \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi v c^2}} \exp(-(y-\mu c)^2/(2vc^2))\,dy.$$

よって $Y = cX$ の分布は $N(c\,\mu, c^2 v)$.

変数変換 $z = x + c$ によって

$$\mathrm{P}[\,Z \leqq a\,] = \mathrm{P}[\,X \leqq a - c\,] = \int_{-\infty}^{a-c} \frac{1}{\sqrt{2\pi v}} \exp(-(x-\mu)^2/(2v))\,dx$$
$$= \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi v}} \exp(-(z-c-\mu)^2/(2v))\,dz.$$

よって $Z = X + c$ の分布は $N(\mu + c, v)$.

(iv) 期待値と独立確率変数の分散の加法性は (29) と (43) で紹介したので，パラメータが $\mu_x + \mu_y$, $v_x + v_y$ になることは既知．ここで新しいのは**正規分布に従う独立確率変数の和は正規分布に従う**という点である．

正規分布の場合は特性関数を使ったほうが証明が圧倒的に早いが，最低限の知識で証明するため，正規分布でなくても一般の密度を持つ確率変数の和の分布で成り立つ次の性質を用意する．

補題 19 X と Y_k がそれぞれ密度 ρ_X と ρ_Y を持つ連続分布に従う独立確率変数ならば，$X + Y$ は密度 $\rho_s(x) = \int_{-\infty}^{\infty} \rho_X(t)\,\rho_Y(x-t)\,dt$ を持つ連続分布に従う． ◇

証明． 独立確率変数なので命題 5 の (48) より，(X_1, X_2) の従う分布の密度は

$$\rho(x, y) = \rho_1(x)\,\rho_2(y), \quad x, y \in \mathbb{R},$$

で与えられる．よって $z = x + y$，$t = x$ によって (x, y) から (t, z) に積分変数変換すると

$$\mathrm{P}[\,X + Y \leqq a\,] = \int\int_{x+y \leqq a} \rho_1(x)\,\rho_2(y)\,dx\,dy = \int_{-\infty}^{a} \left(\int_{-\infty}^{\infty} \rho_1(t)\,\rho_2(z-t)\,dt \right) dz.$$

よって $X + Y$ の分布は密度 $\int_{-\infty}^{\infty} \rho_1(t)\,\rho_2(z-t)\,dt$ に従う． □

注 20 X と Y 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 $X + Y$ 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, 命題 5.6 の注]*. X と Y が独立という仮定なので，$X + Y$ はノルウェー人と日本人のペアごと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． ◇

補題 19 は正規分布でなくても成り立つが，正規分布の場合に戻る．正規分布の場合 X を定数だけずらすと平均がずれるだけで再び正規分布になることはすでに見たので（X と Y が正規分布に従えば $X + Y$ も正規分布に従うことを証明する予定なので），$\mu_x = \mu_y = 0$ の場合だけ証明すればよい．X と Y がそれぞれ $N(0, v_x)$ と $N(0, v_y)$ を分布に持つ独立確率変数とする．$X + Y$ の分布の密度 ρ_s は補題 19 から

$$\rho_s(z) = \frac{1}{2\pi\sqrt{v_x v_y}} \int_{-\infty}^{\infty} e^{-t^2/(2v_x) - (x-t)^2/(2v_y)} dt$$

指数部を t について平方完成すると，

$$-\frac{t^2}{2v_x} - \frac{(x-t)^2}{2v_y} = -\frac{v_x + v_y}{2v_x v_y} \left(t - x\frac{v_x}{v_x + v_y} \right)^2 - \frac{x^2}{2(v_x + v_y)}.$$

t を含む因子はガウス積分 (55) で計算できて

$$\rho_s(z) = \frac{1}{2\pi\sqrt{v_x v_y}} \int_{-\infty}^{\infty} e^{-\frac{v_x+v_y}{2v_x v_y} t^2} dt e^{-\frac{x^2}{2(v_x+v_y)}} = \frac{1}{\sqrt{2\pi(v_x + v_y)}} e^{-\frac{x^2}{2(v_x+v_y)}}.$$

これは $N(0, v_x + v_y)$ の密度である．

(v) これは中心極限定理（定理 8）そのもの． □

補足：第 2 種の過誤．

§5.2，§5.3 と §6 でそれぞれ原理を紹介した（する）検定と推定の関係，および，検定における第 2 種の過誤の意味は，パラメータを持つ母集団分布（母集団分布の族）を考えると明らかになる [小針, §8.1], [楠岡], [辞典, 284A].

無から有は生まれない． まず，明らかにならないことから．

この講義ではもっぱら母集団のパラメータ（§5.2，§5.3 では検定の対象）として平均と分散のみをとりあげ，§7 や §8 では正規分布の族に限る．これらのパラメータ空間（母集団の族）の設定はこの講義でいう統計学的推測からは決まらない．そもそも無数の可能性のある母集団について，有限のできるだけ小数のデータだけで推測したいのだから，もっとも欲張ってもデータ数を越える数のパラメータを決めることはできない．多くの場合，少しでも意味のある結果を導きたければパラメータ数はデータ数よりきわめて小さく取る必要がある．従って，母分布の形の「大部分」は別の理論的根拠から問題毎に決めておかないといけない．しばしば正規分布（の族）が用いられるのは，中心極限定理（定理 8）などの「理論的」根拠（とパラメータ数の少なさ）によるが，硬貨投げや視聴率は問題の定義から 2 項分布が基本だし，他の分布が理論的に適切な場合もある．

もう一点，§5.2 や §5.3 で設定した危険率 α は「母集団が帰無仮説 H に従うとすると，現に今出ているデータが起きる確率は小さすぎて，そんな異常事態が起きるとは信じられない（ので H を棄却する）」という論法である．標本が得られる確率が論じられるのであって,「母集団が H で仮定される特定の分布になる（パラメータがある場合はそのパラメータをとる）確率」では**決してない**．だから H を採択するときは，α に強い意味がなく,「H が棄却できない」という言い方のほうが一般には実体に近い，と注意した．母分布（より正確には母分布のパラメータ）について確率を論じるためには母分布（のパラメータ）空間上に何らかの確率が定義されていなければならない．これを**先験的確率**と呼ぶことがある．無から有は生まれない．先験的確率は統計学の多くの場合「神にしか分からない」確率なので，古来その解釈について議論が重ねられた.「現実的」な理解も進んで，統計学的推測の手法の一般化に役だった面もあるが，議論が果てしなく続くのでこの講義では深入りを避ける．

対立仮説と第 2 種の過誤と検定力． §5.2 で，H が採択された場合，H が本当に正しい確率論的根拠はここまで説明した検定手続に内在していない，と注意した．実はそのような確率を理論に組み込む方法は存在する．

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．このような仮説を**対立仮説**と呼ぶ．§5.2 ではそのことを明示しなかったが，H を棄却すれば,「それ以外のことが起きている」と主張したのだから，暗黙に「それ以外」が存在する．たとえばさいころの例ではパラメータ p （不公平さの度合い）を変えることがが仮説の一群である．

一般に，既存の理論や常識（帰無仮説）と新説や新発見（対立仮説）の対決，正常状態と予想される異常事態，の比較をデータによって行うのが検定である．

話を具体的にするために，想定されうる母集団の分布 P_θ が一つのパラメータ θ だけを持つとし，帰無仮説の分布は P_{θ_0}，対立仮説の分布は P_{θ_1} とする．たとえば，不公平なさいころの例では $\theta = p$，$\theta_0 = 1/6$，そして θ_1 は各目の面で削り取った量から何らかの（筆者の知らない）理論物理学的計算によって予想される値[18]とする．

危険率 α を自分で選び，危険域 A を (57) すなわち $\alpha = P_{\theta_0}(A)$ となるように決め，$x \in A$ ならば H を棄却するのであった．第 1 種の過誤が起きる確率は α そのものである．(57) だけでは A は決まらないことに注意．§5.2 では $\mathbb{R}$ 上の分布の場合 A は通常区間の外側の領域に選ぶとだけ書いた．その決め方がここで問題になる．

さて,「理論」どおり帰無仮説か対立仮説のいずれかが成り立つとすると，第 2 種の過誤は，本当は P_{θ_1} が正しい母集団なのに $x \in A^c$ なるデータを得たとき起きるので，その確率 β は

$$\beta = P_{\theta_1}(A^c) \tag{58}$$

である．P_θ は θ が異なれば違う分布なので，A のとりかたによって α が等しくても一般には β が変わる当然 α が同じなら β が小さくなる A のとりかたのほうが良い．これが危険域の形 A を決める (Neyman–Pearson) [辞典, 284B], [楠岡]．同じ α に対して β の小さい検定方法を**検定力**が強いと言い，$1-\beta$ を検定方式 (α, A) の検定力という．θ_1 を動かすときは，その関数とみて検定力関数ともいう．最強力検定関数の特徴付けについては Neyman–Pearson の定理 [辞典, 284B], [楠岡], [林周二, 14.3] を参照．

例えば θ が母平均を表すパラメータ（今までの記号ならば $\theta = \mu$）で $\theta_1 > \theta_0$ のとき，正規分布や 2 項分布のような分布ならば $A = [a, \infty)$ の形にとるのが明らかに検定力が強い．すなわち，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 β を最小に抑える．これが危険域の形を決める原理である．具体的には，硬貨投げにおける硬貨の公平性の検定において,「一部の硬貨は表が出やすい細工があり，残りは公平で，現場でどちらが使われたかは分からない.」という情報が入った場合に相当する．表が出にくいケースを考える必要が無くなるので，片側危険域をとることになる．同じ信頼係数 α （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く，検定力が高い．当然裏が多めに出たデータの場合は「偶然多く出ただけで，細工の事実なし」と結論することになる．この例から，危険域が一般に区間の補集合にとられる理由も分かる．正規確率が小さいと言うだけで中ほどの短い区間を気まぐれにくりぬいたりはしない．

通常片側危険域ではなく両側危険域を採るのは，$\theta_1 > \theta_0$ のようなはっきりした偏りのある対立仮説を採れることが少ないからである．つまり，念頭にある対立仮説が $\theta_1 < \theta_0 < \theta_2$ なる θ_1 と θ_2 を半々の確率（先験的確率）でとる状況である．このように，素朴な経験に基づく危険域の決め方は頭の中に対立仮説があると理解できる．

[18] 削り取る量が多いから 1/6 より小さいのだろう．筆者に想像できるのはそれだけ.）

第 1 種の過誤と第 2 種の過誤． 以上のように母分布のパラメータ空間を設定して帰無仮説とともに対立仮説を用意すれば，第 2 種の過誤の確率 β も第 1 種の過誤の確率 α と同様に議論できる．

とすれば α を制御しつつ β への定量的評価をあらわにしないという，伝統的な両者の扱いの落差はどこから来るのか？ここに，人の悪い筆者が教科書の一節を裏読みして気づいた第 2 種の過誤を軽視する「人間性の理由の可能性」を記しておく．以下の段落をどう読むかは読者の判断に任せる．

(i) **科学的禁欲主義** [楠岡, §6.3]．第 1 種の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 2 種の過誤は新理論が正しいのに既存の理論に執着すること．既存の権威にできるだけ逆らわない人間性の本質がある[19]．

(ii) **生産者の損得優先主義** [林周二, 13.3]．抜き取り検査では第 1 種の過誤は良品を不合格とする生産者の損失，第 2 種の過誤は不良品を合格とする消費者の損失．生産者の損得はきちんと計算するが消費者の損得は生産者の都合の範囲内でしか考えない人間性の本質がある[20]．

6 視聴率調査，何人分調べれば十分か？— 区間推定の原理と大標本理論．

正規分布に基づく推定・検定が広く用いられる．期待値と分散で決まる 2 パラメータの分布族という単純さの割に，確率変数の和と定数倍について閉じていること（命題 13），そして中心極限定理（定理 8），すなわち，分散有限な独立同分布確率変数列 $\{X_k\}$ について $S_n = \dfrac{1}{\sqrt{nv}}\displaystyle\sum_{k=1}^{n}(X_k - \mathrm{E}[\,X_1\,])$ が $n \to \infty$ で標準正規分布 $N(0,1)$ に収束することがある．後者は次の理屈で統計学に用いられる．

(i) 制御できない無数の独立な要因によって分布が生じている場合，その分布は正規分布に近づくだろう．従って，種々の要因で分布が起きうる複雑な対象の母集団は正規分布に近いとして差し支えあるまい．（実験誤差や人に関する母集団を正規分布と仮定する「根拠」または「いいわけ」となる．）
正規母集団に対する（残されたパラメータである平均と分散に関する）推定や検定を小標本理論と呼び §7，§8 で紹介する．

(ii) 未知の母集団であっても，そこから無作為抽出で十分大きなデータを集めれば，中止極限定理（定理 8）から標本平均 $\overline{X}_n$ は $N(\mu, \frac{v}{n})$ にほぼ従う（正しくは，$\sqrt{n}(\overline{X}_n - \mu)$ が $N(0,v)$ にほぼ従う）．ここで，μ, v は（正規分布とは限らない）母集団の平均と分散，n はデータの大きさ．母集団分布に関係なく，データが大きいときの標本平均が正規分布にほぼ従うとして行う推定・検定を大標本理論と呼ぶことがある．
既に §5.2 で紹介した検定の原理も 2 項分布を正規分布で近似して計算した．本節 §6 では §3 の続きで視聴率調査の例について，実質上同じ問題を推定の問題として説明する．
なお，大標本理論は主に標本平均について考えるが分散の推定・検定にも応用できる．不偏分散は独立確率変数の和の形に書かれていないが，母平均 μ が既知のときの標本 $\{X_k\}$ の分散を $V_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(X_k-\mu)^2$ とでもおくと，これは独立確率変数列 $Y_k = (X_k - \mu)^2$, $k = 1, 2, \cdots$, の和を n で割った形だから平均 $\mathrm{E}[\,Y_1\,] = \mathrm{E}[\,(X_1 - \mu)^2\,] = v$（母分散），分散 $\dfrac{1}{n}\mathrm{E}[\,(Y_1 - \mathrm{E}[\,Y_1\,])^2\,] = \dfrac{1}{n}\mathrm{E}[\,((X_1 - \mu)^2 - v)^2\,]$ の正規分布に近い（繰り返しだが，これも正しくは $\sqrt{n}(V_n - v)$ が $N(0, \mathrm{E}[\,((X_1 - \mu)^2 - v)^2\,])$ に近い）．

6.1 区間推定の原理．

母集団の分布は何でも（既知でも未知でも）かまわない．その平均 μ が未知で，分散 v は既知（あるいは μ から計算可能）とする．標本平均 $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ の標本 $\overline{X}_n(w)$ が得られたとき，母平均 μ を推定したい．$\mu = \mathrm{E}[\,X_1\,]$，$v = \mathrm{V}[\,X_1\,]$ に注意すると，大標本理論はデータ数 n が大きいとき中心極限定理（定理 8）によって $\sqrt{n}(\overline{X}_n - \mu)$ が $N(0,v)$ に近づくことを根拠とする．推定される区間の幅はデータ数 n による．（母集団の大きさと無関係なことに注意．視聴率調査の「精度」は調査対象数で決まり，日本全国か関東地域かは関係ない．）

[19] 善悪ではなく人間とはそういうものだという自然法則である．そういう逃れがたい呪縛からの脱却だからこそ，創造的研究は ‘challenge’ と呼ぶにふさわしい．

[20] これも自然法則に過ぎない．必然的に消費者も対抗監視する論拠を持たなければならない．

定義 21 母集団が平均 μ のみが未知とする．つまり母集団は μ をパラメータとする分布の族 P_μ に入っている．μ を標本から区間推定するとは以下の手続きを言う．
検定と同様に小さい確率 α を決めておく．つまり確率 α 以下の現象は「あり得ない」として切って捨てる決断をする．$1-\alpha$ を**信頼水準** *(confidence level)* と呼ぶ．標準正規分布の数表から $N(0,1)([-a,a]) = 1-\alpha$ なる a を求めておく．
$\sqrt{\frac{n}{v}}(\overline{X}_n - \mu)$ は n が大きいとき中心極限定理（定理 *8*）から標準正規分布に近いから，もし μ が既知ならば $\overline{X}_n$ は確率 $1-\alpha$ で $-a \leqq \sqrt{\frac{n}{v}}(\overline{X}_n - \mu) \leqq a$ を満たす．これは普通の確率論である．
μ が未知のときこの式を逆に読んで，標本 $x = \overline{X}_n(w)$ が与えられたとき信頼水準 $1-\alpha$ で（**危険率** α で）

$$x - a\sqrt{\frac{v}{n}} \leqq \mu \leqq x + a\sqrt{\frac{v}{n}}$$

と μ を推定する．この区間

$$[x - a\sqrt{\frac{v}{n}},\ x + a\sqrt{\frac{v}{n}}] \tag{59}$$

を**信頼区間**と呼ぶ． ◇

注 22
- 手続きから，信頼区間とは，そこからパラメータ μ がはずれているのになおかつ手元に得られた標本が出てくるのは小さい確率だから信じられないという意味である．α は μ が起きる確率とは無関係である．信頼区間でもってパラメータの推定値としようというのである．候補となる分布の一つ一つについて，それが母集団だった場合に標本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，標本の珍しさの許容限界を示すのが危険率である．危険率の意味は検定と同じである．
- *(59)* から幅は $\sqrt{n}$ に反比例するので n 大なほど信頼区間の幅は狭い．データを集めることの意味が分かるだろう．俗称で**統計誤差**と呼ばれているものがこれである．実験や調査の報告をするときは（たとえば $\alpha = 0.01$ のとき），$\mu = x \pm a\sqrt{v}n$ *(99% confidence level)* 等のように書く．
- 危険率 α 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．
- m が未知なのに v が既知というのは奇異に見えるかもしれないが，実用的には，v の推定値として不偏分散の標本値 $V_n(w)$ をデータから計算して用いればそんなにはずれない．$V_n(w)$ が v に近いというのは大数の法則（定理 *7*）である．ここでもデータの大きさ n が大きいことが必要である（大標本理論）．
- 母集団についての具体的なことが分かれば計算できる場合もある．正規分布の場合は §7 や §8 で μ, v とも未知な場合について厳密に扱う．

◇

6.2 例題 — 視聴率調査．

一人ごとの調査結果は見た $X_1 = 1$ か見ない $X_1 = 0$ で，真の視聴率を p とすると (40) で注意したとおり $\mathrm{P}[\,X_1 = 1\,] = p, \mathrm{P}[\,X_1 = 0\,] = 1-p$ なので，$\mu = \mathrm{E}[\,X_1\,] = p,\ v = \mathrm{V}[\,X_1\,] = p(1-p)^2 + (1-p)(0-p)^2 = p(1-p)$.

例として信頼水準 99% で推定することにすると $\alpha = 0.01$．§5.1 の正規分布の表から $a = 2.576$ とおけば $N(0,1)([-2.576, 2.576]) = 0.99$ となる．視聴率調査で得られた標本（見ていた人と調査対象人数との比）

が $x = \overline{X}_n(w)$ だったとすると，(59) から信頼区間は $[x - 2.576\sqrt{\frac{v}{n}},\ x + 2.576\sqrt{\frac{v}{n}}]$ 大標本理論では v を不偏分散 $V_n = \frac{1}{n-1}\sum_{k=1}^{n}(X_k(w) - \overline{X}_n(w))^2$ で近似するのであった．

このケース（2 項分布）の場合は $v = p(1-p)$ なので $p \fallingdotseq \overline{X}_n(w)$ で近似して $v = \overline{X}_n(w)(1 - \overline{X}_n(w))$ とする方法もある．さらに，

$$x - 2.576\sqrt{\frac{p(1-p)}{n}} \leqq p \leqq x + 2.576\sqrt{\frac{p(1-p)}{n}} \tag{60}$$

は p に関する連立 2 次不等式なので，それを解くことで p の区間を得るのがもっとも理論的にしっかりしている．ただし最後の方法は 2 項分布特有の方法ではある．また，どの方法をとっても実用上数字の上では大差ない．

たとえば視聴率調査はデータの大きさ n が 1 万人，その調査結果上は 30% の大ヒット番組と出たとしよう．$x = \overline{X}_n(w) = 0.3$ なので，安直に $v = \overline{X}_n(w)(1 - \overline{X}_n(w)) = 0.3 \times 0.7 = 0.21$ とすれば信頼水準 99% で求める信頼区間は $0.3 \pm \frac{2.576\sqrt{0.21}}{100} = 0.3 \pm 0.01$ 29% から 31% の間と見ることができる．

信頼区間は母集団の総数によらないことに注意しておく．日本全体の視聴率でも関東の視聴率でも調査すべき人数は変わらない．

(60) に戻って，$0 < p < 1$ のとき $p(1-p) \leqq 0.25$ なので，推定された区間の幅の半分は $2.576\sqrt{\frac{p(1-p)}{n}} \leqq \frac{1.288}{\sqrt{n}}$ なので約 $n^{-1/2}$ である．視聴率の整数部分を（信頼水準の範囲で）決めるためには調査数 n を約 4 万用意したほうがよい．頻繁に変えなければならないことや地域や年齢層などをまんべんなく備えなければならないことを考えると，これは実用上簡単な数字ではないというのが筆者のカンである．もしかしたら信頼水準を 95% くらいに落としているかもしれない．時々視聴率がパーセントで小数点以下表示されているテレビ番組があるが，どう考えてもナンセンスである．

6.3 例題 2 — 硬貨投げ，推定と検定．

例 23 練習のため次の例を考えてみる．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 とすると §6.2 の視聴率と全く同じ問題であることが分かる．

平均 $\mu = \mathrm{E}[\,X_1\,] = p$，分散 $v = \mathrm{V}[\,X_1\,] = p(1-p)$．標本のほうは $n = 400$，$x = \overline{X}_n(w) = \frac{220}{400} = 0.55$ である．§6.2 と同様に $\sqrt{v} = \sqrt{x(1-x)} = 0.497$ で近似しよう．危険率 α に対して $N(0,1)([-a,a]) = 1-\alpha$ なる a を $a = a(\alpha)$ とおくと (59) から信頼区間は $0.55 \pm \frac{0.497}{20}a = 0.55 \pm 0.025a$ §5.1 の正規分布の表から $\alpha = 0.05$ と 0.01 の場合を計算しておこう．

α	0.05	0.01
a	1.960	2.576
μ	0.550 ± 0.049 $[0.501, 0.599]$	0.550 ± 0.064 $[0.486, 0.614]$

信頼水準を高くとるほど信頼区間の幅が広がることに注意．信頼水準を高くとるということは慎重に可能性を残すことを意味するので発言内容は曖昧になる．慎重がよいか曖昧がいけないかは意志決定する側の判断である．

上記の例において，この硬貨は表が出る確率が p_0 か？という検定の問題に変えてみよう．

検定の原理（§5.2）を思い出して，まず帰無仮説 H: $\mu = p_0$ を立てる．

このとき $\mu = \mathrm{E}[\,X_1\,] = p_0$，$v = \mathrm{V}[\,X_1\,] = p_0(1-p_0)$．§6.1 の一般論に戻って，$\sqrt{\frac{n}{v}}(\overline{X}_n - \mu)$ は n が大きいとき $N(0,1)$ にほぼ従う．

$N(0,1)([-a,a]) = 1-\alpha$ なる a を $a = a(\alpha)$ とおいて，危険域 A を $[p_0 - a\sqrt{\dfrac{p_0(1-p_0)}{400}}, p_0 + a\sqrt{\dfrac{p_0(1-p_0)}{400}}]$ の**補集合**にとる．そしてデータ $x = \overline{X}_n(w) = 0.55$ が $x \in A$ ならば H を棄却，$x \notin A$ ならば H を採択する．

たとえば $p_0 = 0.5$，すなわち帰無仮説 H として「硬貨が公平である」をとると，$\alpha = 0.05$ と 0.01 の場合は次のようになる．

α	0.05	0.01
a	1.960	2.576
A^c	$[0.436, 0.564]$	$[0.451, 0.549]$
$x = 0.55$	$x \in A^c$	$x \in A$

標本の値が $x = \overline{X}_n(w) = 0.55$ のとき，表から信頼係数 95% ならば $x \in A$ だから H は棄却される．すなわちこの硬貨は不公平である．信頼係数 99% ならば $x \in A^c$ だから H は採択される．すなわちこの硬貨は不公平とは言えない．

これは推定の問題として扱った際に，仮説 H の値 0.5 が $\alpha = 0.05$ のときは信頼区間 $[0.501, 0.599]$ からはずれており，$\alpha = 0.01$ のときは信頼区間 $[0.486, 0.614]$ に入っている，ということと対応している．

予想があるとき検定，ないとき推定，と言うのが素朴な区分けだが，推定といえども母集団の形について暗黙の（時には理論的な）仮定があるので，つきつめて考える（上記のようにどちらもパラメータを持つ母集団の族を念頭に置いていることに注意する）と，両者に殆ど差がない．古典的に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統合して用いられる感もある．

なお，硬貨やさいころの例ばかり挙げると，賭博師にしか関係ないと思われそうだが，半導体などの大量工業生産品で，ゴミや製造機械の一時的調整不良で不良品がロット（あるまとまり）単位で集中的に発生しやすい場合，そのような不良ロットの検出のための抜き取り検査の設計（何個抜き取れば不良品率を適切な範囲に抑えられるか）等も基本的に 2 項分布に基づく検定である．

6.4 例題 3 — 事故，母集団がポワッソン分布の場合．

中心極限定理があるので §6 冒頭に挙げたいずれかの理由によって正規分布に基づく推定・検定の計算が行われることが多い．実際，率直に言って，たとえば意志決定の上で 96% と 95% の間に何ほどの差もないので，分布の形を少々複雑にしても実用上は変わらないから，正規分布で全てすましてしまうのは思考の節約という意味で良いことかもしれない．**ただし**，本質的な差を生じる場合には気を付けないといけない．

たとえば自動車事故のように「起こりうる場所（道の多さや人や車の多さ）が極めて多いが 1 箇所あたりの起きる頻度が極めて小さく，平均生起数 λ が目に見える量になる（0 でない）」現象は，一定期間内の発生数がポワッソン分布（§10.2）と呼ばれる分布に従うと考えられる．そういう現象が k 件起きる確率は

$$Q_\lambda(\{k\}) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0, 1, 2, \cdots,$$

で与えられる．生起件数という意味づけからは当然だが $\Omega = \mathbb{Z}_+$ 上の分布である．

ちなみに，定数 $e^{-\lambda}$ は $Q(\mathbb{Z}_+) = 1$ と e^x のテーラー展開から決まる．2 項分布で言うと，n が大きく p が小さく，平均 $np = \lambda$，ということである．実際 $B_{n,\lambda/n}$ において $n \to \infty$ とすると，平均 λ のポワッソン分布に弱収束することが分かる．（ここで離散分布の離散分布への弱収束とは，各 k 毎の収束である．）

教科書から例を引用しておく．

例 24 ([小針, §8.2 例題 3]) ある町では 1 日平均 1 人の交通事故死者がある．1 週間の交通事故死亡者数を信頼率 95% で予測せよ．また 1 週間に 14 人以上の交通事故死者が出たら異常な事態と言えるか？

解． 1 週間では死者数は平均 7 のポワッソン分布に従うことになるので 1 週間の死者の分布は

$$Q(\{k\}) = \frac{7^k}{k!}e^{-7}, \quad k = 0, 1, 2, \cdots,$$

で与えられる（図 *1*）．図は *Mathematica* に以下の命令を実行させると得られる．

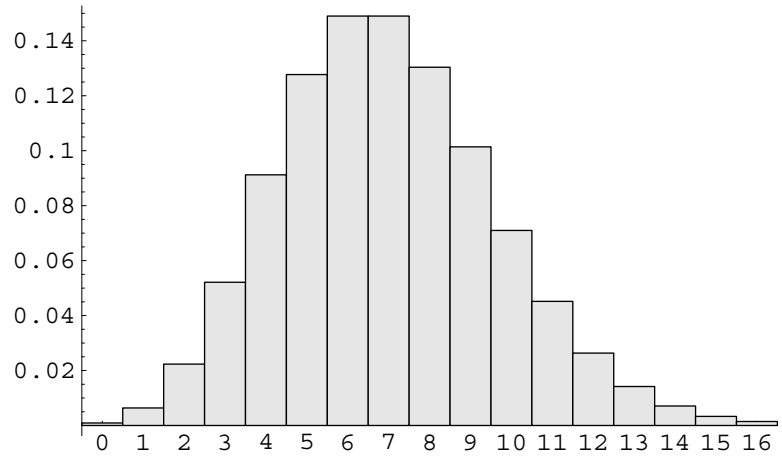

図 1: 平均 7 のポワッソン分布.

```
Q[k_]:=N[7^k/k! * Exp[-7]];
Remove[BarChart,BarSpacing, BarGroupSpacing,BarLabels, BarStyle]
<<Graphics‘Graphics‘;
BarChart[Table[Q[k],{k,0,16}], BarSpacing->0, BarGroupSpacing->0,BarLabels ->
 Table[k,{k,0,16}], BarStyle->{GrayLevel[0.9]}]
```

信頼率 $1-\alpha$ に対する信頼区間を $[a,b]$ の形にとる．$a=a(\alpha)$, $b=b(\alpha)$ は信頼率の関数．ポワッソン分布は左右非対称なので正規分布のように平均 7 を中心として左右対称にとる必然性はない．推定・検定の原則から考えて，平均 7 を含み，確率の大きい k を順次とるのがもっとも偏りがないだろう．そうやってとったとき $Q([a,b]) \geqq 1-\alpha$ となる中でいちばん狭い区間 $[a,b]$ を $[a(\alpha),b(\alpha)]$ と定める．

Mathematica の計算結果

```
Reverse[Sort[
  Table[ {k, N[7^k/k! * Exp[-7]]}, {k,1,15} ],
  OrderedQ[{Last[#1], Last[#2]}]&
           ]
      ]

{{7,0.149003},{6,0.149003},{8,0.130377},{5,0.127717},{9,0.101405},
 {4,0.0912262},{10,0.0709833},{3,0.0521293},{11,0.0451712},{12,0.0263498},
 {2,0.0223411},{13,0.0141884},{14,0.00709419},{1,0.00638317},{15,0.00331062}}

G[ord_]:=Last[Part[%1, ord]]
Sum[G[ord],{ord,1,10}]   0.943364
Sum[G[ord],{ord,1,11}]   0.965705
Sum[G[ord],{ord,1,13}]   0.986988
Sum[G[ord],{ord,1,14}]   0.993371
```

よって信頼率 95% ，すなわち，$\alpha=0.05$ のとき信頼区間は $[2,12]$ となる．2 人以上 12 人以下の死者が予測されることになる．

後半については母分布は決まっている問題なので検定ではなく，分布が決まっているときある標本が起きる予測についての普通の確率計算である．$Q([14,\infty)) = 1-Q([0,13]) = 1-0.987 = 0.013$ だから 95% 以上の確からしさでそんなことは起こるまい，と言えるが，99% の確からしさを要求すると（100 回中 1.3

回は起きそうだから)「あり得ない」とは言えない.「本当に本当に異常なの(状況が悪化したの)?」と問われれば「たまたま多かった可能性を 1% 未満とは見積もれない」というところである[21] ◇

なお,起こる頻度の小さい現象は全てポワッソン分布かというと,それは間違いである!§10 で紹介するように,一定時間内に起こる現象数がポワッソン分布に従うとき,発生時間間隔は指数分布に従う.一方,特定地点で観測する大規模地震や装置・システムの末期故障のように「しばらくの間,事故原因となるエネルギーや疲労が蓄積していき,臨界点に達すると事象が発生する」という場合は,小さい時間間隔で起きる確率が指数分布に比べて非常に小さい.このような場合は筆者はブラウン運動の脱出時間の分布(§11)を使うほうが論理的だと考える.

7 パッケージ 1:1 つの正規母集団の推定・検定— χ^2 分布(分散)と t 分布(平均).

大標本理論(§5, §6)データの大きさが大きいとき,中心極限定理によって(主に)その標本平均が正規分布にほぼ従うことを利用する推定・検定であった.利点としては母分布によらない点があるが,欠点としてデータの大きさが大きくないと中心極限定理という根拠を失う点と,正規分布は平均と分散の 2 パラメータの分布族なので,分散を,やはりデータの大きさが大きいことから大数の法則を根拠として,不偏分散の標本値で近似する必要があった.

母集団が正規分布ならば小数のデータでも標本平均は正規分布に正確に従う(命題 13)ので前者の問題はないが,母分散の推定値を要する点で後者の問題は残り,データ数を大きくしないと使えない.

母集団が正規分布の場合に限定することで,2 パラメータ m, v の族として(一方を点推定で置き換えたりせずに)区間推定したり検定したりする方法が小標本理論である.この節 §7 で一つの分布についての推定・検定,次節 §8 で二つの分布の比較の検定を扱う.母集団が正規分布の場合が頻出する根拠は §6 の冒頭に説明した.

1 つの分散の推定・検定には χ^2 分布(§7.1),1 つの平均の推定・検定には t 分布(§7.3),がそれぞれ用いられる.いずれも正規分布から変数変換で得られる分布である.

7.1 χ^2 分布.

定義 25 標準正規分布 $N(0,1)$ に従う独立な確率変数 $X_1, X_2, \cdots, X_n$ の平方和 $\sum_{i=1}^{n} X_i^2$ が従う分布を,自由度 n のカイ平方分布(χ^2 分布)といい,${\chi^2}_n$ と書く *[小針, §8.3, §7.3–4], [林周二, 15.1], [楠岡, p.212].* ◇

定理 26 自由度 n の χ^2 分布は平均 $m = n$,分散 $v = 2n$ の連続分布で密度関数は

$$f_n(z) = \begin{cases} \dfrac{1}{2}\dfrac{\left(\frac{z}{2}\right)^{n/2-1} e^{-z/2}}{\Gamma(\frac{n}{2})}, & z \geqq 0, \\ 0, & z < 0. \end{cases} \tag{61}$$

で与えられる(図 *2*).ここで

[21] それにしても,**平常**の 2 倍の交通事故死者が出た**から**異常なのだろうか?そもそも毎日交通事故死者が出ている**現実が既に異常**ではないのか?筆者は幼少期に弟を交通事故でなくしている.交通事故死者が統計の数字になるほど平時から交通事故死者がいることそのことが異常だと思う心を忘れてはならない.

まして,雇用者保護などのどんな大義名分があろうとも,リコール隠しで交通事故を引き起こす自動車会社はあってはならない.即刻,経済社会から退場してやり直すべきである.技術は普遍性を持つから,他の自動車会社がシェアを奪うと同時に技術者を雇用できるはずなのだ.そうでなければ科学技術に対する信頼が失われる.そのほうが合計の社会損失ははるかに大きい.

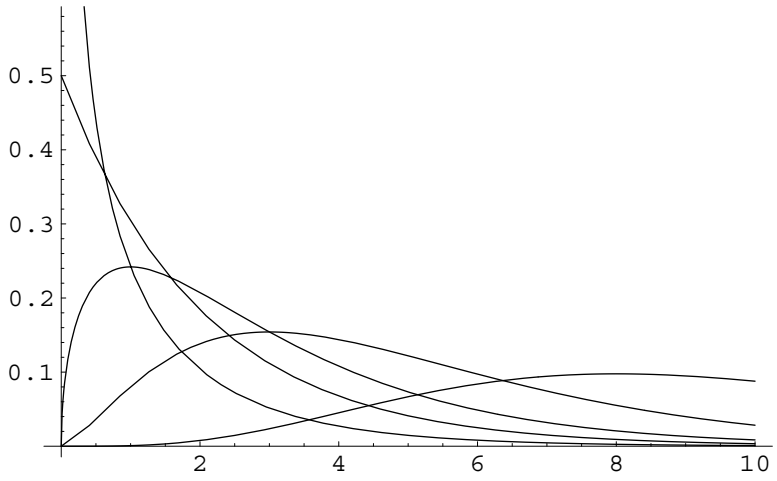

図 2: 自由度 $n = 1, 2, 3, 5, 10$ のカイ平方分布の密度関数．0 寄りが大きいグラフが小さい n に対応．$n = 8$ 近くに最大値のあるのが $n = 10$.

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx, \quad s > 0, \tag{62}$$

とおいた（ガンマ関数と呼ばれる）. ◇

証明． 定義から X_i たちが $N(0,1)$ に従うことを用いると

$$m = \mathrm{E}[\ \sum_{i=1}^n X_i^2\ ] = \sum_{i=1}^n \mathrm{E}[\ X_i^2\ ] = n\mathrm{V}[\ X_1\ ] = n.$$

さらに X_i たちが独立なことも使うと

$$v = \mathrm{V}[\ \sum_{i=1}^n X_i^2\ ] = \sum_{i=1}^n \mathrm{V}[\ X_i^2\ ] = n\mathrm{V}[\ X_1^2\ ] = n(\mathrm{E}[\ X_1^4\ ] - \mathrm{E}[\ X_1^2\ ]^2) = n(\mathrm{E}[\ X_1^4\ ] - 1) = 2n.$$

ここで $\mathrm{E}[\ X_1^4\ ] = 3$ の証明は次のように行う．ガウス積分 (55) で $x = y\sqrt{a}$ と変数変換すると

$$\int_{-\infty}^\infty e^{-ay^2/2}\, dy = \sqrt{\frac{2\pi}{a}}\,.$$

a で 2 回微分して $a = 1$ とおくと $\dfrac{1}{4}\displaystyle\int_{-\infty}^\infty y^4 e^{-y^2/2}\, dy = \frac{3}{4}\sqrt{2\pi}$. 右辺は $\dfrac{1}{4}\mathrm{E}[\ X_1^4\ ] \times \sqrt{2\pi}$ に等しい.

あとは密度関数 f_n を求めればよい．2 乗の和は負にならないから $z < 0$ では $f_n(z) = 0$ なので $z \geqq 0$ の場合だけ考えればよい.

$N(0,1)$ の密度関数は (54) から $\dfrac{1}{2\pi} e^{-x^2/2}$ であり，命題 5 から独立確率変数列の結合分布の密度関数は個々の確率変数の分布の密度関数の（変数を別々にした）積だから，χ^2 分布の定義から，

$$\int_0^a f_n(z)\, dz = \mathrm{P}[\ \sum_{i=1}^n X_i^2 \leqq a\ ] = \int_{\sum_{i=1}^n x_i^2 \leqq a} e^{-\frac{1}{2}\sum_{i=1}^n x_i^2} \frac{d^n x}{\sqrt{2\pi}^n}\,.$$

$F_n(z) = \displaystyle\int_{0 \leqq \sum_{i=1}^n x_i^2 \leqq z} \frac{d^n x}{\sqrt{2\pi}^n}$ とおくと，部分積分と積分順序の変更によって，

$$\int_0^a e^{-z/2} F_n'(z)\, dz = e^{-a/2} F_n(a) + \frac{1}{2}\int_0^a e^{-z/2} F_n(z)\, dz = \int_{\sum_{i=1}^n x_i^2 \leqq a} e^{-\frac{1}{2}\sum_{i=1}^n x_i^2} \frac{d^n x}{\sqrt{2\pi}^n} = \int_0^a f_n(z)\, dz.$$

これが任意の $a \geqq 0$ に対して成り立つから $f_n(z) = e^{-z/2} F_n'(z)$, $z \geqq 0$. 変数変換 $x_i = \sqrt{z} x_i'$, $i = 1, \cdots, n$, を考えると $F_n(z) = z^{n/2} F_n(1)$ を得るから，$f_n(z) = C_n\, e^{-z/2} z^{n/2-1}$, $z \geqq 0$. ここで $C_n = \dfrac{n}{2} F_n(1)$ は z に

よらない．C_n を $F_n(1)$ を直接計算して求めてもよいが，f_n は確率の密度なので $z \geqq 0$ 全範囲で積分すると 1 になることを使ったほうが速い．変数変換 $z = 2x$ によって

$$1 = \int_0^\infty f_n(z)\,dz = C_n \int_0^\infty e^{-z/2} z^{n/2-1}\,dz = 2^{n/2} C_n \Gamma(n/2).$$

これから (61) を得る． □

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

$n \setminus \alpha$	0.99	0.975	0.95	0.05	0.025	0.01	0.005
1	0.00	0.00	0.00	3.84	5.02	6.63	7.88
2	0.02	0.05	0.10	5.99	7.38	9.21	10.60
3	0.12	0.22	0.35	7.81	9.35	11.34	12.84
4	0.30	0.48	0.71	9.49	11.14	13.28	14.86
5	0.55	0.83	1.15	11.07	12.83	15.09	16.75
6	0.87	1.24	1.64	12.59	14.45	16.81	18.55
7	1.24	1.69	2.17	14.07	16.01	18.48	20.28
8	1.65	2.18	2.73	15.51	17.53	20.09	21.96
9	2.09	2.70	3.33	16.92	19.02	21.67	23.59
10	2.56	3.25	3.94	18.31	20.48	23.21	25.19
12	3.57	4.40	5.23	21.03	23.34	26.22	28.30
14	4.66	5.63	6.57	23.68	26.12	29.14	31.32
16	5.81	6.91	7.96	26.30	28.85	32.00	34.27
18	7.01	8.23	9.39	28.87	31.53	34.81	37.16
20	8.26	9.59	10.85	31.41	34.17	37.57	40.00
50	29.7	32.36	34.76	67.50	71.42	76.15	79.49
100	70.1	74.22	77.93	124.34	129.56	135.81	140.17

7.2 分散の推定

[林周二, 15.2], [小針, §8.3].

注 27 小標本理論（§7，§8）に出てくる全ての分布（χ^2 分布，t 分布，F 分布）に共通の注意だが，これらの分布を用いる際は母集団が正規分布であること（正規分布が良い近似であること）が必要である．データの大きさ n を大きくしてもカイ平方分布への中心極限定理はない． ◇

不偏分散の分布 正規母集団から無作為抽出したデータ列 $X_1, \cdots, X_n$ に対して標本平均は $\overline{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$，不偏分散は $V_n = \dfrac{1}{n-1}\displaystyle\sum_{j=1}^n (X_j - \overline{X}_n)^2$ であった．

定理 28 $\dfrac{n-1}{v} V_n$ の従う分布は自由度 $n-1$ の χ^2 分布である． ◇

定理 28 の証明はこの節 §7 の節末の補足に回す.

χ^2 分布の自由度がデータ数 n に比べて見かけ上 1 少ない．これは平均を $\overline{X}$ で推定したために，自由度が 1 減ったと理解すると覚えやすい.

定理 28 で扱う確率変数は母平均 m を含まないので，母平均を知らなくても母分散 v の推定・検定ができる.

検定の場合は母分布を χ^2 分布として検定の原理（§5.2）を適用すればよい．例えば古い工程と新しい工程で製品の品質のばらつき具合が変わらないという帰無仮説を立てて，棄却することにより品質のばらつきが減ったことを統計学的に立証する場合が考えられる．古い工程のデータが多くてデータから求めた不偏分散を古い工程の母分散と思うことができれば，定理 28 から帰無仮説の下で新しい工程による製品の不偏分散 V_n に対して $\dfrac{n-1}{v}V_n$ が自由度 $n-1$ の χ^2 分布に従うはずなので検定ができる．χ^2 分布の数表の使い方等は推定でも検定でも同じなので，推定の例のみ挙げておく.

例 29 *1996* 年 *1* 月 *17* 日に近所のスーパー「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）は茶色が 61, 62, 64, 64, 68，白が 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 v を信頼係数 90% で求めよ.

解. 計算すると，標本平均は 63.7，不偏分散 $V_n(w)$ については $(n-1)V_n(w) = 2.7^2 + 1.7^2 + 0.3^2 \times 3 + 4.3^2 + 5.7^2 + 0.7^2 + 2.3^2 + 3.3^2 = 78.1$ だから，$\dfrac{78.1}{v}$ が自由度 $n = 10 - 1 = 9$ の χ^2 分布に従う.

両側に *5%* ずつの危険域をとることにして $\alpha = 0.95$ と $\alpha = 0.05$ に対応する自由度 9 の χ^2 分布の区間の端点の値は §7.1 の表より，それぞれ 3.33, 16.92 だから，信頼区間は $3.33 < \dfrac{78.1}{v} < 16.92$．即ち，$4.62 < v < 23.45$, すなわち，母分布の標準偏差 $\sigma = \sqrt{v}$ の 90% 信頼係数での信頼区間は $2.15 < \sigma < 4.84$[22]. ◇

標本分散の分布. 定義から明らかに，一般の正規分布 $N(\mu, v)$ に従う独立同分布確率変数列 $X_1, X_2, \cdots, X_n$ に対して，命題 13 から $\dfrac{X_i - \mu}{\sqrt{v}}$ は標準正規分布 $N(0,1)$ に従うので，$\dfrac{1}{v}\displaystyle\sum_{i=1}^{n}(X_i - \mu)^2$ は自由度 n の χ^2 分布に従う．特に,

正規母集団 $N(\mu, v)$ から無作為抽出で得られた大きさ n のデータ $X_1, \cdots, X_n$ に対して

$$\frac{1}{v}\sum_{i=1}^{n}(X_i - \mu)^2 \tag{63}$$

は自由度 n の χ^2 分布に従う.

このことから，母平均 μ が既知のとき，たとえば，μ の値に関して帰無仮説を置いて検定するとき，は不偏分散ではなく標本分散 (63) を χ^2 検定できる．検定のしかたは不偏分散に対する検定と同様なので省略する.

7.3 t 分布

定義 30 Z が $N(0,1)$ に従い，Y が自由度 n の χ^2 分布に従う確率変数で，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 *[*林周二*, 16,17], [*小針*,* §8.4, §7.5–6*].* ◇

[22]結果を見ると，殆どデータのばらつきの範囲をなぞっているだけで，おもしろくないが...

Z が $N(0,1)$ に従うならば，Z^2 は自由度 1 の χ^2 分布に従うので，T が T_n に従うならば，T^2 は後述の F 分布 F_n^1 に従う（§8.1）．よって， T は F_n^m の特別な場合（の平方根）として得られる．

ここでは §8.1 の結果を先取りして利用する．

補題 31 Z の分布の密度関数 ρ_Z が左右対称 *(*すなわち $\mathrm{P}[\, Z > z\,] = \mathrm{P}[\, Z < -z\,]$*)* で，$X = Z^2$ の分布の密度を ρ_X とするとき，ρ_Z は

$$\rho_Z(t) = \rho_X(t^2)|t|, \quad t \in \mathbb{R},$$

で与えられる． ◇

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

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

最後の変形で変数変換 $x = t^2$ を行った． □

補題 31 と §8.1 の (66) から T_n の密度 f_n は

$$f_n(t) = \frac{1}{\sqrt{n}B(\frac{1}{2},\frac{n}{2})}\left(\frac{t^2}{n}+1\right)^{-(1+n)/2} \tag{64}$$

と求まる．B はベータ関数．詳しくは §8.1 を参照．

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 §5.1 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
19	1.729	2.093	2.5395	2.861
20	1.725	2.086	2.528	2.845
30	1.697	2.042	2.457	2.750
40	1.684	2.021	2.423	2.704
∞	1.6448	1.9600	2.326	2.5758

7.4 平均値の推定．

命題 32 正規分布に従う独立同分布確率変数列 $X_1, \cdots, X_n$ について，標本平均 $\overline{X}_n$ と不偏分散 V_n は独立な確率変数である． ◇

命題 32 の証明は (41), (49), (50) から直接機械的に証明できる．証明はこの節 §7 の節末の補足に回す．

定理 33 正規母集団 $N(\mu, v)$ から大きさ n のデータを選び，その標本平均 *(49)* を $\overline{X}_n$，不偏分散を *(50)* V_n，とすると，

$$T = \sqrt{\frac{n}{V_n}}\,(\overline{X}_n - \mu) \tag{65}$$

は自由度 $n-1$ の t 分布 T_{n-1} に従う． ◇

証明． 命題 13 から $\dfrac{\sqrt{n}}{\sqrt{v}}\,(\overline{X}_n - \mu)$ は $N(0,1)$ に従い，定理 28 から $\dfrac{(n-1)V_n}{v}$ は自由度 $n-1$ の χ^2 分布に従う．さらに命題 32 から V_n と $\overline{X}_n$ は独立である[23]．よって T_n の定義から主張が従う． □

定理 33 の T は母分散 v を含まないので，母分散を知らなくても母平均 m の推定・検定ができる．その手続きは χ^2 分布による分散 v の推定・検定と同様である．母分散を既知とするか不偏分散等で根拠無く近似する大標本理論（§6）と比較せよ．

例 34 §7.1 で母分散の推定を行った例を取り上げる．例 *29* の例で茶色の卵と白の卵に区別がないとして，この 10 個の卵の母平均 μ を信頼係数 95% で求めよ．

解． 例 *29* で見たように，データの大きさ $n = 10$，標本平均 $\overline{X} = 63.7$，不偏分散 V_n は $(n-1)V_n = 78.1$ である．危険率は $\alpha = 0.05$．定理 *33* から $T = \sqrt{\dfrac{n}{V_n}}\,(\overline{X} - \mu)$ は T_{n-1} に従う．t 分布の表より，$0.05 = \mathrm{P}[\,|T| > 2.262\,]$．即ち，信頼係数 95% の信頼区間は $\sqrt{\dfrac{90}{78.1}}|63.7 - \mu| < 2.262$，すなわち，$\mu = 63.7 \pm 2.1$ *(95% CL*[24]*)*．§7.1 の標準偏差の推定 $2.15 < \sigma = \sqrt{v} < 4.84$ *(90% CL)* と合わせて，小標本理論における正規母分布の推定の例が完成した． ◇

7.5 例題．

平成 14 年人口動態調査 [人口] によると，1 年間に日本で生まれた女児 100 に対する男児の数（性比）は 2002 年までの 20 年間次の一番左の表のようになっていた（残りの表は §8.5 で使う）．毎年のばらつきが独立で分散未知の正規母集団の標本であると仮定して出生性比の平均を区間推定せよ．

[23] 正規母集団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, p.236] では F に関しては言及している．

[24] confidence level，信頼係数

西暦	性比
1983	105.7
1984	105.4
1985	105.6
1986	105.9
1987	105.8
1988	105.6
1989	105.6
1990	105.4
1991	105.7
1992	106.0
1993	105.6
1994	105.6
1995	105.2
1996	105.6
1997	105.2
1998	105.4
1999	105.6
2000	105.8
2001	105.5
2002	105.7

西暦	性比
1961	105.9
1962	106.1
1963	105.7
1964	105.9
1965	105.3
1966	107.6
1967	105.3
1968	107.1
1969	107.2
1970	107.1
1971	106.7
1972	106.5
1973	106.2
1974	106.4
1975	106.2
1976	106.2
1977	106.1
1978	106.0
1979	106.2
1980	106.0

西暦	性比
1910	103.9
1911	104.0
1912	104.1
1913	104.4
1914	104.9
1915	104.2
1916	104.3
1917	104.2
1918	104.3
1919	104.9
1920	104.5
1921	104.5
1922	104.0
1923	104.4
1924	104.2
1925	103.5
1926	105.8
1927	103.7
1928	104.4
1929	104.0

解. 表から，データの大きさ $n=20$，標本平均 $\overline{X}_n=105.595$，不偏分散 $V_n=0.0426053$．定理 33 から μ を母平均とすると $T=\sqrt{\dfrac{n}{V_n}}\,(\overline{X}_n-\mu)=21.666\times(105.595-\mu)$ は自由度 $n-1=19$ の t 分布 T_{19} に従う．§7.3 の表から $T_{19}([-a,a])=1-\alpha$ となる a は，$\alpha=0.05$ のとき $a=2.093$，$\alpha=0.01$ のとき $a=2.861$．

以上より，$\mu=105.595\pm0.097$ (95% CL)，$\mu=105.595\pm0.132$ (99% CL).

補足：定理 28 の証明.

定理 28 の証明. $W_j=\dfrac{X_j-\mu}{\sqrt{v}}$ および $\overline{W}_n=\dfrac{1}{n}\sum_{j=1}^{n}W_j$ とおくと，W_j の分布は $N(0,1)$ で $j=1,\cdots,n$ は独立．

$w^T S\,w=\sum_{j=1}^{n}(w_j-\overline{w}_n)^2$ と書くと，

$$\mathrm{P}[\,\frac{n-1}{v}V_n\geqq a\,]=\mathrm{P}[\,\sum_{j=1}^{n}(W_j-\overline{W}_n)^2\geqq a\,]=\int_{w^T S\,w\geqq a}\exp(-\frac{1}{2}\sum_{j=1}^{n}w_j^2)\,\frac{d^n w}{\sqrt{2\pi}^n}$$
$$=\int_{v^T D\,v\geqq a}\exp(-\frac{1}{2}v^T\,v)\,\frac{d^n v}{\sqrt{2\pi}^n}\,.$$

ここで，$S_{ij}=S_{ji}=\delta_{ij}-1/n$ で定義される $n\times n$ 行列 S を対角化する直交行列を $O:OSO^T=D$ および $OO^T=1$，とし，対角化された結果を D，および，$v=O\,w$ と置いた．右肩添字 T は転置を表す．

S の固有値は $\sum_j S_{ij}O_{kj}=d_k O_{ki}$ を解くと，$d_k=0$，$O_{ki}=O_k$，または $d_k=1$，$\sum_j O_{kj}=0$．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，

$$\mathrm{P}[\,\frac{n-1}{v}V_n\geqq a\,]=\int_{\sum_{i=1}^{n-1}v_i^2\geqq a}\exp(-\frac{1}{2}\sum_{i=1}^{n-1}v_i^2)\,\frac{d^{n-1}v}{\sqrt{2\pi}^{n-1}}\,.$$

よって，$\frac{n-1}{v}V_n$ の従う分布は自由度 $n-1$ の χ^2 分布．□

補足：命題 32 の証明．

命題 32 の証明． $X_1, \cdots, X_n$ の結合分布の密度は

$$\mathrm{P}[\,(X_1, \cdots, X_n) \in A\,] = \int_A \exp(-\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2v})\, d^n x =: \int_A \rho(\{x_i\})\, d^n x\,.$$

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

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

よって，

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

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

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

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

8 パッケージ 2：正規母集団の比較の検定— F 分布（等分散の検定）と t 分布（等平均の検定）．

前節 §7 では 1 つの正規母集団のパラメータ（母平均と母分散）の推定・検定を χ^2 分布と t 分布を用いて行った．この節では 2 つの正規母集団のパラメータが等しいかどうかの検定を F 分布（等分散の検定，平均任意）と t 分布（等平均の検定，等分散または大標本理論）を用いて行う．

8.1 F 分布．

定義 35 X が自由度 m の χ^2 分布に従い，Y が自由度 n の χ^2 分布に従う確率変数で，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_n^m と書く *[林周二, 16,17], [小針, §8.4, §7.5–6].* ◇

補題 36 ([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

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

で与えられる． ◇

証明.

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

途中で変数変換 $ax = byt$ を行った. □

定理 37 F_n^m の密度 g_n^m は

$$g_n^m(z) = \chi(z \geqq 0)\, \frac{m^{m/2} n^{n/2}}{B(m/2, n/2)}\, \frac{z^{m/2-1}}{(mz+n)^{(m+n)/2}} \tag{66}$$

である．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．さらに

$$F_n^m((0,a]) = F_m^n([1/a, \infty)) \tag{67}$$

および

$$T_n([-a,a]) = F_n^1([0,a^2]) \tag{68}$$

が成り立つ．後者は T が T_n に従うとき，T^2 は F_{n-1}^1 に従うという意味である. ◇

証明. 補題 36 において，$a = 1/m$, $b = 1/n$, また，p, q に (61) を代入．整理して，(62) および公式

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

を用いると (66) を得る.

残りは容易. □

F が F_n^m に従うとき，$\alpha = \mathrm{P}[\, F > f \,] = \int_f^\infty g_n^m(x)\, dx$ 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_n 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
∞	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
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64

8.2 等分散の検定.

定理 38 2 つの正規母集団 $N(\mu, v)$, $N(\mu', v')$ からそれぞれ，大きさ m, n のデータを選び，不偏分散 *(50)* をそれぞれ V_m, V'_n とすると，

$$\frac{v'}{v}\frac{V_m}{V_n} \tag{69}$$

の分布は F_{n-1}^{m-1} である.
特に，分散が等しい 2 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. ◇

証明. §7.1 から，$\frac{m-1}{v}V_m$, $\frac{n-1}{v'}V'_n$ の従う分布はそれぞれ自由度 $m-1$ および $n-1$ の χ^2 分布である．よって F 分布の定義から結論を得る. □

定理 38 から，特に母平均を知らない 2 つの正規母集団から得られた標本の組について，母分散が両者で等しいという帰無仮説を検定できる．2 組の正規母集団のパラメータについて何の知識も必要ない！

定理 38 に基づく F 分布による検定をF **検定**と俗称する．一般に，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がこの範疇に入る.（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．数学的には区別する理由はほとんどない.）

例 39 §7.1 の例 *29* と §7.4 の例 *34* で母分散の推定と母平均の推定を行った例を再度取り上げる．これまで茶色と白色の卵を区別しなかったが，今度は茶色の卵と白色の卵の分散の違いを信頼係数 90% で検定する.
解. 茶色と白の分布をそれぞれ，$N(\mu, v)$, $N(\mu', v')$ として帰無仮説 *H:* $v = v'$ をおく.

茶色と白色の卵のデータ数と標本平均と不偏分散はそれぞれ $m = n = 5$, $\overline{X} = 63.8$, $V = \frac{1}{4}(2.8^2 + 1.8^2 + 0.2^2 * 2 + 4.2^2) = 28.8/4 = 7.2$, $\overline{X}' = 63.6$, $V' = \frac{1}{4}(5.6^2 + 0.6^2 + 0.4^2 + 2.4^2 + 3.4^2) = 49.2/4 = 12.3$. 特に分散比は $V/V' = 0.59$. 90% での検定で，両側区間検定として $\alpha = 5\%, 95\%$.

定理 *38* から，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， *(67)* より，$\mathrm{P}[\, F_n^m > x\,] = 1 - \mathrm{P}[\, F_n^m < x\,] = 1 - \mathrm{P}[\, F_m^n > 1/x\,]$ 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 $0.156 < V/V' < 6.39$. データは $V/V' = 0.59$ だか

ら危険率 0.1 で H は棄却されない．すなわち，茶色と白の卵の間で母分散は等しいとして統計的に異常な判断ではない． ◇

8.3 等平均の検定．

§8.2 の F 検定で帰無仮説が採択された場合等のように，母分散が等しい 2 つの正規母集団から無作為抽出で得られたデータについて母平均が等しいかどうかは t 分布で検定できる [林周二, 16.3]．母分散が等しい正規分布に従う確率変数の差は母平均の差を平均とする正規分布に従うからそれが 0 かどうかを検定することになり，1 つの正規母集団の母平均の検定の問題に帰着するから，§7.4 によって t 検定の問題にすぎないことが分かる．

定理 40 分散の等しい二つの正規母集団 $N(\mu, v)$, $N(\mu', v')$ からそれぞれ大きさ m, n のデータを無作為抽出して，その標本平均 *(49)* と不偏分散 *(50)* をそれぞれ $(\overline{X}_m, V_m)$, $(\overline{X'}_n, V'_n)$ とすると，

$$T = \frac{\overline{X} - \overline{X}' - \mu + \mu'}{\sqrt{(m-1)V_m + (n-1)V'_n}} \sqrt{\frac{mn(m+n-2)}{m+n}} \tag{70}$$

は自由度 $m+n-2$ の t 分布 T_{m+n-2} に従う． ◇

証明． 命題 13 から $\dfrac{\overline{X} - \overline{X}' - \mu + \mu'}{\sigma\sqrt{\frac{m+n}{mn}}}$ は $N(0,1)$ に従い，§7.1 から $\sigma^{-2}(m-1)V_m$ と $\sigma^{-2}(n-1)V'_n$ はそれぞれ自由度 $m-1$ と $n-1$ の χ^2 分布に従うので，両者の和は自由度 $m+n-2$ の χ^2 分布に従う． □

例 41 §7.1 の例 *29* 以来の例を再度取り上げる．§8.2 の例 *39* から分散は茶色の卵と白色の卵で等しいと仮定して，帰無仮説

H: $\mu = \mu'$

を信頼水準 95% で検定する．

解． $T = \dfrac{\overline{X} - \overline{X}' - \mu + \mu'}{\sqrt{(m-1)V_m + (n-1)V'_n}} \sqrt{\dfrac{mn(m+n-2)}{m+n}}$ は定理 *40* から T_{m+n-2} に従う．

既に見たように，$\overline{X} = 63.8$, $\overline{X}' = 63.6$, $V = 7.2$, $V' = 12.3$, $m = n = 5$ なので，$T = 0.101$．t 分布の表より，$0.05 = \mathrm{P}[\,|T_8| > 2.306\,]$．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，白い卵と茶色の卵で重さの分布に差があるというデータ上の根拠はない． ◇

8.4 例題．

下表は仙台，東京，名古屋の 2003 年と 2002 年の 7 月 26 日から 2 週間の毎日の最高気温である．2003 年のこの期間について，最高気温の分布が都市間で平均は異なるかもしれないが分散は等しい正規分布に従っていると仮定して，各都市間の気温差を推定せよ．「名古屋は東京より暑く，仙台は東京より涼しい」と言ってよいか？2002 年についてはどうか？

2003 年	仙台	東京	名古屋
7 月 26 日	17.8	26.7	29.8
7 月 27 日	22.4	27.3	29.5
7 月 28 日	19.3	25.7	26.4
7 月 29 日	22.4	25.4	27.2
7 月 30 日	22.5	27.1	27.0
7 月 31 日	26.0	30.9	31.8
8 月 1 日	24.3	28.4	31.8
8 月 2 日	26.6	31.4	32.9
8 月 3 日	29.4	32.0	32.2
8 月 4 日	29.7	33.4	34.6
8 月 5 日	27.9	33.3	35.0
8 月 6 日	24.9	30.9	34.1
8 月 7 日	24.9	31.3	32.5
8 月 8 日	26.3	31.0	32.0
データサイズ	14	14	14
標本平均	24.6	29.63	31.2
不偏分散	12.11	7.66	7.91

2002 年	仙台	東京	名古屋
7 月 26 日	32.3	33.9	33.8
7 月 27 日	30.6	34.5	34.9
7 月 28 日	26.4	29.6	33.5
7 月 29 日	26.9	30.9	32.9
7 月 30 日	32.7	33.2	35.5
7 月 31 日	32.1	34.8	36.8
8 月 1 日	36.1	35.6	37.1
8 月 2 日	31.7	35.2	35.0
8 月 3 日	23.5	32.3	36.1
8 月 4 日	26.1	31.4	35.9
8 月 5 日	30.3	34.1	36.4
8 月 6 日	35.2	35.7	38.2
8 月 7 日	32.6	34.3	37.6
8 月 8 日	33.6	35.7	35.7
データサイズ	14	14	14
標本平均	30.72	33.66	35.67
不偏分散	13.62	3.72	2.40

略解と解説.

2003 年（仙台－東京）の平均気温差の推定値は，-5.0 ± 2.0 (90% CL), -5.0 ± 3.3 (99% CL),（東京－名古屋）は -1.6 ± 1.8 (90% CL), -1.6 ± 2.9 (99% CL)．仙台と東京の差ははっきりしているが，東京と名古屋には等温の帰無仮説を棄却できる差はない.

どの欄も $\sqrt{V} \sim 3$ 程度のばらつきのあるデータなので，一見 5 度という標本平均の差もばらつきのうちに見えるが，$n = 14$ あってそれが正規分布と思っている（仮定により）ので，そこから母平均の差を再構成する相対精度は $1/\sqrt{14/2} \sim 2$ 程度上がって，5 度の差は意味を持つ.

一方 2002 年（仙台－東京）の平均気温差の推定値は，-2.9 ± 1.9 (90% CL), -2.9 ± 3.1 (99% CL),（東京－名古屋）は -2.0 ± 1.1 (90% CL), -2.0 ± 1.8 (99% CL)．差はあるが 99% までは言えない.

しかし，問題文や，名古屋は暑い，仙台は寒い，という常識に引きずられてもっと重要なことを見落としてはならない，と思う．表の中で顕著な差は地域差ではなく年の差である．仙台の（2003 年－ 2002 年）の平均気温差の推定値は，-6.1 ± 2.3 (90% CL), -6.1 ± 3.8 (99% CL)，で各年度の都市間の差より顕著である！2002 年と 2003 年のどちらに異常があったのか（あるいは両方か）は平年値（長い年月にわたる平均として気象庁が発表している）と比べねばならない（2003 年の夏が異常低温だった）が，異常な年があったことが表から最初に読み取るべき内容であろう.（自分で経験していなければ，ひょっとしたら表に誤りがあるのかもしれないと疑うくらいの大きな差である.）

最も問われるのは計算技術ではなく**データから何を読み取るか**である．そこを常識や政府公式見解や既存の権威に引きずられていては統計学に何の意味もない．さらに言えば，**どのようなデータを集めるか**が読み取るべき内容と両輪なのは言うまでもない.

8.5 例題 2.

§7.5 の出生比の表の右 2 つ，大正時代と昭和の高度成長期の 20 年間の出生比を比べよ．ただし，それぞれの期間で毎年のばらつきが独立で分散未知の正規母集団の標本であると仮定する.

略解. 性比の高度成長期の値と大正時代の差の推定値は $(2.0 \pm 0.3)\%$ (90% CL)，$(2.0 \pm 0.5)\%$ (99% CL).

おおむね 2% 程度高度成長期のほうが男児過剰である．生物としてのヒトの平均属性が短期間に変動するとは思いにくいから，社会的要因によるものであろう．社会情勢がホルモンバランスなどに影響して母体の内部で性比に影響を与えたのか，間引き（出生後の性別による殺人の差）によるものかはこの統計だけでは分からない．最近は出生前に性別判断ができるので，流産による人為的性比制御が容易になっている.

ただ，人為的制御は行き過ぎやすいので（特に 20 年以上経たないとどちらの性が有利か分からないのだから）内的なものか人為的なものかの判断が必要ではある.（この時点でどうすればいいのか筆者には分からない．性比と一見無関係そうな胎児死亡率の国際比較などから間接的に嬰児殺しを判断するしかない.）

9 メンデルの法則に隠された真実 — 統計学その他の話題.

9.1 その他の話題.

少し経路の変わった興味深い例をまとめておく.

9.1.1 メンデルの実験と 100 年後のフィッシャーの発見.

メンデルは 1865 年，エンドウ豆の形質遺伝を調べて次の結果を得た [渡辺浩].

表現型	丸い黄色	しわがある黄色	丸い緑色	しわがある緑色	合計
観測度数	315	101	108	32	556
理論比	9	3	3	1	16
理論度数	312.75	104.25	104.25	34.75	556

このデータは「メンデルの法則」の正しさを証明しているか？

適合度の検定 §9.2.1．観測値は $n = 556$ 個の独立同分布確率変数の標本値で，分布は $\mathrm{P}[\, X = \text{丸い黄色}\,] = \dfrac{9}{16}$ などと読む.

$\chi^2 = 0.470$，自由度 $A = 3$ で定理 42 を適用．99% CL で $\chi^2 \leqq 11.34$（§7.1）だから，余裕で適合している．χ^2 検定では通常大きいほうを危険域にとる片側検定（§9.2.1）．よく合っている小さい χ^2 を棄却するのは統計学の常識に反する.

しかし，硬貨を 8 回投げて正確に 4 回表が出るより，理論度数からばらつくのが普通．χ^2 はおおむね自由度（小さいときはやや小さい値）に近い値になるのが自然．$n = 3$ の χ^2 のグラフを見ると（§7.1）小さい χ^2 は少ない．χ^2 分布に対する「下側検定」$\mathrm{P}[\, \chi^2 < 0.58\,] = 0.1$ 90% CL の下側検定で小さすぎる.

何年かデータを積み重ねて，その中でいちばん自説（帰無仮説 H そのもの）を主張するのに都合のよい年のデータだけを発表した，という「物語」が考えられる（小標本理論を完成させた統計学の大家フィッシャーの考察）.

9.1.2 分布のばらつきと標本のばらつき.

実験データが少ないために標本平均がばらつくということと母分布自体が分散 0 でないということは区別を要する．[渡辺浩] による次の問題，特に巧妙に作られた選択肢は勉強になる.

児童生徒の健康状態サーベイランス事業報告書（1998 年）によると学童の睡眠時間について次の調査結果が得られた.

学年	（人数）	平均値	標準偏差
小学 3, 4 年生男子	(350)	9 時間 03 分	0 時間 40 分
小学 3, 4 年生女子	(322)	9 時間 01 分	0 時間 40 分
小学 5, 6 年生男子	(530)	8 時間 56 分	0 時間 42 分
小学 5, 6 年生女子	(481)	8 時間 45 分	0 時間 42 分
中学生男子	(992)	7 時間 34 分	1 時間 07 分
中学生女子	(941)	7 時間 10 分	1 時間 02 分
高校生男子	(1073)	6 時間 55 分	1 時間 21 分
高校生女子	(1985)	6 時間 42 分	1 時間 15 分

以下のうち正しいことを言っているのは誰だろうか.

A: 大まかには，睡眠時間は小学 3, 4 年生の男子が最も長く，高校生の女子が最も短いということだね.
B: 睡眠時間を聞いて，年齢と性別を当てることもできそうだ.
C: そうかな．たとえば高校生の男子と女子の睡眠時間の分布はほとんど重なっているよ.
D: つまり高校生の男女の睡眠時間には統計的な差がないということさ.

（母分散が異なる 2 つの正規分布の等平均の検定は Welch の検定などを用いるのが本来だが，ここでは母分散がほぼ等しいと見積もって t 検定で考えれば十分である.）

9.1.3 統計量の選択の重要性.

データから読み取るべきは何か，それに応じて適切な統計量の選択は異なる．たとえば [渡辺浩] に次の問題があった.

Hayes は，1943 年から 1958 年までの 16 年間に，ノースカロライナのある病院における記録をもとに，急性白血病の発症数を月別に集計して，次のような表を得た. この結果において，急性白血病の発症に季節変動が認められるか. 有意水準 5% で検定せよ.

月	1	2	3	4	5	6	7	8	9	10	11	12	計
頻度	23	21	15	20	14	8	11	11	14	17	10	20	184

全度数を N，m 月における度数を x_m $(m = 1, 2, \cdots, 12)$ とし，

$$C = \sum_{m=1}^{12} x_m \cos \frac{2\pi m}{12} \ , \ S = \sum_{m=1}^{12} x_m \sin \frac{2\pi m}{12}$$

とおいて，統計量

$$R = \frac{2}{N}(C^2 + S^2)$$

を考える. このとき,「季節変動が存在しない」という仮定のもとで，R は自由度 2 の χ^2 分布に従う. この方法を用いると，有意水準 1% で，Hayes のデータに季節変動が認められることを示せ.

9.1.4 統計調査.

政府はいろいろな統計を取っている.（通常発表されるのは平均値だけなので統計学の腕のふるいどころは少ない.）たとえば，人口動態調査 [人口] に，初婚のカップルの男女年齢差の変遷の表がある．筆者が 1976 年に受けた大学 1 年の統計学の講義の 11 月 10 日に，林周二先生 [林周二] が「初婚男女の結婚年齢の男女差は平均 3（男性が上）の正規分布によく合っている」と言われたことが私のノートに残っている．結婚年齢は年によって変動しても年齢差は変わらないという主張に聞こえた記憶がある.

しかしこれが正しかったのは 1970 年ころまでの 20 年弱だけである（図 3）. その後確実に年齢差の平均は 3 から同年齢の方向にずれていった．2000 年以降は 1.8 まで落ちている．1970 年から約 4 年減り続けていたので，講義した時点でもその傾向を真剣に見ればすばらしい予言者たり得たのに ! 何に注目するかがだいじだということを思わせるエピソードであった.

正規分布という観察も人口動態調査からある程度「検定」できる（図 4）. 1970 年頃はたしかに平均 3 の正規分布と言えるが，1980 年には早くも崩れ始め，1997 年には非常にとがった正規分布と似ても似つかぬ分布になっている.

定量的なことはきわめていいにくいが，このグラフを伝統的な「3 歳差結婚」と「同年齢結婚」の重ね合わせ（実際にはそれぞれ標準偏差が 3, 1 程度の分布）と考えておおざっぱに見積もると，1980 年は「同年齢（±1 歳）結婚」が 10%，1990 年は 20% 強，1997 年は 30% 強となる．学校の同級生や入社したときの同期の間での結婚がそういう比率で着実に増えてきたのが 1970 年代以降の傾向と仮定すれば，分布の変化の社会的背景の説明として説得力を感じる.

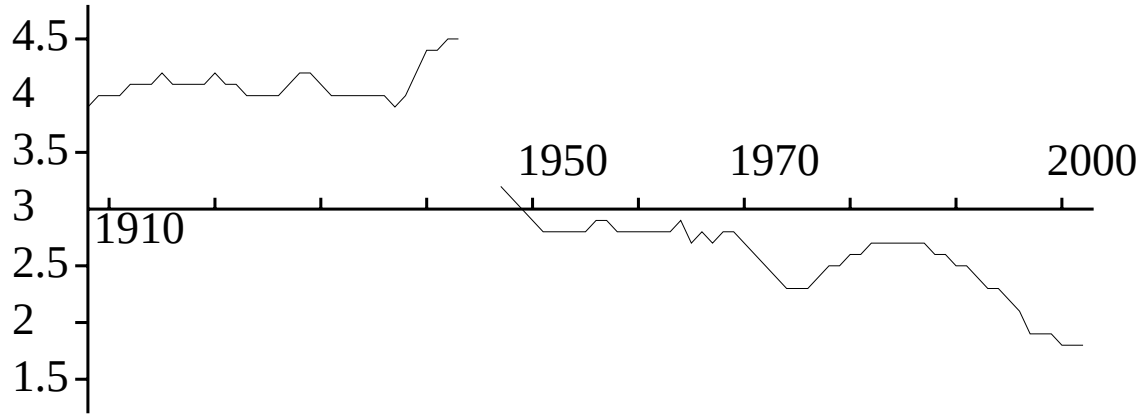

図 3: 初婚男女年齢差の年間平均．縦軸は年齢差で男性が年上を正にとっている．

9.2 χ^2 分布や F 分布を用いた種々の統計的推測．

9.2.1 適合度の検定．

母集団分布形の理論的仮定を帰無仮説として，母集団から得られたサイズ N のデータ $X_1, \cdots, X_N$，（母集団を分布とする独立同分布確率変数列）の標本 ω によって検定したい．

母集団が離散的な分布で，各変数値に十分データの度数があれば以下のようにカイ平方分布が適用できる [小針, §8.3, 例題 5]．母集団が密度 f を持つ連続分布のときもデータの標本値は有限 (N) 個の数値なので，その分布（経験分布の標本値）は密度を持たないが，母集団の密度の近似として，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．母集団が離散分布の場合でもデータ数が少ないときは，同様にいくつかの変数の値をまとめて bin を取り直す．

Bin の選び方[25]：$Np_a \geqq 4$ になるように bin 数 A を減らすことと $\inf_{a,a'} p_a/p_{a'} \approx 1$.

$a \in \{1, 2, \cdots, A\}$ 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 42 $X = \displaystyle\sum_{a=1}^{A} \frac{(Z_{N,a} - Np_a)^2}{Np_a}$ 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． ◇

[25] [林周二, 15.3] の見解とは必ずしも一致しない．

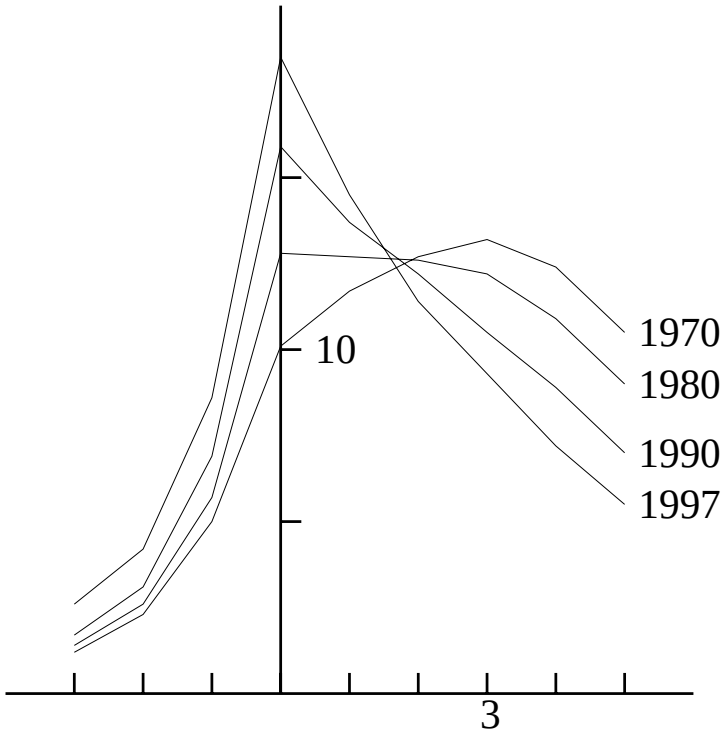

図 4: 初婚男女年齢差の分布．横軸は年齢差で男性が年上を正にとっている．縦軸は割合%

注 43 *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，

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

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 $\sum_{a=1}^{A} k_a = N$ は $\sum_{a=1}^{A} z_a\sqrt{p_a} = 0$ 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，$f = \log N - \sum_a k_a/N \log(k_a/p_a)$ とおくと ($\sum_a p_a = 1$ に注意),

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

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，

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

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 $\sum_a z_a\sqrt{p_a}=0$ を用いた．よって，

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

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

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

□

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 42 から $\chi^2=\sum_{a=1}^{A}\dfrac{(g_a-g_a^*)^2}{g_a^*}$ が自由度 $A-1$ のカイ平方分布に従う $X=\sum_{a=1}^{A}\dfrac{(Z_{N,a}-g_a^*)^2}{g_a^*}$ の信頼区間にはいるかどうかで，**適合度の検定**ができる [林周二, 15.3].

適合度の検定は通常 χ^2 の大きい側だけを棄却域とする片側検定を行う．t 検定などと違って，データの理論値（帰無仮説）からの 2 乗を検定の対象にしているので，理論値からのずれは大きいほうに出るから．0 の近くはよく合っていることになる．しかし，§9.1.1 も参照せよ．

注 44 *(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) 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]*.

(iii) 一般に実数空間でなくても Ω 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．

(iv) カイ平方の値が小さすぎたら *bin* のとりかたも一応確認する． ◇

9.2.2 回帰分析．

$p+1$ 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

$$Y=a_0+\sum_{j=1}^{p}a_jX_j$$

で近似（**回帰**）したい [林周二, 19.3], [渡辺浩]．言い換えると，$Z=(X_1,X_2,\cdots,X_p,Y):\ \Omega\to\mathbb{R}^{p+1}$ のデータ（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

$$E=\sum_{k=1}^{n}(Y_k-a_0-\sum_{j=1}^{p}a_jX_{kj})^2 \tag{71}$$

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．

$$\begin{aligned} a_j &= \sum_{k=1}^{p} C_{jk}^{-1} d_k \\ a_0 &= \overline{Y} - \sum_{k=1}^{p} a_k \overline{X}_k \end{aligned}$$

ただし共分散 $C(X,Y) = \mathrm{E}[\,(X-\mathrm{E}[\,X\,])(Y-\mathrm{E}[\,Y\,])\,]$,

$$\begin{aligned} C_{jk} &= C(X_j, X_k) = \sum_i (X_{ij} - \overline{X}_j)(X_{ik} - \overline{X}_k), \\ d_k &= C(Y, X_k) = \sum_i (Y_i - \overline{Y})(X_{ik} - \overline{X}_k). \end{aligned}$$

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，

$$Y \text{ の変動 } S_{\mathrm{total}} = \sum_{k=1}^{n} (Y_k - \overline{Y})^2$$

は，X の変動に起因する部分（**回帰変動**） $S_r = \sum_{k=1}^{n} (\underline{Y}_k - \overline{Y})^2$ 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，帰無仮説 H:「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 α_i, $i=0,1,2,\cdots,p$ の推測値であるが，

定理 45 (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_{n-p-1}^{p} に従う．$\diamond$

上では (71) において理論式からのずれの 2 乗の和を最小にするようパラメータを決めたので**最小 2 乗法**と呼ばれる．化学実験のようにデータが誤差を伴い，その誤差がデータ点によって異なる場合は偏差の 2 乗を誤差（分散）で割った量の和を最小にする（ §9.2.1 の統計量に対応）．χ^2-fit とも呼ばれる．

9.3 その他の方法．

9.3.1 最尤法．

古典的な点推定の方法の一つ [林周二, 14.2]．§4 の点推定と異なるのは，§4 は母分布（族）を仮定せず平均や分散などの一般性のある母数を推定したのに対して，最尤法は母分布の族を仮定してパラメータを点推定する．

母分布の族 P_θ の密度を f_θ とする：$P_\theta((-\infty, a]) = \displaystyle\int_{-\infty}^{a} f_\theta(x)\,dx$.

データサイズ n の標本 w，すなわち，母分布に従う独立同分布確率変数列 $X_1, \cdots, X_n$ の w での値に対して，$L(\theta) = \displaystyle\prod_{j=1}^{n} f_\theta(X_j(\omega))$ を**尤度関数**と呼ぶ．

最尤法による点推定の原理. L の最大値を与える θ を採る：$\dfrac{d\log L}{d\theta}=0$. （パラメータが複数ある時は当然偏微分になる.）

最尤法に関する一般的結果は [辞典, 287M] 参照.

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

例 46 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，μ, 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. ◇

統計的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．問題の立て方でも立場や考え方でも異なる.

10 バスはなぜ来ないか — ポワッソン確率過程.

10.1 ばらつきか間引きか？

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

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

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

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

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

即ち，

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

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

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

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

つまり，

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

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

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

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

10.2 ポワッソン分布．

今までの想定ケースは等間隔か集中的到着かという意味では極端な違いがあったが，どちらも「規則的に」到着した．現実のバスは不規則だから困るのであった．そこで極端なケースとしてバスが全く「でたらめ」に運行する場合に平均待ち時間がどうなるかを知りたい．このための数学的準備としてまずポワッソン分布を紹介する．

実数区間上に一様分布で針を何本か落とす.（$[n, n+1)$ 上に一様分布で落とす，ということを全ての n について独立に行う，という意味.）平均 λ 個落とすことと，異なる針の間では落としかたは独立ということが決まっているとする．そういう確率分布は存在するか？実はそれがポワッソン過程と呼ばれるもの．落とした本数の分布がポワッソン分布，複数本落としたときの針の間の間隔が指数分布になる．

$[0, 1)$ を n 等分，n が十分大きくて各 bin に落ちる本数は高々1 本と近似すると，bin 当たり落ちる確率を $p = \lambda/n$ で，bin 間は独立とすると，幅 $1/n$ を単位にした一様性と独立性，そして本数の平均値が条件に合っている．$n \to \infty$ とした極限が確率になっていれば，それが求めるものになるだろう．区間全体で落ちた針の本数は，bin を n 等分したときは $B_{n,p} = B_{n,\lambda/n}$ なので，k 本落ちる確率は (4) から

$$Q_n(\{k\}) = {}_n\mathrm{C}_k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k},\ k = 0, 1, \cdots, n-1, n.$$

λ と k を固定して n の極限を取ると，Stirling の公式 (14) などから

$$Q(\{k\}) = \lim_{n\to\infty} Q_n(\{k\}) = e^{-\lambda}\frac{\lambda^k}{k!}, k = 0, 1, 2, \cdots, \tag{72}$$

となる．これが $\Omega = \mathbb{Z}_+$ を全体集合とする確率測度（分布）を与えていることを見るのはやさしい.（確かめるべきは $Q(\Omega) = 1$ だけ.）これをポワッソン分布と呼ぶ.

細かいことを言うと §1.2 で離散分布を定義したとき，全体集合 Ω が有限集合の場合に限った．ポワッソン分布は $\Omega = \mathbb{Z}_+$ なので無限集合の場合に拡張しないといけない．拡張すべきは，無限個の確率を足すことの定義だけである．たとえば (6) のように全部足すと 1 という性質は確率が当然持つべき性質である．自然に思いつくように，これを級数の和として定義すれば有限集合の場合に持っていた確率らしいイメージは全て保存される.

Ω がその要素を非負整数で番号づけられる（一列に並べられる）集合とする．簡単のためにその番号付けをそのまま利用して $\Omega = \mathbb{Z}_+$ と置こう．$Q(\{k\}), k \in \mathbb{Z}_+$, が非負で与えられていて，$\sum_{k=0}^{n} Q(\{k\}) = 1$ を満たしているとき，$A \subset \Omega$ に対して $Q(A) = \sum_{k\in A} Q(\{k\})$ とおいたものを離散分布と呼ぶ．ポワッソン分布は離散分布の例である.

ポワッソン分布に戻って，(72) の正定数 λ は Poisson 分布の平均値である；

$$\sum_{n=0}^{\infty} nQ[\{n\}] = \lambda\,. \tag{73}$$

10.3 練習問題.

定義 (72) に出てくる（n によらない）定数 $e^{-\lambda}$ は全事象の確率が 1 になるように決まっているはずである：$\sum_{n=0}^{\infty} Q[\{n\}] = 1$．この式を証明せよ.

また (72) を用いて (73) を証明せよ.

10.4 ポワッソン確率過程と指数分布.

ポワッソン分布からでたらめに到着するバスまでの間に，もう一つ用意しておく数学がある.

無限個の確率変数を確率過程と呼ぶ．特に $X_t, t \geqq 0$, のように確率変数 X_t を区別するパラメータ t が連続で，しかも標本 w に対して $X_t(w)$ が t の関数として興味深いものが興味の対象となる．t はしばしば時間パラメータとして，ばらつきをともなう時間変化の数学モデルとして確率過程が用いられる.

Poisson 確率過程は，1 個単位で時間的にでたらめに発生して累積していく量の数学的モデルである．バスの例では，特定の停留所で見張っていて，時刻 t までに到着（発車）したバスの累積台数を X_t とおく．その性質が Poisson 過程でよく表されることが分かれば,「でたらめに」到着していると判断できる．応用上重要な例として，電話交換台が処理する電話の累積回数やサービスカウンター（レジ，案内カウンター，チケット売場など）への客の累積到着数，通信回線や画像などのビットやピクセル単位のノイズの累積個数，等がある.

時間が連続な確率過程 $X_t, t \geqq 0$, 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.

(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 47 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 47 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

$$\mathrm{P}[\, X_t - X_s = n \,] = \exp(-\lambda\,(t-s))\,\frac{1}{n!}\,\{\lambda\,(t-s)\}^n\,, \quad n = 0, 1, 2, \cdots, \tag{74}$$

に従う，ということである．ここで正定数 λ は単位時間幅 1 内の到着数の期待値であることは Poisson 分布の説明から明らかであろう．Poisson 過程という名前はこの事実に由来する．

一見計算の役に立たないように見える Poisson 過程の定義（4 つの条件）から，具体的な公式 (74) が得られることに注意．

定理 47 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 (74) という具体形が出る「気持ち」を示す．

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

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

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．${}_N\mathrm{C}_n = \dfrac{N!}{n!\,(N-n)!}$ と分数にして，さらに分母と分子をそれぞれ $\sqrt{2\pi N}\,N^N\,e^{-N}$ で割っておく．公式 $\lim\limits_{N\to\infty} \dfrac{N!}{\sqrt{2\pi N}\,N^N e^{-N}} = 1$ を $N!$ と $(N-n)!$ に適用する．$\left(\dfrac{\lambda\,(t-s)}{N}\right)^n$ からも分母に N^n が出ることに注意すると，右辺は

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

と変形される．a, b が N によらないとき成り立つ公式 $\lim\limits_{N\to\infty}(1-\dfrac{a}{N})^N = e^{-a}$ 及び $\lim\limits_{N\to\infty}(1-\dfrac{a}{N})^b = 1$ を用いれば (74) を得る． □

（時刻ゼロから測定開始して）初めて event が発生した時刻を T_0, 以下 n 件目と $n+1$ 件目の event 発生の時間間隔を T_n $(n \geqq 1)$ とおく．T_n たちは $\{X_t\}$ で定まる確率変数である．これについて次の性質が証明できる．

定理 48 X_t が *Poisson* 過程ならば，T_0, T_1, T_2, $\cdots$, は独立同分布確率変数列になる．T_0 の分布は平均 $\mathrm{E}[\, T_0 \,] = 1/\lambda$ の指数分布になる（どの T_n でも同じ）．ここで λ は *(74)* の λ，即ち，単位時間内の到着数の期待値 $\mathrm{E}[\, X_1 - X_0 \,]$ である． ◇

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

$$\rho_\lambda(t) = \lambda \exp(-\lambda\, t) \tag{75}$$

で与えられるものを言う．例えば，時間間隔 T_n が a 以上である確率は

$$\mathrm{P}[\, T_n \geqq a \,] = \int_a^\infty \rho_\lambda(t)\, dt \tag{76}$$

で計算される．

定理 48 の証明． $\{T_n\}$ が独立同分布であることは，Poisson 過程の加法性と一様性から導かれるのだが，省略する（以下の証明にかなり含まれている）．T_n が平均 $1/\lambda$ の指数分布に従うことだけ証明しよう．定理 47 により，Poisson 過程の定義から (74) 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，(74) から (75) を得ることができればよい．$a \geqq 0$ を定数とする．$T_0, T_1, \cdots, T_{n-1}$ をそれぞれある値に固定するという条件の下で $T_n > a$ となる条件付き確率（これを $P[T_n > a \mid T_0, T_1, \cdots, T_{n-1}]$ と書く）を計算する．$S = \sum_{i=0}^{n-1} T_i$ とおいておく．T_n の定義から，求める確率は，時間間隔 $(S\,,\ S+a]$ の間に event が発生しない確率，即ち，この時間に X_t が変化しない確率に等しい．従って，

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

最後の等号は (74) において $n = 0$, $s = S$, $t = S + a$ とおいて得られる．右辺は S によらないから，T_0, $T_1, \cdots, T_{n-1}$ の値によらない（即ち，これらの変数と T_n は独立である）．だから，条件をはずしても等号が成り立つ；$P[T_n > a] = \exp(-a\,\lambda)$. この式と，(76) を見比べれば，$T_n$ の分布の密度は $e^{-a\,\lambda}$ 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）(75) の右辺に他ならない．□

定理 48 の逆も成り立つ．

定理 49 $\{T_n\}$, $n = 0, 1, 2, \cdots,$ が独立同分布で，平均 $1/\lambda$ の指数分布ならば，

$$X_t = \min\{n \mid \sum_{i=0}^{n} T_i > t\} \tag{77}$$

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

定義 (77) の右辺は時刻 t までに発生した event の総数を表している（各自確認せよ）．定理 49 の証明は省略する．X_t の加法性と独立性を証明すればいいのだが，$\{T_n\}$ が独立同分布であることだけでなく，指数分布に従うことも用いる必要がある．

定理 49 は event 発生の累積観測データを得たとき，それが Poisson 過程になっている（つまりでたらめな event）かどうかを確かめるのに便利である．(i) まず，累積到着数 sample path $X_t(\omega)$ を多数集める．１時間毎に（毎日，など）データを異なる sample path とみなす.（t は 0 から１時間（１日）しか動かないことになる.）(ii) それぞれ，最初の event 発生時刻を $T_0(\omega)$ とし，以下順に発生間隔を $T_n(\omega)$ に割り振る．(iii) 各 n ごとに T_n の sample として割り振られたデータのヒストグラム（度数分布）または，分布関数，を作って，それが指数分布になることを確認する．(iv) 異なる n の間の T_n が確率変数として独立（第４回講義参照）なことを確認する．

以上を満たせば，定理 49 より，この観測データは Poisson 過程になっていると結論される．Poisson 過程の定義（事実上，加法性と一様性）をチェックする方法もある．

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

10.5 練習問題．

問 1． 指数分布の密度 (75) についても，全事象の確率 $\int_0^\infty \rho_\lambda(t) = 1$ であること，及び，平均が $\dfrac{1}{\lambda}$ になること，を証明せよ．最後に，平均 $\dfrac{1}{\lambda}$ の指数分布に従う確率変数 T が，定数 a 以上になる確率 $\mathrm{P}[\, T \geqq a \,]$ を計算せよ．

問 2． 計算機の乱数を利用して Poisson 過程 X_t の sample path を発生して見よ．乱数の初期値を変えていろいろ発生させ，いくつかを，X_t を縦軸 t を横軸にして図示せよ．標本を多数作って，適当な s, t をいくつかとって，時間 $(s,t]$ の間の到着数 $X_t - X_s$ の統計分布（0 の標本がいくつ，といったふうに）をとってみよ．また，Event 間隔 T の統計分布も図示して見よ．

標本の発生には，event 発生間隔 T が指数分布 (75) に従うことを使うのが便利かも知れない．(計算機で，区間 $[0,1)$ の一様分布に従う乱数 w から，指数分布に従う乱数 u を作るには，$u = -\log(w)$ と置けばよい.)

10.6 極端な仮想のケース — でたらめに到着するバス．

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

「運行間隔を表す確率変数 T の分布の密度が指数分布

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

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

T が指数分布に従うということは，もちろん，t_0 と t_1 を $0 < t_0 < t_1$ を満たす定数とするとき $t_0 < T < t_1$ となる確率 $\mathrm{P}[\, t_0 < T < t_1 \,]$ が

$$\mathrm{P}[\, t_0 < T < t_1 \,] = \int_{t_0}^{t_1} \lambda \exp(-\lambda t)\, dt\,, \tag{78}$$

で計算されるということである．

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

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

このことを逆に見ると，

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

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

乗客の平均待ち時間が $1/\lambda$ になることは次のように証明される．

定理 48 より λ が単位時間当たり通る車の台数の平均 $\mathrm{E}[\, X_{t+1} - X_t \,]$ に等しく，さらに，平均通過時間間隔の逆数 $1/\mathrm{E}[\, T_n \,]$ に等しいことを示した．平均待ち時間と言ったのは，ある時刻 s に停留所に到着した乗客がバスがくるまで待つ時間の平均値 $\mathrm{E}[\, \sum_{i=0}^{X_s} T_i - s \,]$ のことである（この式がそういう意味を持つことは各自確認せよ）．この値は，$T_n > s'$ なる条件下での $T_n - s'$ の期待値（条件付き期待値） $\mathrm{E}[\, T_n - s' \mid T_n > s' \,]$ に等しい．

ある特定のバスに乗れるケースに注目する．それは直前のバスが通り過ぎてから問題のバスが到着する直前までに乗客が停留所に来た場合である．仮に乗客は直前のバスが通り過ぎてから時間 s たって停留所に着いたとしよう．ということは，問題のバスと直前のバスは間隔 T が s 以上だということである．この

条件の下で待ち時間 $T-s$ の期待値 $\mathrm{E}[\,T-s \mid T>s\,]$ を計算すればよい.(| $T>s$ は $T>s$ という条件付きで期待値を取ることを表す.) T の分布の密度 (78) から

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

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

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

と変形され, これを計算すると

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

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

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

Poisson 確率過程 $X(t)$ は次の性質を持つ.

(i) 標本が階段関数, 即ち, ときどき (バスが到着するごとに) 1 ずつ増え, それ以外のときは時間的に一定の値をとる.
(ii) 加法的, 即ち時間 $(t_0, t_1]$ の間の $X(t)$ の増分 $X(t_1)-X(t_0)$ (何台バスが来たか) が t_0 以前の $X(t)$ と独立である.
(iii) 時間的一様, 即ち, 時間幅 $(s,t]$ の間に到着するバスの台数 $X(t)-X(s)$ の分布は $t-s$ だけで決まる.

逆に以上の性質を持つ確率過程は Poisson 確率過程である. (これが本来の定義.) 以上の事実を用いると, いろいろな量を式変形で計算することができる. 例えば, (78) に出てくる λ が実際に平均運行間隔の逆数 (即ち単位時間当たり通るバスの台数の平均) $\mathrm{E}[\,X(a+1)-X(a)\,]$ に等しいことなども示すことができる. また, 平均待ち時間が $1/\lambda$ に等しいことを先ほど証明したが, この証明も実は加法性と時間的一様性が本質的な役割を果たしている.

10.7 練習問題.

問 1. 平均運行間隔が定数 a で与えられるバスのバス停において, 乗客の到着時刻の分布が一様分布のとき, この乗客の平均待ち時間を次の各場合に求めよ:

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

問 2. あなたはバス会社 (その他どんな顧客サービス会社でも当てはまりうるが) の顧客相談窓口に配属されているとする. 乗客から次のような苦情が来た.「停留所の時刻表では 10 分に 1 本となっているのに 20 分待たされることもざらだ. 平均を取ったら 15 分待たされている. 間引き運転しているのではないか?」あなたの会社が間引き運転をしていない良心的な会社の場合, この苦情に正しく説明するにはどう返事をすればよいかを考えよ. (先方は確率論の専門家でもないし, この講義も聞いていない, という前提で何とか分かるように説明せよ.)

問 3–1． バスが，本文説明の記号で $\lambda = 10$ 分なるポワッソン過程に従って到着するとする．ある日バス停に来て 2 分後，バスが来ないうちに忘れ物に気づいて急いで取りに戻り，3 分後に再度バス停に戻った．あと平均何分待てばバスは来るか？

問 3–2． バスはともかく，放射性物質が放射線を出す現象はポワッソン過程であることが知られている．いま $\lambda = 10$ 分のポワッソン過程に従って放射線を出す放射性物質の放射線を研究室で測定していた[26]．実験を始めて 2 分後，一時的に 3 分間夜食のカップ麺を作っていて測定し損なった．次に放射線を検出するまで平均何分待つことになるか？

11 宮城沖地震発生期間の分布 — ブラウン運動の脱出時刻．

[Karatzas–Shreve, p.197 (5.1.2)] によれば，W_t を原点 0 を出発する 1 次元ブラウン運動，ν と m を正定数とし，$X_t = W_t + \nu t$ という法則に従う確率過程（drift 付きブラウン運動）が $x = m\nu$ を hit する時刻を T とおくと T の分布は

$$\mathrm{P}[\, T \in [t, t+dt] \,] = \frac{m\nu}{\sqrt{2\pi t^3}} e^{-\frac{(t-m)^2\nu^2}{2t}} dt, \quad t > 0, \tag{79}$$

で与えられる．$\nu = \sqrt{\dfrac{m}{v}}$ と変換すると

$$\mathrm{P}[\, T \in [t, t+dt] \,] = \frac{1}{\sqrt{2\pi v (t/m)^3}} e^{-\frac{(t-m)^2}{2v(t/m)}} dt, \quad t > 0. \tag{80}$$

Laplace 変換の公式

$$I(p, a) = \int_0^\infty e^{-pt - \frac{a^2}{4t}} \frac{dt}{\sqrt{t^3}} = \frac{2}{a}\sqrt{\pi} e^{-a\sqrt{p}}, \quad a > 0,\ p > 0,$$

からラプラス変換を計算できるので，

$$\mathrm{P}[\, T \in [0, \infty) \,] = 1, \quad \mathrm{E}[\, T \,] = m, \quad \mathrm{V}[\, T \,] = v \tag{81}$$

が順次求まる．X_t を，時刻 t における地震（一般に製品の故障やシステムの破壊）を引き起こす力や局所的な欠損の総量などの総合的な指標とみなす．たとえばプレート説では日本の近くで沈む太平洋プレートに引きずられた日本の乗っている地盤が，応力限界に達して元に戻るときに地震が起こるとされているが，その「ひずみ」の大きさが指標の主要部分となるであろう．地震が起こるたびにひずみは 0 にリセットされ，種々の条件で揺動を受けつつおおむね一定速さでひずみが蓄積してある臨界ひずみに達すると地震となる．これをモデル化するのに原点を出発点とする確率過程で，ひずみに応じた drift と揺動に応じたブラウン運動の和の脱出問題を考えるのは自然であり，したがって (79) や (80) が地震発生間隔の分布の簡単なモデルとしては推薦できる．

このモデル理論分布への標本の当てはめは，(81) によって可能である．2004 年 6 月現在の地震調査研究推進本部の web page「宮城県沖地震の長期評価」

`http://www.jishin.go.jp/main/chousa/00nov4/miyagi.htm`

の表 2 によれば宮城県沖地震の過去の発生記録は次のようになっている．

[26] 筆者は 1978 年に物理学科の学生実験で本当に放射線の測定実験をやっていた．「魔の放射線」と呼ばれていた．いちばん運の悪い物質に当たると信号より背景ノイズ（誤差）のほうが大きくて，統計処理が不可能になったようだ．総元締めは 2003 年にノーベル物理学賞をとった小柴教授で，彼の指揮したカミオカンデという検出装置は，設計時点で意図していた現象は何一つ検出しなかったが，退職直前偶然超新星爆発が起きてそれに伴うニュートリノを検出した．元同僚の教授が「幸運の星の下に生まれた男がいる」と評したと言われている．「制御できないばらつき」そのもののような先生であった．
ところで「魔の放射線」学生実験は今もやっているのだろうか，それとも管理上の理由でできなくなっているのだろうか．

発生年/月/日	間隔（年）
1793/02/17	
1835/07/20	42.4
1861/10/21	26.3
1897/02/20	35.3
1936/11/03	39.7
1978/06/12	41.6

この地震発生間隔を，大きさ $n = 5$ のデータとして §4.4 で標本平均 $\overline{X}$ と不偏分散 V を計算してあった．

$$n = 5,\ \ \overline{X} = 37.06,\ \ V = 43.74\ \ (\sqrt{V} = 6.61).$$

詳しくやるためには区間推定を行うべきかもしれないが，データ数も少ないので点推定で話を進めることにして，

$$m = 37.06,\ \ v = 43.74$$

と (80) で地震発生間隔分布の推定とする．なお，地震調査研究推進本部は

`http://www.jishin.go.jp/main/chouki2/shuhou/node10.html`

において種々の分布への当てはめを検討しているが，なぜかここで説明したもっとも自然かつ簡単と思われるモデルに 2004 年 6 月現在言及がない．あろうことか指数分布（ポワッソン確率過程）まで採用しているが，この分布は前回の地震と無関係に次回の地震が起きる，という描像であり，ひずみの蓄積と解放という地震の基本原理を無視している．統計モデル担当部署はよほど軽視された閑職なのであろう．同情を覚える．

ここから先は，地震調査研究推進本部上記 web page の手法にしたがって，最後の地震発生から現在まで地震が起きなかったという条件付き確率で現在から何年か後までに地震が起きる確率を計算する．まず，現在を 2004/06/17 とし，この講義の主な受講生（大学 2 年生）諸君が学部を卒業する 2007/03/31，そのまま東北大学大学院博士前期過程に進学した場合に修了する 2009/03/31，さらに東北大学大学院博士後期程に進学した場合に修了する 2012/03/31，までの地震発生確率をそれぞれ求める．最後の地震発生日 1978/06/12 から現在までの時間を $t_0 = 26.0$，最後の地震発生日から求める時刻までの時間を t，とおくと，条件付き確率は

$$q(t) = \frac{\mathrm{P}[\, t_0 < T < t\,]}{\mathrm{P}[\, t_0 < T\,]} \tag{82}$$

で与えられる．Mathematica に数値積分させると，

$$\mathrm{P}[\, T > t_0\,] = \int_{26.0}^{\infty} \frac{1}{\sqrt{2\pi v (t/m)^3}} e^{-\frac{(t-m)^2}{2v(t/m)}}\, dt = 0.97\ \ (m = 37.06,\ v = 43.74,\,)$$

および

期限	前回からの 経過時間 t	期限までの 発生確率 $q(t)$
2007/03/31	28.8	6.5%
2009/03/31	30.8	15%
2012/03/31	33.8	32%

を得る．

統計学の教科書

[小針] 小針あき宏，確率・統計入門，岩波書店，1973. 日見

確率論から説き起こして論理的整合性と統計学の最低限の基礎（正規母集団の推定・検定）の両方を配慮した教科書．問題多数．丁寧な解答付き．きちんと自習するのに最適.

[宮沢光一] 宮沢光一，経済分析のための数学入門 5　経済分析と決定理論，東洋経済新報社，1971.

[楠岡, 林周二] がこの著者を引用していてよい教科書らしいが，未入手のまま絶版.

[林周二] 林 周二，統計学講義，丸善，1973 .

古典的な統計学の基礎教科書の 1 例．筆者が教養教育科目で学んだ先生.

[渡辺浩] 渡辺 浩，日本医科大学応用数学講義録，1995, 2002.

古典的な内容に向けて，読みやすく題材を選んだ講義ノート.

読み物的教科書

[楠岡] 楠岡 成雄，確率・統計，森北出版（新数学入門シリーズ 7），1995.

確率論の大家が，統計学の基礎的な問題意識（何をやりたい学問か，どこが論争の種になりやすいか）を極めてわかりやすく解説した薄い本．数学の詳しい知識が無くても読める．統計学について考えることが好きな人の必読書.

[服部] 服部 哲弥，数理解析特論講義録（宇都宮大学大学院博士前期課程情報工学科），1985.

確率論の基礎的事項に関するトピックスを読み切り式にしてみたもの.

統計データ

[人口] 人口動態調査表，厚生労働省統計表データベース

`http://wwwdbtk.mhlw.go.jp/toukei/index.html`

から統計調査一覧のページに入ればリンクがある．(2004 年時点で平成 15 年 (2003 年) のデータは出そろってない．出生率が低下したことを年金法の変更が成立するまで隠していたといういわくのデータももちろん未掲載．平成 14 年のものを利用した.)

[気温] 気温データのページ

`http://www.nakamuu.jp/kion/kion.htm`

個人で新聞等で調べたデータらしい.

確率論の教科書

[熊谷] 熊谷 隆，確率論，共立出版（新しい解析学の流れ），2003.

[西尾] 西尾 真喜子，確率論，実教出版，1978.

[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），1993.

[熊谷, 西尾] は本格的な安心して勉強できる基礎教科書だが，とりあえず薄いのを欲する向きのための 1 冊.

[辞典] 日本数学会編，岩波数学辞典第 3 版，岩波書店，1985.

[熊谷, 西尾] に載っていない単語はこちらへ.

確率過程論の教科書

[Karatzas–Shreve] I. Karatzas, S. E. Shreve, *Brownian motion and stochastic calulus,* 2nd ed., Springer,1991.

確率過程論，特にファイナンスの数学的基礎であるマルチンゲール，伊藤積分，確率微分方程式の非常に詳しい教科書.（統計力学と関連する確率偏微分方程式には対応していない.）

乱数生成アルゴリズムの教科書

[Devroye] L. Devroye, *Non-Uniform Random Variate Generation,* Springer–Verlag, New York, 1986.

乱数生成アルゴリズムの辞書本の決定版．こういう本でも絶版になるとは恐ろしい.

[Knuth] D. E. Knuth, *The art of computer programming,* 2nd ed., vol. 2, Addison-Wesley, Reading, MA, 1981.

計算機科学の伝説的大家であり，本書原稿も利用する組版ソフト TEX の発明者としても有名な著者の伝説的名著．計算機による乱数の生成の基礎.

[伏見] 伏見 正則, 確率的方法とシミュレーション, 岩波書店（岩波講座応用数学 方法 10），1994.

[楠岡] と同様，とりあえず薄いのを欲する向きのための 1 冊.