



演習 確率と統計

さいころからランダムウォークまで



初級 慣れよう

定型的な問題で，筋道がほぼ決まっている



中級 頑張ろう

思考力を高め，本質を理解しよう



応用 チャレンジしよう

問題発掘，問題解決能力の向上，実社会で活きている数学体験
深い理解や簡単なプログラミング

松葉育雄^{*1}

2020 年 7 月 29 日

©matsuba.i

- 教員養成学部学生（数学），看護学部の講義資料，および工学部情報系学生向けの教科書「シリーズ工学のための数学 5 確率（朝倉書店）」などを参考
- 大学入試問題，教員採用試験問題（定型問題），各種出版本を参考
- texlive2019(コンパイル upLaTeX)，emath など各種スタイルファイル
計算用補助ツール sagemath9.0, Python, R3.6
- 絵文字は twitter を参照
- ロゴはフラクタルを用いて作成

^{*1} 秀明大学 matsuba@mailg.shumei-u.ac.jp

目次

第 1 章	確率	3
§ 1	確率とは	3
§ 2	確率の基本法則	20
§ 3	応用問題 	37
§ 4	確率変数	40
§ 5	いろいろな確率分布	45
§ 6	確率変数の関数	55
§ 7	モーメント, 母関数	66
§ 8	応用問題 	68
§ 9	確率過程	70
第 2 章	統計	83
§ 1	統計の基本	83
§ 2	標本と母集団	90
§ 3	推定と検定	94
§ 4	区間推定	96
§ 5	検定	101

第1章

確率

§ 1 確率とは

◆ 1 確率統計は社会のいろいろな場面で使われている。



1. 下記の項目で確率統計が用いられている事項を2つ書きなさい。

自然 _____

社会 _____

医療 _____

産業 _____

経済 _____

スポーツ _____

2. 確率統計を用いてノーベル賞受賞に結びついた有名な例など、以下の項目について調べ、レポートにまとめよ。

コルモゴロフ，ベイズ，マルコフ，伊藤清の業績

アインシュタインのブラウン運動（1905 年）

シュレディンガー方程式（1926 年）

ブラック・ショールズ方程式（1973 年）

エングルの ARCH モデル（1982 年）

ニュートリノ（2015 年）

物理と順列組合せ

n 個の状態に r 個の粒子を分配するとき、以下の条件での場合の数

- 箱も球も区別ができる： n^r （マクスウェル・ボルツマン統計）

- 箱の区別はできるが球は区別できない： ${}_{n+r-1}C_{n-1}$ (ボーズ・アインシュタイン統計)
- 箱も球も区別できず，さらに一つの箱には1つの球しか入らない： ${}_nC_r$ (フェルミ・ディラック統計)

3. 確率統計ではしばしばパラドクス（正しそうに見える前提と妥当に見える推論からでは，受け入れがたい結論が得られる事をいい，逆説，背理とも言われる）と名づけられた問題が数多くある．本演習でもいくつか取り上げるが，どういうものがあるか調べてみよう．

さいころ  の目表し方：     

コイン  の表裏の表し方：H, T

事象と要素の数

$\emptyset$ (空事象)：起りえない集合

$A \cup B$ (和事象)： A, B の少なくとも一方が生起する事象

$A \cap B$ (積事象)： A と B がともに生起する事象

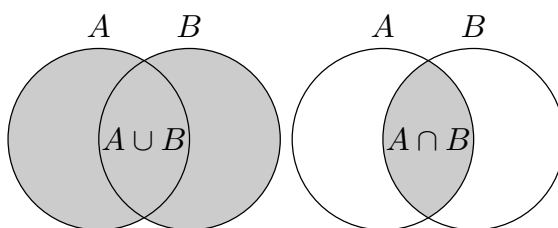
$\bar{A}$ (余(補)事象)： A が起らない事象, $\bar{A} = \Omega - A$

$A - B$ (差事象)： A に属して B に属さない事象 $A - B = A \cap \bar{B}$

排反：事象 A と B が共通の要素を持たない $A \cap B = \emptyset$

特に $B = \bar{A}$ とすると， $A \cap \bar{A} = \emptyset$

互いに排反：複数の事象 $A_i (i = 1, 2, 3, \dots)$ があり， $A_i \cap A_j = \emptyset (i \neq j)$



- 事象 A の要素数を記号 $|A|$ で表す．
- 空事象には要素がないので， $|\emptyset| = 0$
- 任意の事象 A, B に対して， $|A \cup B| = |A| + |B| - |A \cap B|$
特に，排反の場合． $A \cap B = \emptyset (|A \cap B| = 0)$ で， $|A \cup B| = |A| + |B|$
- 積事象， $|A \cap B| \leq |A|, |A \cap B| \leq |B|$
- 差事象， $|A - B| = |A| - |A \cap B|$
- $A \subseteq B$ に対して， $|A| \leq |B| (A \subseteq B)$

ド・モルガンの法則： $\overline{A \cap B} = \bar{A} \cup \bar{B}, \overline{A \cup B} = \bar{A} \cap \bar{B}$

◆2 算術的確率 (いろいろな言い方がある)

さいころを振る実験において、 (1 の目) から  (6 の目) の目が出る **事象** は **根元事象** という。根元事象 $\omega_i (i = 1, 2, \dots, 6)$ で表した **標本空間** $\Omega = \{\omega_1, \omega_2, \dots, \omega_6\}$ の確率は、必ずどの目かは出るので $P(\Omega) = 1$ 。 $|\Omega| = 6$ で、各根元事象は同じ確率で起こるとすると、どの根元事象も $|\omega_i| = 1$ なので、確率は $P(\omega_i) = \frac{|\omega_i|}{|\Omega|} = \frac{1}{6}$ 。コインの場合、H と T は同様に確からしいとすると、 $P(H) = P(T) = \frac{1}{2}$ で、標本空間 $\Omega = \{H, T\}$ の確率は $P(\Omega) = 1$ 。

一般化すると、可能な実験結果が n 種類あり、それらの結果は等しい確率で現れるとする。このとき、ある事象になるのが m ($0 \leq m \leq n$) 種類あったとすると、その事象の確率は、 $P = \frac{m}{n}$ である。

◆3 以下の例題などから、順列組合せ、標本空間、事象、同様に確からしい、という確率の基本事項を身につけよう。 

10 本中あたりが 2 本含まれているくじがある。このくじを 10 人が順に引くとき、10 番目に引く人があたりを引く確率を求めよ。

◇確率は $1/5$ になるのは当たり前だと思うだろうが、事象、標本空間、前提条件など正確に述べた上で、結果を導く過程を説明できるだろうか。

1 番目から 10 番目までの人が引いたくじの結果を表す 事象 を列

$$(x_1, x_2, \dots, x_{10})$$

で表そう。ここで、 $x_i = 0$ (はずれ), $x_i = 1$ (あたり) とする。このなかで、2 人が当たりくじを引くので、標本空間 Ω は、 $(x_1, x_2, \dots, x_{10})$ の中に 1 が 2 個、0 が 8 個含まれる列の集合である。そのような列は $|\Omega| = {}_{10}C_2$ 個あり。これ以上分解できないので、各事象は根元事象である。また、各根元事象が現れるのは 同様に確からしい (それらはすべて等確率で生起する) と仮定する。次に、10 番目の人があたりを引く 事象 を A とすると、事象 A は、

$$(x_1, x_2, \dots, x_9, 1)$$

となる列で表され、 $(x_1, x_2, \dots, x_9)$ の中に 1 が 1 個現れるので、 $|A| = {}_9C_1$ である。よって、事象 A の確率は $P(A) = \frac{|A|}{|\Omega|} = \frac{{}_9C_1}{{}_{10}C_2} = \frac{1}{5}$ となる。なお、何番目に引いても同じ計算になるので確率も同じ。



1個のさいころを6回続けて投げるとき、1の目が1回、3の倍数の目が2回、その他の目が3回出る確率を求めよ。

◇【2018 神奈川県中高】 6回さいころを投げた時の目の出方を列

$$(x_1, x_2, \dots, x_6)$$

と表そう。ここで、 x_i は $(1, 2, 3, 4, 5, 6)$ のいずれかである。標本空間を Ω とすると、すべての列の個数から $|\Omega| = 6^6$ で、各列で表される根元事象は同様に確からしいと仮定する。問題の事象を A とすると、

$$(\square, \square, \square, \square, \square, \square)$$

のような列である。 $\square$ の目が1回出るのは ${}_6C_1 = 6$ 通り、残り $6-1=5$ 回の内、3の倍数の目 ($\square$ か $\square$) が2回出るのは ${}_5C_2 \times 2^2 = 40$ 通り、残り $6-(1+2)=3$ 回の内、その他の目 ($\square$ か $\square$ か $\square$) が3回出るのは ${}_3C_3 \times 3^3 = 27$ 通りで、 $|A| = 6 \times 40 \times 27$ 。よって、確率は $P(A) = \frac{|A|}{|\Omega|} = \frac{6 \times 40 \times 27}{6^6} = \frac{5}{36}$ 。

4. コインを $n(n \geq 1)$ 回投げたとき、全て同じ面 (H あるいは T) になる事象に関して以下の問いに答えよ。

(a) 標本空間、根元事象はどのようなものか。

(b) 問題の事象を A として、その事象の確率 $P(A)$ を求めよ。

5. さいころを2回振る実験で、2番目に振ったさいころの目が最初のさいころの目よりも大きくなる事象に関して以下の問いに答えよ。

(a) 標本空間、根元事象はどのようなものか。

(b) 問題の事象を A として、その事象の確率 $P(A)$ を求めよ。

➡ (b) 条件を満たす場合の数を数える。この例は簡単であるが、表 (縦 i , 横 j) に書いて数えるのも分かりやすい。

6. 3本の当たりくじを含む10本のくじがある。最初に1本引き、もとに戻さないで次に1本引くとき、2本とも当たる確率を求めよ。

➡ 簡単と思うだろうがいろいろな考え方がある。答えがあえば終わりというのではなく、確率の理解を深めるためにも解答、別解をよく読んでほしい。

7. 目の順番 1個のさいころを3回投げて出る目の数を順に a, b, c とする。

(a) 1回目に出た目 a よりも2回目に出た目 b の方が大きくなり、さらに3回目 c に出た目が2回目に出た目 b よりも大きくなる確率を求めよ。

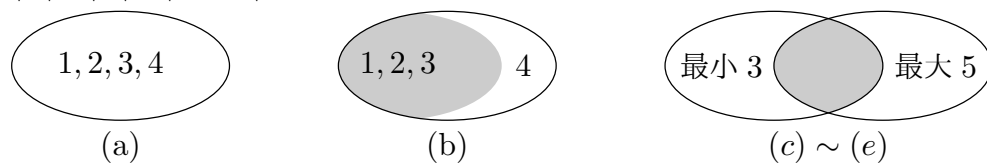
(b) $a \leq b < c$ となる確率を求めよ。

➡ すべての場合を数えることでもできるが、一度解答のように解くとテクニックは身につくが、それだけで終わらないように！

8.  目の最小最大 次の確率を求めよ.

- (a) 3 個のさいころを同時に投げるとき, 出た目の最大値が 4 以下である確率
- (b) 3 個のさいころを同時に投げるとき, 出た目の最大値が 4 である確率
- (c) 1 個のさいころを 3 回続けて投げるとき, 出る目の最小値が 1 で, かつ最大値が 6 である確率
- (d) n 個のさいころを同時に投げるとき, 出る目の最小値が 3 であり, かつ最大値が 5 である確率. ただし, n は 2 以上の自然数とする.
- (e) 3 個のさいころを同時に投げるとき, 出る目の最大値を M , 最小値を m とする. $M - m = 3$ となる確率

 ベン図を書いて場合の数を求めよう. (c)~(e) $|A \cap B|$ を求める問題であるが, 余事象 $|\overline{A \cap B}|$ の方の計算をする. ド・モルガンの法則を利用すると, $|\overline{A \cap B}| = |\overline{A} \cup \overline{B}| = |\overline{A}| + |\overline{B}| - |\overline{A} \cap \overline{B}|$.



9.  目の和 次の確率を求めよ.

- (a) 3 個の異なるさいころを同時に投げるとき, 目の和が 8 になる確率
- (b) 大中小 3 つのさいころを同時に投げるとき, 目の和が 9 になる確率

 すべての場合を数え上げればよいが, (a) では 8 個を 3 つに分ければよいので, ○○○○○○○○ をどのように区切ればよいかを考えよ.

10.  目の積 以下の確率を求めよ.

- (a) 3 個のさいころを同時に投げるとき, 出た目の積が偶数になる確率
- (b) 2 個のさいころを同時に投げるとき, 出た目の積が 6 の倍数になる確率
- (c) 3 個のさいころを同時に投げるとき, 出た目の積が 6 の倍数になる確率
- (d) 3 個のさいころを同時に投げるとき, 出る目の積が 15 の倍数になる確率

 (a) 余事象 (積が奇数) を利用. (b) $6 = 2 \times 3$ から, A:2 で割切れる事象, B:3 で割切れる事象とすると, 問題の事象の要素数は $|A \cap B| = |\Omega| - (|\overline{A}| + |\overline{B}| - |\overline{A} \cap \overline{B}|)$ と表せる. (c)(d) もほとんど同じ. (c) は 6 の目を含む場合, 3 と 2 を含む場合などと規則正しく数え上げる事でも可能. (d) $15 = 3 \times 5$ から, 3, 6, 5 で割り切れればよい.

11. $1 \square 2 \square 3 \square 4 \square 5$ の $\square$ に $+$ または $-$ のいずれかを無作為に選んで数式を作るとき, その答えが 3 の倍数になる確率を求めよ.

 $\square 3$ の $\square$ には $\pm$ いずれを入れても 3 の倍数.

12.  へんな 目の出方にかたよりのあるさいころがある. このさいころを 2 回投げ

たとき、出た目の和が2である確率は $\frac{1}{64}$ 、3である確率は $\frac{1}{40}$ であった。 $\square$ が出る確率を求めよ。

13. 2次方程式と絡めた確率

(a) さいころを3回投げて、出た目の数を順に a, b, c とするとき x の2次方程式 $abx^2 - 12x + c = 0$ が重解をもつ確率を求めよ。

(b) 大小2つのさいころを同時に投げ、大きいさいころの出た目の数を a 、小さいさいころの出た目の数を b とする。このとき、2次方程式 $x^2 + ax + b = 0$ の解が有理数となる確率を求めよ。

👉 判別式、覚えている？後は数え上げ。

14. 白玉、赤玉 以下の確率を求めよ。

(a) 白玉6個、赤玉4個が入っている袋から、玉を同時に3個取り出すとき、白玉1個と赤玉2個が出る確率

(b) 白玉3個、赤玉6個が入っている袋から、玉を同時に4個取り出すとき、白玉2個と赤玉2個が出る確率

(c) 白玉3個、赤玉2個、青玉1個が入っている袋から、玉を同時に3個取り出すとき、全部異なる色である確率

15. A, B, C, D, E の5人が、くじ引きで順番を決めて一列に並ぶとき、両端がAとBである確率を求めよ。

16. A, B, C, D, E, F, G の7人が、くじ引きで順番を決めて横一列に並ぶとき、AはBよりも左、BはCよりも左となる確率を求めよ。

👉 まずA, B, Cは同じもののだとして(○)として、○, ○, ○, D, E, F, Gの並びを考える。そこで、AはBよりも左、BはCよりも左になるように、3つの○に左から順にABCを入れる。

17. 男子3人と女子3人が1列に並ぶとき、男子と女子が交互に並ぶ確率を求めよ。

18. 委員になりた〜い 5人から2人の委員を選ぶことになった。5枚のカードを用意して、その内の2枚にあたりと書いたくじをつくり、5人が順に引くこととした。委員になりたい人は何番目に引くのがよいだろうか(ただし、2人の委員を決めるのが目的なので、カードは元に戻さない)。

👉 1番目に委員になる、2番目に委員になる確率など順に計算していこう。

19. じゃんけん👊👉👌

(a) 3人でじゃんけんを1回するとき、あいこになる確率を以下の考え片にしたがって求めよ。

(ai) 余事象はあいこにならない事象で、1人が勝つ、2人勝つ場合である。

(aii) 全員が同じ手になるときと、全員がそれぞれ異なる手になるとき。

☞ (ai) の方が単純で分かりやすいかも。

(b) 4人でじゃんけんをして、勝者1人を決めたい。1回目のじゃんけんで1人に決まらなかった場合には、負けた人を除いて2回目のじゃんけんをする。

(i) 1回目のじゃんけんで、勝者1人が決まる確率を求めよ。

(ii) 1回目のじゃんけんで、あいこになる確率を求めよ。

☞ (a) と同じようにするか、余事象（勝者が出る場合で、2種類の手が出る）を用いる。

(iii) ちょうど2回目のじゃんけんで勝者1人が決まる確率を求めよ。

☞ 1回目のじゃんけんで残るのは2人、3人、4人(iiの結果を使う)のいずれか。

20. 【モンティ・ホール問題, 1990年】 昔アメリカのテレビで話題になった。3つの扉があり、いずれかの扉の後ろに車と残りの2つの扉の後ろにはヤギがいる。君はいずれかの扉を選んで、あたれば車がもらえる。君は1つの扉を選んだ。司会者は「ヒントを差し上げましょう」と言って、1つの扉をあけヤギを逃がした(司会者は車のある扉を開けたりしない)。そしてこう言う「今なら扉を選びなおしてもいいです」。どうしようか。

☞ 条件付き確率を用いて正確に導くことは後の問題で扱うが、ここでは直感的に考えよう。直感が必ずしも正確でない一例である。

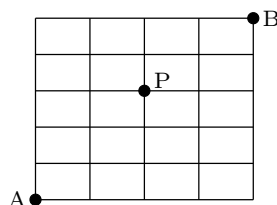
◆4 以降は、最短路で行くときの確率を求める標準的な問題である。図の左下から右上に向かう場合は、最短で行くので逆戻りはしなく、各交差点では上方向(↑)か右方向(→)に移動する。

21. A地点からB地点まで最短の道順で行くとき、以下の問いに答えよ。

(a) 標本空間、根元事象はどのようなものか。

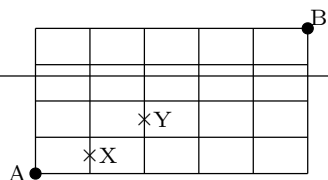
(b) P地点を通る確率を求めよ。

☞ もっとも標準的な問題で答えはすぐ分かるだろうが、何が、そしてなぜ同様に確からしいのか明確にできるだろうか。

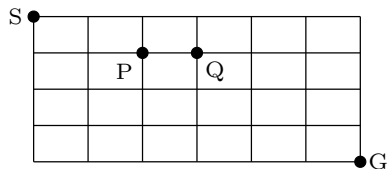


22. XとYが通行止めで通れないとき、A地点からB地点まで最短の道順で行く確率を求めよ。

☞ 余事象として、XあるいはYを通る道を考えよう。このとき、 $|X \cup Y| = |X| + |Y| - |X \cap Y|$ 。



23. S 地点から G 地点まで最短の道順で行くとき，以下の各場合について確率を求めよ．



- (a) 必ず区間 $P \rightarrow Q$ を通って，S から G まで行く場合
 (b) $P \rightarrow Q$ が工事中で通れないとき，S から G まで行く場合

これまでの問題では，前提として同様に確からしい(equally possible)という条件がある．大抵の問題ではこの前提を明確にしないているが，同様に確からしいをどのように解釈するかによって，確率が異なる場合がある．たとえば，2 枚のコインを投げて表の出る回数は $\{0,1,2\}$ の 3 通りあり，それらが同程度に起こり得るならば，確率は $1/3$ である．もちろん同程度で起こり得るのは公正なコインの根源事象で，確率は $1/4$ である．この場合は問題にならない程容易だが，最短路の場合，各道順が道程に選ばれることを前提にするか，各交点でどの方向に移動するかを考えた場合，異なる結果になる．



各面に 1, 1, 1, 2, 2, 3 と書かれたサイコロがある．

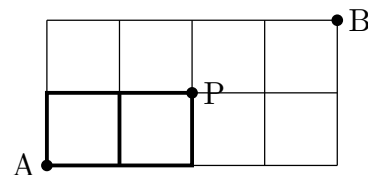
- (a) 目の出方は全部で何通りか．
 (b) 1 の目が出る確率を求めよ．

◇(a)3 通り，(b)1 の目が出る事象を A とする．各面の出る根元事象は同様に確からしいと仮定すれば，確率は $P(A) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}$ ．むろん， $\frac{1}{3}$ ではない．

24. コインを 2 個投げる．表の出る回数は，0（表が出ない），1（表が 1 回），2（表が 2 回）の 3 通りで，それぞれの確率を求めよ．



図のような街がある．A 地点から B 地点まで最短の道順で行くとき，P 地点を通る確率を以下の仮定の下で求めよ．ただし，最短で行くので，各交差点では上方向（↑）か右方向（→）に移動する．



- (a) 最短路として同様に確からしい
- (b) 各交差点で同様の確からしさで道を選ぶ

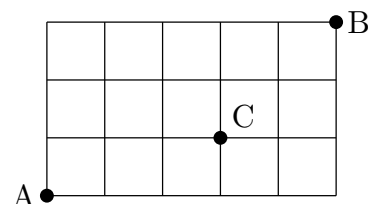
◇(a) A 地点から B 地点までは ${}_6C_2$ 通り，そのうち P を通るのは ${}_3C_2 \times {}_3C_2$ 通りある．よって， $\frac{{}_3C_2 \times {}_3C_2}{{}_6C_2} = \frac{3}{5}$

(b) A から P までの 3 通りに道があるが，すべて同じ確率 $\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$ なので $\frac{3}{8}$ ．P 地点から B 地点までは必ず到達するので，その確率は 1(具体的に計算しても， $\frac{1}{2} + \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = 1$ となる)．したがって， $\frac{3}{8} \times 1 = \frac{3}{8}$

この問題が「最短路の中で P 地点を通る確率を求めよ」ならば，答えが 2 通りある．つまり，同様に確からしいの解釈によって異なる結果が導かれる．この意味で，同様に確からしさをいつも意識して確率を考えてほしい．



図のように東西に 4 本，南北に 6 本の道があり，各区画は正方形である．P, Q の二人はそれぞれ A 地点、B 地点を同時に出発し，最短距離の道順を取って B 地点 A 地点に向かった．ただし，2 通りの進み方がある交差点では，それぞれの選び方の確率は $\frac{1}{2}$ であるとする．P, Q が C 地点で出会う確率を求めよ．



◇【北里大入試より】 (a) 最短路として同様に確からしいとして求める．A 地点から B 地点までは ${}_8C_2$ 通りある．よって，確率は， $\left(\frac{{}_4C_1 \times {}_4C_2}{{}_8C_5}\right)^2 = \left(\frac{3}{7}\right)^2 = \frac{9}{49}$

(b) 下図を参照にしよう．

P が A → C に至る確率(各交差点で上方向（↑）か右方向（→）に移動)：各交点における確率は，特に太字にした確率は以下のように計算できる．

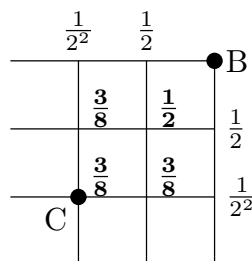
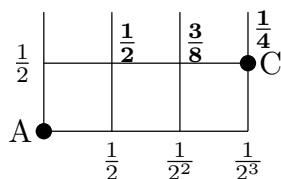
$$\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2} \quad \frac{1}{2} \times \frac{1}{2} + \frac{1}{2^2} \times \frac{1}{2} = \frac{3}{8} \quad \frac{3}{8} \times \frac{1}{2} + \frac{1}{2^3} \times \frac{1}{2} = \frac{1}{4}$$

Q が B → C に至る確率(各交差点で下方向 (↓) か左方向 (←) に移動) :

$$\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2} \quad \frac{1}{2^2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{3}{8} \quad \frac{1}{2} \times \frac{1}{2} + \frac{1}{2^2} \times \frac{1}{2} = \frac{3}{8} \quad \frac{3}{8} \times \frac{1}{2} + \frac{3}{8} \times \frac{1}{2} = \frac{3}{8}$$

以上から, P, Q が C 地点で出会う確率は $\frac{1}{4} \times \frac{3}{8} = \frac{3}{32}$

この問題では, 勿論 (b) の方が求められている.

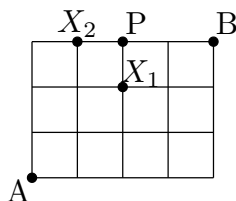


25. 下図において, 以下の問題に答えよ.

(a) A 地点から B 地点まで最短の道順で行くとき, P 地点を通る確率を求めよ.

(b) ある地点で, 東行きと北行きの 2 つの選択肢があるとき, どちらかのコースを選ぶかは同様に確からしいとする. この場合, A から B まで行く行き方の内, P を通る確率を求めよ.

👉 (b) 図の X_1, X_2 は解答の説明のため.



以下は少々難しい問題



26. X, Y, Z の 3 人がさいころを同時に投げるときに出る目をそれぞれ x, y, z とする. このとき, $100x + 10y + z$ の値が 3 で割り切れる確率を求めよ. ただし, さいころの 1 から 6 までのどの目が出ることも同様に確からしいものとする.

👉 $100x + 10y + z = (99x + 9y) + x + y + z$ から, $x + y + z$ が 3 で割り切れる場合を考えればよい.

27. いびつなサイコロがあり, 1 から 6 までのそれぞれの目が出る確率が $\frac{1}{6}$ とは限らないとする. このサイコロを 2 回ふったとき同じ目が出る確率を P とし, 1 回目に奇数, 2 回目に偶数の目が出る確率を Q とする.

(a) $P \geq \frac{1}{6}$ であることを示せ. また, 等号が成立するための必要十分条件を求めよ.

👉 シュワルツの不等式 $(\sum_{i=1}^n a_i^2)(\sum_{i=1}^n b_i^2) \geq (\sum_{i=1}^n a_i b_i)^2$ を使うこともできる.

(b) $\frac{1}{4} \geq Q \geq \frac{1}{2} - \frac{3}{2}P$ であることを示せ.

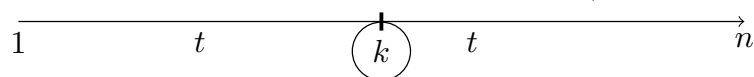
28. 【秘書問題，結婚問題，最良選択問題】 いろいろな呼ぶ名があるこの問題は，以下の仮定および手順で最良の応募者を決める．

1. n 人の応募者に対して 1 名ずつ面談をし秘書を 1 名（のみ）採用する．
2. 応募順に番号が付けられ，複数の応募者が同じ順になることはない（1 から n まで重複無く番号付けする）．
3. 無作為な順序で 1 人ずつ面接を行う．次に誰を面接するかは同じ確率である．
4. 毎回の面接後，その応募者を採用するか否かを即座に決定する．
5. その応募者を採用するか否かは，それまで面接した応募者の相対的番号に基づいて決定する．
6. 不採用にした応募者を後から採用することはない．

以上の手順で最良の応募者を選択する最良選択問題と呼ばれている方法がある*¹．この方法では，面接者は最初の $k-1$ 人の応募者を不採用にし，次の k 番目の応募者がそれまで面接した中で最もよい応募者なら採用する．いま，任意の k 番目に最良の人を選ぶ確率を P_k とおく．

(a) 確率 P_k を $x = \frac{k}{n}$ の関数として表せ．

👉 どの順番でも最良な人である確率は同じ $1/n$ ． $t \leq k$ では 0， k 番目の人の相対順位が 1 位で最終順位も 1 位になる（最良の）確率は，それ以降の $k+1, k+2, \dots, n$ 番目の人は全員相対順位が 2, 3, $\dots$, 位 になる確率である． $k < t$ では $\frac{1}{n} \times$ 相対順位 で，
 $\frac{1}{n} \times 1 \Leftarrow (k \text{ 番目}) + \frac{1}{n} \times \frac{k-1}{k} \Leftarrow (k+1 \text{ 番目}) + \frac{1}{n} \times \frac{k-1}{k+1} + \dots$



(b) n が大きいとき， $P(x)$ の最大値を求めよ．

👉 調和級数の和が $\sum_{j=1}^k \frac{1}{j} \simeq \log k$ と近似できることを用いる．

(c) 最大値が存在することの意味を考察せよ．

29. 多数のくじをいくつかの販売店に配ると，どの販売店で購入するかによって当たりくじを引く確率が異なるだろうか．各販売店にはくじが 1 本あるとする． N 本のくじの内， n ($1 \leq n \leq N$) 本当たりくじがあるとする． k 番目に引いたくじが，当たりくじを引く確率 p_k ($1 \leq k \leq N$) を求めよ．結論は簡単だが，少々理屈っぽく考えよう．

30. 【ベンフォードの法則，1938 年】

膨大な数字が載っている冊子，例えば，会計報告書などから最初の桁の数値

*¹ 他の方法，いろいろな拡張問題もあるので興味あれば調べよう．

$d \in 1, \dots, 9$ (0 は含まれない) の確率は $P(d) = \log_{10} \frac{d+1}{d}$ と表せる事が知られている. この式はベンフォード以降も各種のデータで確認されているが, なぜこの式が成立するかは知られていない. さて, 無作為に数値を選ぶ. その数の最初の桁 5, 6, 7, 8, 9 の中の数字であれば賞金を勝ち取り, 1, 2, 3, 4 の中にあれば掛け金を失う. あなたはこのゲームに賭けますか. 2^n , $n = 1, 2, \dots, 1000$ として数値的に求め以下の表を埋めて, 考えよ.

- (a) ネットなどで調べ実データを用いて, 先頭の数字の頻度を調べよ. 調べた結果を $P(d) = \log_{10}((d+1)/d)$ と比較せよ. また, 成り立たない場合についても調べよ.

数字	1	2	3	4	5	6	7	8	9	計
相対頻度										1
$P(d)$	0.301	0.176	0.125	0.0969	0.0792	0.0669	0.0580	0.0512	0.0458	1

- (b) 2^k (k は整数) と表せるような数の場合は正確に導ける. つまり, 数の集合は, $2, 2^2, 2^3, \dots$ に限った場合である. 先頭の数字が $d = 1$ となる条件を不等式で表すと, $1 \times 10^m \leq 2^k < 2 \times 10^m$ である. たとえば, $k = 10$ とすると, $10^3 < 1024 < 2 \times 10^3$ を満たすのは $m = 3$ があるが, $k = 11$ では m はない. このように k を変えていって m を調べる.

$$\frac{m}{\log_{10} 2} \leq k < \frac{m + \log_{10} 2}{\log_{10} 2} \Rightarrow k \text{ の範囲} = \frac{m + \log_{10} 2}{\log_{10} 2} - \frac{m}{\log_{10} 2} = 1$$

同様に, 先頭の数字が d となる条件は,

$$\frac{m + \log_{10} d}{\log_{10} 2} \leq k < \frac{m + \log_{10}(d+1)}{\log_{10} 2} \Rightarrow k \text{ の範囲} = \frac{\log_{10}(d+1) - \log_{10} d}{\log_{10} 2}$$

以上のことを参考にして理論分布 $P(d)$ を導け.

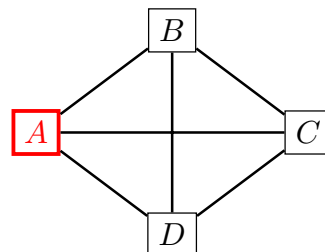
- (c) 【未完 スケーリング則: $P(d) = aP(ad)$ 】

- (d) ベンフォードの法則は, 企業等の経理の不正を暴くために用いられているという. 調査し, レポートにまとめよ.

31. 異なる 4 つの都市 A, B, C, D の間を n 回渡り歩くとき, 次の都市へは等確率で訪れることにする. ただし, 連続して同じ都市を訪れることはしないとし, k 回 ($3 \leq k \leq n$) 渡り歩いて初めてすべての都市を訪れる確率を p_k とする.

- (a) 初めの都市を A (もちろんどの都市でもいいが) として, $k = 3$ 回渡り歩いて初めてすべての都市を訪れる確率は $p_3 = \boxed{}$

- (b) 確率 p_k を求めよ.



32. 一般的な最短路問題



上の問題を一般化した問題に**巡回セールスマン問題**がある。 n 個の都市をちょうど一回ずつ通って最初の都市に戻ってくる最も短い巡回路を求める問題である。ゲノム解析、天体望遠鏡の制御など様々な応用があるが、数学的に解を見つけることが超難解で、現在もなお、近似的な手法を含め様々な挑戦が続いている。興味があれば調べてみよう。

さらに、一般の経路に対する最短路を探索する問題は、情報科学の一分野として扱われている。その基本アルゴリズムとして**ダイクストラの方法**はよく知られている。興味があれば調べてみよう。

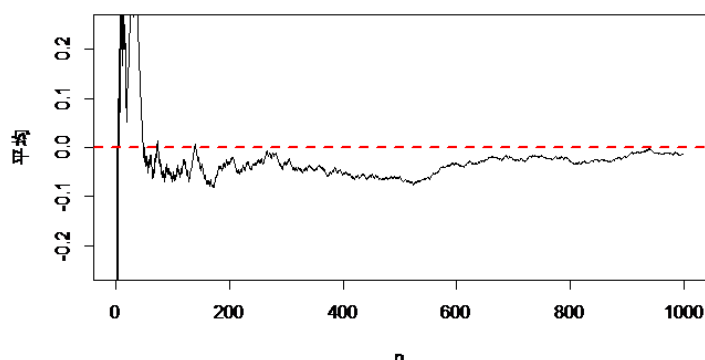
◆ 5 統計的確率（実験的確率などいろいろな言い方がある）

同様に確からしさとはいったいなんだろうか。仮定であるが、その妥当性は結局実験で確かめるしかないだろう。コインを投げて H が出たときに勝ちとするゲームを n 回行い、 n 回のゲームのうち k 回勝ったとしよう。 k は n とともに変化するので $k(n)$ と書くと、ゲームに勝つ確率は、

$$P = \frac{k(n)}{n}$$

と考えることができる。つまり、確率をゲームに勝った回数をゲームを行った回数で割った値、つまり相対頻度として定義する。ただし、 n が有限な限り P は変動するので、 $n \rightarrow \infty$ で一定になったとして、その値を確率として**モデル化**する。

模擬的に -1 と 1 の値をとる一様乱数を用いて、出てきた乱数の値が -1 の時を裏、 1 の時を表とする。結果は、下図に示すように、 P は n の値によって異なる。しかし図をよく眺める、 n が小さいときは値は大きく変動するが、 n を大きくすると、次第に 0 （つまり $P(H) = P(T) = \frac{1}{2}$ ）に近づくように見える。といっても、 $n = 1000$ あたりでも十分とは言えない。実際のコインで実験すればもっと変動するであろう。



なお、実際のコインをいろいろな条件で 1 万回以上投げた実験（ご苦労様！）がイン

ターネットで公開されているようなので、調べてみよう。で、結局 $P = \frac{k(n)}{n} = \frac{1}{2}$ といっているだろうか。

A 君 だいたい $\frac{1}{2}$ だろうということに関しては、たぶん異論のある生徒はあまりいないと思う。しかし、 $\frac{1}{2}$ に近いからと言って正確に $\frac{1}{2}$ ではない。だから、同様に確からしいと言ってるのであって、同じ確かさとは言っていない。つまり、正確に言うとする、ボロがでそうなので、ごまかしているのだろう。

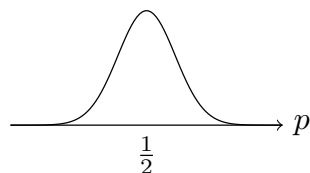
B 君 では、どうせ仮定なのだから、同じ確かさ、つまり、 $\frac{1}{2}$ と言ってしまえばどうか。

A 君 そうなると、おかしくなってくる。というのは、正確に $\frac{1}{2}$ だとすると、表も裏も区別なく模様も同じでないといけない、そうなるとそもそも表と裏を区別が意味なくなる。

B 君 ならば、 $\frac{1}{2}$ に近いであろうということは認め、あとは実際に実験して結果を考えてはどうか。

33. 硬貨を n 回投げたとき、 k 回表の面が出たとしよう。この硬貨を続けて投げるとき、表の出る確率を求める問題の設定では2通り解釈が可能である。😞

- (a) 表の出る確率は未知数であるが定まった値であるとする。この場合、過去の観測結果と関係ない未知数であるが、ある決まった値であると考え、 n が大きいとすると、 $P = \frac{k}{n}$ は $\frac{1}{2}$ に近づくものと期待してよいだろう。Rなどでプログラムを作成し、上図のように、確率が $n \rightarrow \infty$ で $\frac{1}{2}$ に近づく様子を確認よ。



- (b) B君の提案を具現化しよう。表の出る確率は、 P を実験の結果から決まる確率変数考える。つまり、固定しないで、実験を行った後で決まると考える。後述する条件付き確率（ベイズの定理）を用いて妥当な確率の値を求めよ。

◆6 幾何学的確率

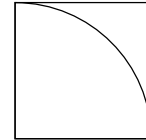
面積を用いた定義がある。たとえば、面積 A の領域 D_A のダーツを考える。その領域内に面積 B ($0 \leq B \leq A$) の小さな領域 D_B の的がある。いま、領域 D_A の中に入るように何度も矢を射る。各領域に当たる矢の数はそれぞれの面積に比例するものと考えられるので、矢がダーツの的に入る確率は、 D_B の D_A に対する面積比として

$$P = \frac{B}{A}$$

で表せるだろう. D_B の半径が D_A の $\frac{1}{2}$ の場合は $B = \frac{A}{4}$ で, $P = \frac{A/4}{A} = \frac{1}{4}$ となる.



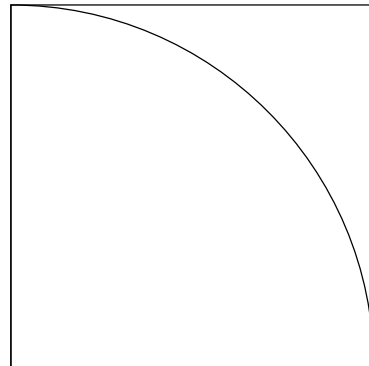
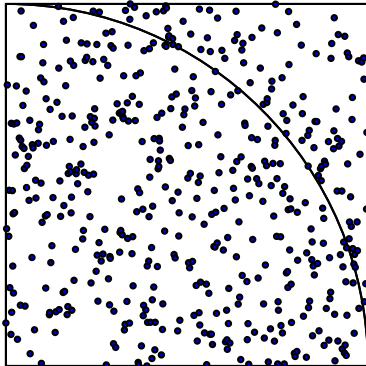
図のように正方形の中に正方形に接するように四分円 (半径を 1 とする) がある. 正方形の中の点を無作為に選んだとき, 四分円の中に入る確率を求めよ.



◇ 面積の比から, $\frac{\pi 1^2/4}{1^2} = \frac{\pi}{4}$. この問題は以下のような観点から, 積分を近似的に求める方法 (モンテカルロ法) として知られている. 四分円の面積は積分を用いて表すと

$$\int_0^1 \sqrt{1-x^2} dx = \frac{\pi}{4} = 0.785398\dots$$

となる^{*1}. つまり, 確率を用いて積分を求めることが出来る. 以下は点数を変えて積分を数値的に求めた例である. プログラム^{*2} は簡単であるが, できなければ下図に点を打って数えてやってみよう.



^{*1} sagemath では `integral(sqrt(1-x^2), x, 0, 1)` と入力するだけで積分できる

^{*2} 図は texlive で描画し, 計算は Python を用いた

```
import random; random.seed(1234);
def monte(num):
    counter = 0
    for i in range(NUM):
        x=random.uniform(0,1);y=random.uniform(0,1)
        if x*x+y*y < 1.0:
            counter +=1
    return counter
NUM=10000
pi = 4.0*monte(NUM)/NUM; print(pi)
```

全点数	四分円内点数	比 (確率)
100	83	3.320
1000	784	3.184
10000	7846	3.1588
100000	78535	3.14708

全点数	四分円内点数	比 (確率)

34. 棒をランダムに折って 2 本の断片を作るとき、長いほうの長さが短い方の 3 倍以上ある確率を求めよ.

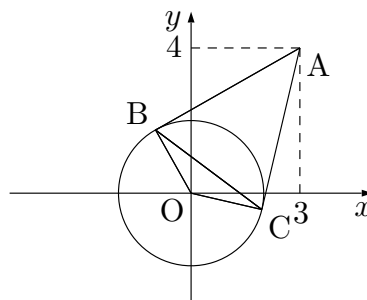
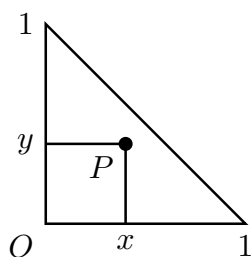
☞ 長い方の長さが短い方の長さを比較せよ.

35. 長さ 1 の線分上に 2 点を無作為にとり、線分を 3 つの線分に分ける. この 3 つの線分で三角形が作成できる確率を求めよ.

☞ 各線分の長さは $\min(x, y)$ によってきまる. 3 つの線分が三角形をなすためには、長さがすべて $\frac{1}{2}$ より小さければよい.

36. 以下の確率を求めよ.

- (a) 直角二等辺三角形の内部に 1 点 P を無作為に選び、直角をはさむ 2 本の等辺に P から垂線を下ろして長方形を作る. この長方形の面積が直角二等辺三角形の面積の $\frac{1}{3}$ 以上になる確率
- (b) 点 $A(3, 4)$ から円 $x^2 + y^2 = 2^2$ に引いた接線ある. 図の三角形 ABC の中で、無作為に選ばれた点が囲まれた ABC で囲まれた部分に落ちる確率



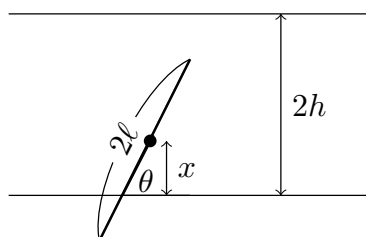
37. 区間 $[0, 1]$ から無作為に 2 点 x, y をとるとき、両点間の距離が $\frac{1}{2}$ 以上になる確率を求めよ.

38. ある 2 人は午後 1 時から午後 2 時の間に 10 分間だけ公園に立ち寄るのが日課である. ただし、公園に到着する時刻はお互いに無関係であるとする. この 2 人が公園で会う確率を求めよ.

☞ 2 人の午後 1 時からの到着時刻 (分) は $[0, 50]$ 上の一様分布にしたがう.

39. 【ビュホンの針, 18 世紀】 長さ 2ℓ の針が間隔 $2h$ ($\ell < h$ とする) の平行線のひとつと交わる確率を求めよ.

☞ 平行線が針と交わる条件は $x < l \cos(\theta)$

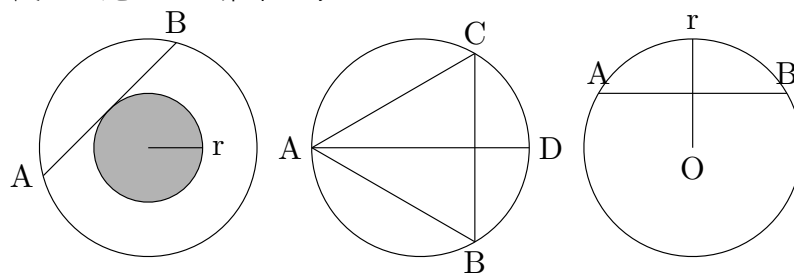


◆7 無作為の意味を考える.

確率の問題では, しばしば無作為に選択すると言った表現をする. 無作為の考え方によっては常に同じ確率を与えるのではなく, 場合によっては異なる値になることもあるので十分注意する必要がある. 以下に, 統計的定義には曖昧さが残る有名な例を用い考えよう.

40. 【ベルトランのパラドックス】 以下の図を参考にして問いに答えよ.

- 面積的に捉える. 弦 AB の中心が半径 $r/2$ の円の中にあれば, $\overline{AB} > \sqrt{3}r$ になる. . これから, 弦 AB の中心が円 C 中の面積 $\pi r^2/4$ から無作為に選択されればよい. 面積の比として確率を求めよ.
- 円周の長さとして捉える. 円 C の円周の一点を固定して, それを弦 AB の端 A とする. 端 B は中心円 O から見る角度が $2\pi/3$ 以上で π 以下であれば $\overline{AB} > \sqrt{3}r$ になる. したがって, 端 B が円周の長さ πr の内, $2\pi r/3$ の中に入るように無作為に選択されればよい. 比として確率を求めよ.
- 円の直径として捉える. 弦 AB をある固定された直径に平行になるように移動する. すると, 円の直径に, 沿った線の長さの比として確率を求められる. 長さの比として確率を求めよ.



このように, 無作為には曖昧さがあり, その解釈によっては確率が異なる場合がある

◆8 公理的確率

コルモゴロフの定義について調べよ. 事象 A に対して数値 $P(A)$ を以下の条件を満たすものを確率とした. (i) 以外の条件を調べよ. ☹️

- $0 \leq P(A) \leq 1$, A は任意の事象

(ii)

(iii)

なお、この公理は具体的な確率を定義するものではないが、幾何学的確率などはこの公理に当てはまる。

§2 確率の基本法則

確率の基本はしっかり身につけよう。

◆1 補（余）事象，加法定理 排反

補事象 $\bar{A}$: $P(\bar{A}) = P(\Omega - A) = 1 - P(A)$

差事象 $A - B$: $P(A - B) = P(A) - P(A \cap B)$

事象 A, B に対する **加法定理** :

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

特に排反なら、 $P(A \cap B) = 0$ なので、 $P(A \cup B) = P(A) + P(B)$

事象 A, B, C に対する **加法定理** :

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

特に排反なら、 $P(A \cup B \cup C) = P(A) + P(B) + P(C)$

ド・モルガンの法則 : $P(\overline{A \cap B}) = P(\bar{A} \cup \bar{B})$, $P(\overline{A \cup B}) = P(\bar{A} \cap \bar{B})$



(a) 1 から 20 までの整数の中から、無作為に取り出した整数が 6 または 8 の倍数である確率を求めよ。

(b) 1 から 200 までの整数の中から、無作為に取り出した整数が 6 または 8 の倍数である確率を求めよ。

◇ (a) 標本空間は $\Omega = \{1, 2, \dots, 20\}$ で、 $|\Omega| = 20$ 、6 の倍数の事象を A 、8 の倍数の事象を B とすると、 $A = \{6, 12, 18\}$ 、 $B = \{8, 16\}$ 、 $A \cup B = \{6, 8, 12, 16, 18\}$ で、 $P(A \cup B) = \frac{|A \cup B|}{|\Omega|} = \frac{5}{20} = \frac{1}{4}$ 。あるいは、 A と B は排反で $A \cap B = \emptyset$ で、 $P(A \cup B) = P(A) + P(B) = \frac{3}{20} + \frac{2}{20} = \frac{5}{20} = \frac{1}{4}$ とすることもできる。(b) $\Omega = \{1, 2, \dots, 200\}$ で、 $|\Omega| = 200$ 、 $|A| = 33$ 、 $|B| = 25$ 、 $|A \cap B| = 8$ から、 $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{33 + 25 - 8}{200} = \frac{50}{200} = \frac{1}{4}$



赤玉 5 個，白玉 3 個，黒玉 1 個が入っている袋から，玉を同時に 4 個取り出すとき，赤玉が 2 個以上である確率を求めよ。

◇ 【2019 広島県高】 赤玉が 0 個，1 個の場合を除けば， $1 - \left(\frac{1}{9C_4} + \frac{5C_1 \cdot 4C_3}{9C_4} \right) =$
 $1 - \left(\frac{1}{126} + \frac{20}{125} \right) = \frac{5}{6}$

41. $P(A) = \frac{1}{3}$, $P(B) = \frac{1}{2}$, $P(A \cup B) = \frac{2}{3}$ とする. このとき, $P(A \cap B)$, $P(A \cap \bar{B})$, $P(A - B)$, $P(A|B)$, $P(A|\bar{B})$ を求めよ.

☞ 確率の基本法則（条件付き確率は後で）を，ベン図も参考にして，十分理解しておこう.

42. 52 枚のトランプから無作為の 1 枚のカードを取り出す. 絵札またはスペードである確率を求めよ.

☞ 絵札は何枚あったかな.

43. 10 から 99 の書かれた 90 枚のカードの中から 1 枚を取り出すとき，そのカードに書かれた数字が 4 または 6 で割り切れる確率を求めよ.

☞ 加法定理の応用.

44. 1 から 100 の書かれた 100 枚のカードの中から 1 枚を取り出すとき，次の問いに答えよ.

(a) カードに書かれた数字が 2 または 3 で割り切れる確率を求めよ.

(b) カードに書かれた数字が 2 でも 3 でも割り切れない確率を求めよ.

45. 5 人の男子と 3 人の女子を無作為に一列に並べるとき，隣り合う女子が存在する確率を求めよ.

☞ 余事象は女子の隣り合わない場合であるが，先に男子を並べて考えよう.

46. カードの和，積 1 から 8 までの番号のついた 8 枚のカードがある. この中から同時に 3 枚のカードを取り出すとき，次の問いに答えよ.

(a) 3 枚のカードに書かれた数の和が 10 以下になる確率を求めよ.

(b) 3 枚のカードに書かれた数の積が 2 の倍数になる確率を求めよ.

(c) 3 枚のカードに書かれた数の積が 4 の倍数になる確率を求めよ.

47. 再度さいころ

(a) サイコロを何回か振る実験を行う. 何回さいころを振れば，少なくとも 1 回 が出る確率が $1/2$ 以上になるだろうか.

(b) n 個のさいころの出た目の積が 6 の倍数になる確率を求めよ.

(c) (b) の確率が 0.8 を超える n を求めよ.

☞ 余事象を用いよう.

48. 3個のさいころを投げるとき、次の確率を求めよ.

- (a) 出た目の積が偶数
- (b) 出た目の積が4の倍数

☞ 余事象を用いよう

49. さいころをくり返し n 回振って、出た目の数を掛け合わせた積を X とする. すなわち、 k 回目に出た目の数を Y_k とすると、 $X = Y_1 Y_2 \cdots Y_n$

- (a) X が3で割り切れる確率を求めよ.
- (b) X が4で割り切れる確率を求めよ.
- (c) X が6で割り切れる確率を求めよ.

50. 1から9までの番号を付けた9枚のカードがある. この中から無作為に4枚のカードを同時に取り出し、カードに書かれた4つの番号の積を X とおく.

- (a) X が5の倍数になる確率を求めよ.
- (b) X が10の倍数になる確率を求めよ.
- (c) X が6の倍数になる確率を求めよ.

51. じゃんけん

(a) 4人でじゃんけんをして、勝者1人を決めたい. 1回目のじゃんけんで1人に決まらなかった場合には、負けた人を除いて2回目のじゃんけんをする【以前の問題参考】

- (i) 1回目のじゃんけんで、勝者1人が決まる確率を求めよ.
- (ii) 1回目のじゃんけんで、あいこになる確率を求めよ.
- (iii) ちょうど2回目のじゃんけんで勝者1人が決まる確率を求めよ.

(b) n を2以上の整数とする. n 人のじゃんけんであいこになる確率を求めよ.

(c) 3人がじゃんけんをして、ちょうど n 回目で1位から3位までの順位が決まる確率 P_n を求めよ. ただし、3人の手の出し方は、どれも同様に確からしいとする.

52. 誕生日のパラドクス (birthday paradox) とは「何人集まれば、その中に誕生日が同一の2人(以上)がいる確率が50%を超えるか?」という問題から生じるパラドックスで、結果が一般的な直感と反していると思うという意味でのパラドックスである (Wikipedia 参照). 1年を365日とする.

(a) n 人のあるクラスに自分と同じ誕生日の人がいる確率を求めよ. また、 $\frac{1}{365}$ は

小さいので、確率は $\frac{n-1}{365}$ と近似できることを示せ.

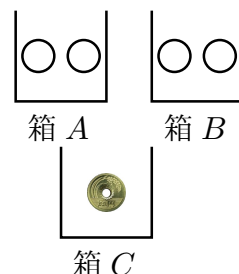
(b) n 人のあるクラスに同じ誕生日の人がいる確率を求めよ. また、 $n = 20$ のとき、確率はいくらになるか.

☞ 余事象を用いよ.

以下は少々難しい問題



53. コインが5枚ある. さいころを振って出た目によって, これらのコインを1枚ずつ3つの箱 A, B, C のいずれかに入れていく. 出た目が1であればコインを1枚, 箱 A に入れる. 出た目が2か3であればコインを1枚, 箱 B に入れる. 出た目が4か5か6であればコインを1枚, 箱 C に入れる. さいころを5回振ったとき, 次の問いに答えよ.



- (a) 箱 A と箱 B にコインがそれぞれちょうど2枚ずつ入っている確率を求めよ.
 (b) A, B いずれの箱にもコインが1枚以上入っている確率を求めよ.
54. 【加法定理の一般化】
- (a) 3つの事象 A, B, C に対する加法定理, $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C)$ が成り立つことを証明せよ.
 (b) $A_i (1 \leq i \leq n)$ に対する加法定理を導け.
55. 【ド・モンモールの問題, 1708年】 $n (\geq 2)$ 本の手紙とその宛て先を書いた n 個の封筒がある. 今, n 本の手紙をでたらめにいずれかの封筒に入れる.
- (a) 少なくとも一つの手紙と宛て先を書いた封筒が一致する確率を $g_1(n)$ とする.
 (i) $g_1(3)$ を求めよ.
 (ii) $g_1(n)$ を n を用いて表せ.
 (iii) $n \rightarrow \infty$ とした場合の $g_1(\infty)$ を求めよ.
 ➡ 加法定理 (上問の (a), (b)) を利用しよう.
- (b) 同じ問題で, (a) の余事象の確率を計算する. 今, 手紙と宛て先を書いた封筒のどれもが一致しない総数を $D(n)$, その確率を $g_2(n)$ とする.
 (i) $n \geq 4$ に対して $D(n) = (n-1)\{D(n-2) + D(n-1)\}$ が成り立つことを証明せよ.
 (ii) $g_2(n)$ を n を用いて表せ. また, $g_1(n) + g_2(n) = 1$ を確かめよ.
 (iii) $n \rightarrow \infty$ とした場合の $g_2(\infty)$ を求めよ. また, $g_1(\infty) + g_2(\infty) = 1$ を確かめよ.
 ➡ (a) と (b) は余事象の関係にあるが, どちらからアプローチしても難しい.

◆2 条件付き確率とベイズの定理

条件付き確率: $P(B|A) = \frac{P(A \cap B)}{P(A)}$

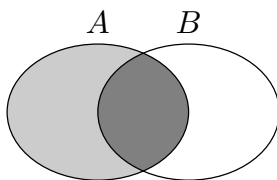
乗法定理: $P(A \cap B) = P(B|A)P(A) = P(A|B)P(B)$

分配則: $P(A) = P(A \cap B) + P(A \cap \bar{B}) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$

ベイズの定理 : $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\bar{B})P(\bar{B})}$

意義 : 事前確率 : $P(B)$, 事後確率 : $P(B|A)$

観測 A によって, 事象 B の確率 $P(B)$ が $P(B|A)$ へと変化



1 から 10 までの番号をつけた 10 枚のカードから 1 枚のカードを取り出す試行において, 取り出したカードの番号が奇数であることがわかっているとき, その番号が 3 の倍数である確率を求めよ.

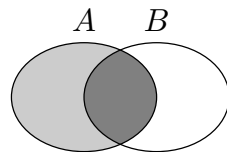
◇ A : 奇数である事象 (薄い灰色), B : 3 の倍数である事象,

$A \cap B$: 奇数でかつ 3 の倍数である事象 (濃い灰色) とする.

$A = \{1, 3, 5, 7, 9\}$, $B = \{3, 6, 9\}$, $A \cap B = \{3, 9\}$ から,

$$P(A) = \frac{5}{10}, P(A \cap B) = \frac{2}{10}. \text{ 条件付き確率から,}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{2/10}{5/10} = \frac{2}{5}$$



ある疾病にかかる確率を p ($0 < p < 1$) とする. いかなる検査も完全ではない. ある検査で疾病にかかっている人をポジティブと判定する確率を 95%, 疾病でない人を誤ってポジティブと判定する確率が 10% であったとする. ポジティブと判定された人が実際に疾病にかかっている確率を求めよ.

👉 条件付き確率とベイズの定理の意味を考えよ.

◇ P をポジティブと判定された事象, D を疾病にかかる事象とする. この例題は $P(D|P)$ を求めること.

事前確率 $P(D) = p$: 予め分かっている確率

事後確率 $P(D|P)$: ポジティブと判断されたときに実際に病気である確率

条件付き確率 :

$$P(P|D) = 0.95 \text{ (病気の人をポジティブと判定} \Leftarrow \text{検査精度)}$$

$$P(P|\bar{D}) = 0.1 \text{ (病気でない人をポジティブと判定} \Leftarrow \text{これも検査精度)}$$

事前確率として $P(D) = p$ が与えられているので, 余事象から $P(\bar{D}) = 1 - p$

ベイズの定理から、ポジティブと判断されたときに実際に病気である確率は、

$$P(D|P) = \frac{P(P \cap D)}{P(P)} = \frac{P(P|D)P(D)}{P(P|D)P(D) + P(P|\bar{D})P(\bar{D})} = \frac{0.95p}{0.95p + 0.1(1-p)}$$

となる．たとえば $p = 0.5$ とすると $P(D|R) = \frac{16}{21} \approx 0.905$ となるので，検査結果は信頼できる．しかし， $p \ll 1$ と小さいと， $P(D|R) \simeq 0.095p \ll 1$ となるので，検査結果は信頼できない．

56. 1 から 12 までの番号をつけた 12 枚のカードから 1 枚のカードを取り出す試行において，取り出したカードの番号が 3 の倍数である事象を A ，番号が一桁である事象を B とする．3 の倍数であることがわかっているとき，その番号が一桁である事象に関して，以下の空欄を埋めよ．

$$\begin{aligned} |\Omega| &= \boxed{}, & |A| &= \boxed{}, & |B \cap A| &= \boxed{} \\ P(A) &= \frac{|A|}{|\Omega|} = \boxed{}, & P(B \cap A) &= \frac{|B \cap A|}{|\Omega|} = \boxed{} \\ P(B|A) &= \frac{|B \cap A|}{|A|} = \frac{|B \cap A|/|\Omega|}{|A|/|\Omega|} = \frac{P(B \cap A)}{P(A)} = \boxed{} \end{aligned}$$

57. 条件付き確率の計算に慣れよう 以下の問いに答えよ．

- (a) $P(A) = \frac{1}{8}$, $P(B) = \frac{1}{7}$, $P(A \cap B) = \frac{1}{10}$ のとき， $P(B|A)$ を求めよ．
 (b) $P(A) = \frac{1}{8}$, $P(B) = \frac{1}{8}$ かつ A, B が独立のとき， $P(A \cup B)$ を求めよ．
 (c) $P(E) = \frac{1}{3}$, $P(F) = \frac{1}{2}$, $P(E \cup F) = \frac{2}{3}$ のとき， $P(\bar{E})$, $P(E \cup \bar{F})$, $P(\overline{E \cup \bar{F}})$, $P(E|F)$, $P(E|\bar{F})$, $P(E \cap F|E \cup F)$ を求めよ．

☞ 余事象，加法定理，条件付き確率を組合わせて計算．ベン図を利用しよう．

58. ベイズの定理（条件付き確率）の基本 赤玉 5 個と白玉 7 個が入っている袋から，玉を 2 個取り出すとき，次の問いに答えよ．

- (a) 同時に取り出すとき，玉の色が異なる確率を求めよ．
 (b) 1 個目を取り出し袋に戻さないで 2 回目を取り出す．2 個目に取り出した玉が白玉であったとき，1 回目に取り出した玉も白玉である確率を求めよ．

☞ (a) 同時に取り出さなくても，1 個目を取り出し，元に戻さないで 2 回目を取り出しても確率は同じ．

59. 白玉 7 個，赤玉 3 個が入っている袋から，玉を 1 個取りだし，それを袋に戻さないで，続けてもう 1 個取り出す．2 個目に取り出した玉が赤玉であるとき，最初に取り出した玉も赤玉である確率を求めよ．

☞ 簡単な絵を書き，題意を確認しながら進めよう．

60. 3 つの箱 A , B , C があり， A には黒球 3 個と白球 2 個， B には黒球 1 個と白球 5

個, C には黒球 2 個と白球 2 個が入っている. 3 つの箱から 1 つの箱を選び, 選んだ箱から球を 1 つ取り出す. 取り出した球が黒球であるとき, 選んだ箱が A である確率を求めよ.

☞ どの箱を選ぶかは $1/3$ である. 3 つの事象に関する分配則 $P(X) = P(A \cap X) + P(B \cap X) + P(C \cap X)$.

61. 袋の中に, 1, 2, 3, 4, 5, 6, 7, 8 の数字が 1 つずつ書かれた 8 枚のカードが入っている. この袋の中から, 最初に 1 枚のカードを取り出し, 取り出したカードに書かれた数を x とする. このカードを袋の中に戻さずに, さらに袋の中から 2 枚のカードを取り出し, 取り出した 2 枚のカードそれぞれに書かれた数の和を y とする.

(a) 積 xy が偶数である確率を求めよ,

(b) 積 xy が偶数であったとき, y が偶数である条件付き確率を求めよ.

☞ (a) 余事象は xy が奇数つまり x, y はともに奇数. (b) y が偶数になるのは (奇数 + 偶数) 以外の場合で, これも余事象.

62. 袋 A には赤玉 2 個と白玉 5 個, 袋 B には赤玉 2 個が入っている. まず, 袋 A から 3 個の玉を同時に取り出し, 玉の色は確認せず, そのまま袋 B に入れ, よくかき混ぜて, 袋 B から 2 個の玉を同時に取り出す. 次の問いに答えよ.

(a) 袋 A から取り出された 3 個の玉が, 赤玉 1 個と白玉 2 個である確率, 白玉 3 個である確率をそれぞれ求めよ.

(b) 袋 B から取り出された玉が 2 個とも白玉である確率を求めよ.

(c) 袋 B から取り出された玉が 2 個とも白玉であったとき, 袋 B に白玉が残っている条件付き確率を求めよ.

☞ (b) は (a) の結果を利用. (c) 袋 B から取り出された玉が 2 個とも白玉で, 袋 B に白玉が残っているのは袋 A から取り出された 3 個の玉が白玉 3 個のとき.

63. 1000 人に 1 人がかかるという病気がある. ある検査薬では陽性, 陰性ともに 99% の精度で判定できる. ある人がこの検査薬を用いたところ陽性と診断された. ある人が感染している確率を求めよ.

☞ 陽性と診断された事象 P , 感染している事象を B とすると, 求めたい条件付き確率は $P(B|P) = P(P \cap B)/P(P)$. ここで, $P(P) = P(P \cap B) + P(P \cap \bar{B}) = P(P|B)P(B) + P(P|\bar{B})P(\bar{B})$

64. 機械 A は 2% の確率で不良品を作る. 機械 B は 6% の確率で不良品を作る. 機械 C は 1.5% の確率で不良品を作る. A, B, C がそれぞれ 4800 個, 1200 個, 4000 個製品を作った. 10000 個のうち 1 個サンプルを取り出したら不良品であった. それを作ったのは A である確率を求めよ.

65. 箱が 4 つあり, 第 1 の箱には 2000 個の部品があり, そのうち 5% が不良品であった. 第 2 の箱には 500 個の部品があり, そのうち 40% が不良品であった. 第 3, 4

の箱には、それぞれ 1000 個の部品があり、そのうち 10% が不良品であった。

- (a) 上記の箱を 1 つ無作為に選び、無作為に部品を 1 個選ぶ。選ばれた部品の不良品である確率はいくらか。
- (b) つぎに、選んだ部品が不良品と分かったとする。その不良品が第 2 の箱から取り出される確率を求めよ。
66. 【保険会社】 保険契約者が 25 歳未満のとき 1 年間に事故にあう確率は a で、25 歳以上のときは b であることを知っている。被保険者の内、25 歳未満の人の割合を c とする。
- (a) ランダムに選ばれた保険契約者が事故にあう確率を求めよ。
- (b) 事故にあった被保険者が 25 歳未満の確率を求めよ。

➡ (a) 保険契約者が事故にあう事象 A の確率は $P(A) = P(A|25 \text{ 歳未満})P(25 \text{ 歳未満}) + P(A|25 \text{ 歳以上})P(25 \text{ 歳以上})$ 。(b) $P(25 \text{ 歳未満} | A) = P(A|25 \text{ 歳未満})/P(A)$

以下は少々難しい問題



67. 【CRAPS ゲーム】 さいころを 2 つを使う。最初に振って和が 2 か 3 か 12 ならば負け、7 あるいは 11 ならば勝ち、4 ならば和が 4 か 7 になるまで振り直し、最後に 4 が出たら勝ち、7 が出たら負け。5, 6, 8, 9, 10 の場合は 4 と同じように最初に出た目か 7 が出るまで振り続け、最後に 7 が出たら負け、さもなければ勝つ。このゲームに勝つ事象 A の確率 $P(A)$ を求めよ。また、なぜこのゲームが面白いのか、結果から考えよう。

➡ $P(A) = \sum_{k=4,5,6,7,8,9,10,11} P(A \cap X = k) = \sum P(A|X = k)P(X = k)$

68. 1 から 10 の番号が付いている 10 枚のチケットがある。このうち 1 番と 2 番が当たりくじとなっている。A 君は 4 枚のチケットを買った。
- (a) A 君は 1 番をもっていると告げた。このとき、残りの 6 枚にあたりが残っている確率を求めよ。
- (b) A 君は少なくとも 1 枚のあたりをもっていると告げた。このとき、残りの 6 枚にあたりが残っている確率を求めよ。

➡ (b) では少なくとも 1 枚持っていると言ってるのだから 2 枚持っている可能性もあり、残り 6 枚にはあたりがない可能性が大きいと思う。結果は、(a) の確率 < (b) の確率で、直感にあわない!

69. 【選択問題の意義】 選択技が c ($c > 1$) 通りある問題がある。受験者は確率 p で正答できるとする。分らなければ確率 $1 - p$ で無作為に選択して回答するだろう。図のように司法試験（短答式）では記述 5 個から正しい記述を 2 個選ぶ問題が課されるので、選択枝としては $c = {}_5C_2 = 10$ 通りある。

- (a) 受験者が正答したとき、その受験者が問題を真に理解していた確率を 0.8 以上にしたいとすれば、 p の値はいくらなるか。
- (b) もし選択した解答が正解ならば、その受験者がその問題を真に理解していた確率を 0.8 以上にするためには、選択肢 c をいくらにすればいいだろうか。

【第4問】
 選択の自由に関する次のアからイまでの各記述について、最高裁判所の判例の趣旨に照らして、正しいものには○、誤っているものには×を付した解答の組合せを、設問1から8までの中から選びなさい。【解答欄】
 ア、作問時既に正解の選択肢を正確に把握する者は、他の選択肢を正解と信じて心理的圧迫を受けることになるので、これにより自己の心算ないし素朴上の感情が害され不快の念を抱いた者は、国の素朴活動に就いた憲法第21条第1項の憲法上の権利の侵害を受けることを理由として国に損害賠償を請求することができる。
 イ、憲法第21条第1項の憲法上の権利の侵害を受ける者として国に損害賠償を請求するものは、自己の権利を侵害しない権利を侵害する者に対しては憲法上の権利を侵害するものとして損害賠償を受けるものではない。
 ウ、憲法第21条第1項の憲法上の権利の侵害を受ける者として国に損害賠償を請求するものは、自己の権利を侵害しない権利を侵害する者に対しては憲法上の権利を侵害するものとして損害賠償を受けるものではない。
 エ、憲法第21条第1項の憲法上の権利の侵害を受ける者として国に損害賠償を請求するものは、自己の権利を侵害しない権利を侵害する者に対しては憲法上の権利を侵害するものとして損害賠償を受けるものではない。
 オ、憲法第21条第1項の憲法上の権利の侵害を受ける者として国に損害賠償を請求するものは、自己の権利を侵害しない権利を侵害する者に対しては憲法上の権利を侵害するものとして損害賠償を受けるものではない。
 カ、憲法第21条第1項の憲法上の権利の侵害を受ける者として国に損害賠償を請求するものは、自己の権利を侵害しない権利を侵害する者に対しては憲法上の権利を侵害するものとして損害賠償を受けるものではない。
 キ、憲法第21条第1項の憲法上の権利の侵害を受ける者として国に損害賠償を請求するものは、自己の権利を侵害しない権利を侵害する者に対しては憲法上の権利を侵害するものとして損害賠償を受けるものではない。
 ク、憲法第21条第1項の憲法上の権利の侵害を受ける者として国に損害賠償を請求するものは、自己の権利を侵害しない権利を侵害する者に対しては憲法上の権利を侵害するものとして損害賠償を受けるものではない。
 ケ、憲法第21条第1項の憲法上の権利の侵害を受ける者として国に損害賠償を請求するものは、自己の権利を侵害しない権利を侵害する者に対しては憲法上の権利を侵害するものとして損害賠償を受けるものではない。
 コ、憲法第21条第1項の憲法上の権利の侵害を受ける者として国に損害賠償を請求するものは、自己の権利を侵害しない権利を侵害する者に対しては憲法上の権利を侵害するものとして損害賠償を受けるものではない。
 1. ア○イ○ウ○ 2. ア○イ○ウ× 3. ア○イ×ウ○
 4. ア○イ×ウ× 5. ア×イ○ウ○ 6. ア×イ×ウ×
 7. ア×イ×ウ○ 8. ア×イ×ウ×

☞ A を問題が正確に解答された事象、S をその学生が正答を理解していた事象とする。問題は $P(S|A)$ を求めること。

70. 【3 囚人の問題, 1959 年】3 人の囚人 A, B, C がいる。3 人とも処刑されること



になっていたが、王子が結婚するというので王様が 1 人だけ恩赦にしてやることになった。誰が恩赦になるのか決定されたが、まだ囚人たちには知らされていない。結果を知っている看守に、囚人 A が「B と C のうち、どちらかは必ず処刑されるのだから、処刑される 1 人の名前を教えてください、私に情報を与えることにはならないだろう。1 人を教えてください」と頼んだ。看守は、その言い分に納得して「囚人 B は処刑されるよ」と教えてやった。囚人 A は、「はじめ自分の助かる確率は $1/3$ だった。今、助かるのは自分と C のどちらかになったので、助かる確率は $1/2$ になった」と喜んだ。さて、看守の返事を聞いた後の囚人 A の助かる確率はどれだけか？ ただし、王様は意図的に恩赦にする者を決めているわけではない（等確率で恩赦にする者を選んでいる）こと。看守はうそをつかないこと。囚人 B, C がともに処刑される場合には $1/2$ ずつの確率でどちらかの名前を言うものとする。

☞ B が処刑されるということを聞いたという条件のもとで、A が助かる条件付き確率を求める問題。

71. 【モンティ・ホールの問題】先にあまり深入りすることはなかったが、ここで、条件付き確率の問題としてもう少し詳しく調べよう。

中が見えない部屋 A, B, C のいずれか一つに車が入っている。ゲストは車が入っている部屋を言い当てることができればその車がもらえる。司会者は、ゲストが部屋を一つ指定した後、指定された部屋以外で、車が入っていない部屋を空け、「部屋をかえますか？」とゲストに尋ねる。ゲストは部屋をかえた方がいいだろう

か？

ゲストが指定した部屋は A（何でもいいが）とし、司会者は車が部屋 B にあるときは部屋 C を空け、車が部屋 C にあるときは部屋 B を空ける。車が部屋 A にあるときは、司会者は確率 p ($0 \leq p \leq 1$) で部屋 B を空け、確率 $1-p$ で部屋 C を空けるとしよう。

72. 【クイズ番組の優勝者チャレンジコーナー】 ガレージが 4 つあり、ドアがすべて閉まっている。4 つのガレージのうち、ひとつには商品の外車が入っているが、残りはすべて空で、「外れ」のガレージである。挑戦者はひとつのガレージを選ぶ。すると、挑戦者が選んだガレージ以外に最低でも 2 つのガレージは空のはずである。空のガレージがどれかを前もって知っている司会者は空のガレージのうちの 1 つを任意にあけて、挑戦者に選んだガレージの変更を迫る。変更を迫られた挑戦者はドアが閉まっている他のガレージに変更してもよいし、司会者に変更しないことを伝えてもよい。いずれにせよ、挑戦者が最後に選んだガレージに外車が入っていたら、その外車を挑戦者が獲得できる。以下の問いに答えよ。

- (a) 司会がガレージを開けたにもかかわらず、選んだガレージの変更をしなかった場合に挑戦者が外車を獲得できる確率はいくらか？
- (b) 司会がガレージを開けるのを見て、選んだガレージの変更をした場合に挑戦者が外車を獲得できる確率はいくらか？

◆3 独立、反復試行

二つの事象 A と B が関係なく生起するならば、 B に関係なく A が決まる。条件として B でも $\bar{B}$ でも、条件付確率は同じ、つまり $P(A|B) = P(A|\bar{B})$ となる。これを分配則 $P(A) = P(A|B)P(B) + P(A|\bar{B})P(\bar{B})$ に代入すると、

$$P(A) = P(A|B)(P(B) + P(\bar{B})) = P(A|B)P(\Omega) = P(A|B)$$

となる。逆に、 $P(A) = P(A|B)$ が成り立てば、事象 A と B は関係なく起る。さらに、これを乗法定理 $P(A \cap B) = P(A|B)P(B)$ に代入すると、

$$P(A \cap B) = P(A)P(B)$$

が成り立つ。この関係が成り立つ事象 A, B は、確率的に**独立**であると言う。独立と排斥の違いには注意しよう。同様に、任意の事象 $A_1, A_2, \dots, A_n$ に対する独立性は

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2) \dots P(A_n)$$

と表せる。



A, B の 2 人がテニスの試合をする。A が B に勝つ確率は p とする。先に 2 勝した方を優勝とすると、3 試合目までに A が優勝する確率を求めよ。

◇○○, ○●○ (○: A 勝ち, ●: B が勝ち, 箱の中は入れ替わる場合も含める) の場合である. A が優勝する確率は $\left(\frac{1}{2}\right)^2 + {}_2C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{2-1} \times \frac{1}{2} = \frac{1}{4} + 2 \times \frac{1}{8} = \frac{1}{2}$

73. 2つの事象 A , B が独立なとき, 以下のことを示せ.

- (a) A と $\bar{B}$ は独立
- (b) $\bar{A}$ と $\bar{B}$ は独立

74. 独立性 次の事象 A , B は独立かどうかを調べよ.

- (a) トランプ (52 枚) において「スペード (13 枚) が出る」という事象を A , 「1 のカード (4 枚) が出る」という事象を B とする. c
- (b) 1 から 15 までの番号を書いた 15 枚のカードの中から無作為に 1 枚抜き出す試行において, 抜き出したカードの番号が偶数である事象を A , 抜き出したカードの番号が 10 以下である事象を B とする.
- (c) n 本のくじのうち, あたりくじが m ($m < n$) 本入っている. 2 人がくじを引くとき, 最初の人引いたくじは戻さないとする. 最初に引いた人があたりを引く事象と, 後で引いた人があたりを引く事象

75. さいころを何回か振る実験を行う. 何回さいころを振れば, 少なくとも 1 回 ●● が出る確率が $\frac{1}{2}$ 以上になるだろうか.

76. 1 個のさいころを 3 回続けて投げるとき, 次の確率を求めよ.

- (a) 1 回目は偶数の目, 2 回目は 3 以下の目, 3 回目は 5 以上の目が出る確率
- (b) 少なくとも 1 回は偶数の目が出る確率

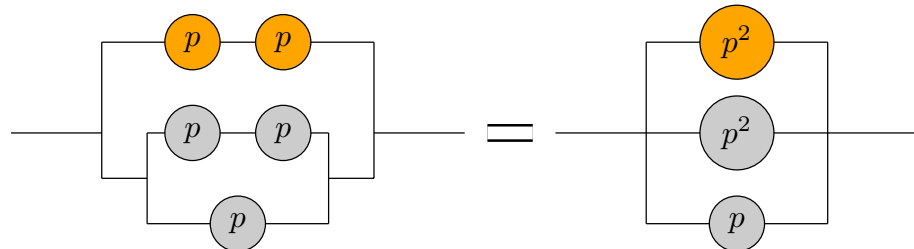
77. 試合の勝敗

- (a) A , B の 2 チームがテニスの試合をする. A が B に勝つ確率は $\frac{1}{3}$ で, 引き分けはないものとする. 先に 3 勝した方を優勝とすると, 4 試合目までに A が優勝する確率を求めよ.
- (b) A と B の試合で, A , B の勝つ確率がそれぞれ $\frac{1}{3}, \frac{2}{3}$ であるとする. この試合を繰り返すとき, A が B よりも先に 3 回勝つ確率を求めよ.
- (c) A , B の 2 人があるゲームをくり返し行う. 1 回のゲームで A が B に勝つ確率は $\frac{2}{3}$, B が A に勝つ確率は $\frac{1}{3}$ である. 先に 3 勝した方を優勝とする.
 - (i) B が優勝する確率を求めよ. 【問題 (b) と同じ】
 - (ii) B が優勝するとき, 3 回目に A が B に勝つ条件付き確率を求めよ.

78. 白玉 2 個, 赤玉 4 個が入っている袋から玉を 1 個取りだし, 色を調べてからもとに戻す. この試行を 6 回続けて行うとき, 次の確率を求めよ.

- (a) 白玉が 5 回以上出る確率

- (b) 6 回目に 3 度目の白玉が出る確率
79. 【先手必勝】 A, B, C の 3 人がこの順に 1 個の硬貨を投げ、最初に表を出したものを勝ちとしよう. それぞれの者が勝つ確率を求めよ.
- ☞ A が最初に勝つ確率は, $P(\text{表}) + P(\text{裏} \cap \text{裏} \cap \text{裏} \cap \text{表}) + \cdots = P(\text{表}) + P(\text{裏})P(\text{裏})P(\text{裏})P(\text{表}) + \cdots$. B が最初に勝つ確率は, $P(\text{裏} \cap \text{表}) + P(\text{裏} \cap \text{裏} \cap \text{裏} \cap \text{裏} \cap \text{表}) + \cdots = P(\text{裏})P(\text{表}) + P(\text{裏})P(\text{裏})P(\text{裏})P(\text{裏})P(\text{表}) + \cdots$. 級数の和を求める.
80. 【補助電源】 電力を供給するシステムを考える. 通常の電力供給源と, 非常用電力供給源が並列に並んでいて, 電力供給源には確率 p で動作する機器 2 個が直列に, 非常用電力供給源には, 確率 p で動作する機器 2 個が直列に, さらに同じ確で動作する機器が並列に置かれている.
- (a) 各機器が独立として, システムが動作する確率 $g(p)$ を p を用いて表せ.
- (b) 補助電源をもうけることで, 動作の確率がどのくらい向上するか考察せよ. この例では, 図の上も機器 2 個を直列につないだのがもとのシステムで, 動作する確率は p^2 である.
- ☞ 2 つの独立な機器の場合, それらが動作する事象を A, B とする. 2 個が直列につながっている場合両方が動作しなければならないので, $P(A \cap B) = P(A)P(B) = p^2$, 並列ならどちらかが動作すればよいので $P(A \cup B) = 1 - P(\overline{A \cup B}) = 1 - P(\bar{A} \cap \bar{B}) = 1 - (1 - p)^2$. 左図と右図の電力回路は等価.



以下は少々難しい問題



81. コインを 2 回投げて, 最初と 2 回目に投げたコインの表か裏かによって以下のような確率変数を定義する.

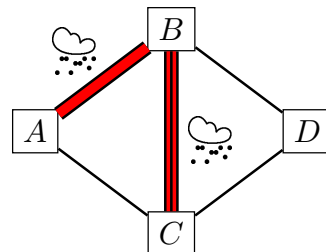
$$x = \begin{cases} 1 & : \text{最初のコインは表} \\ 0 & : \text{最初のコインは裏} \end{cases} \quad y = \begin{cases} 1 & : \text{2 回目のコインは表} \\ 0 & : \text{2 回目のコインは裏} \end{cases}$$

いま, $z = x + y(\text{mod}2)$ と定義する. よって, 標本空間は $\Omega = \{(x, y, z) | x, y \in \{0, 1\}, z = x + y(\text{mod}2)\}$ である. A を x が 1 となる事象 $A = \{\omega \in \Omega | x = 1\}$ とし, 同様に事象 B, C を $B = \{\omega \in \Omega | y = 1\}, C = \{\omega \in \Omega | z = 1\}$ とおく. このとき, 三つの事象 A, B, C は独立かどうかを調べよ.

☞ 標本空間は $\Omega = \{000, 011, 101, 110\}$. たとえば事象 A, B を考えると, $P(A \cap B) =$

$P(A)P(B)$ となるので独立だが, $P(A \cap B \cap C)$ はどうであろうか. どの2つの事象も独立であったとしても, 3つの事象は必ずしも独立とは言えない.

82. ある地域の天気予報によると雪の降る確率は p で, 雪のため図に示す各道路は独立に確率 p で封鎖される. 除雪の経費が利用でき, AB あるいは BC のどちらかを除雪することにした. どちらの道路を除雪した方が, A 町から D 町へドライブできる確率が高いかな.



☞ AB を除雪した場合: $A=B$ と考える. $A(B) \rightarrow D$ と $(ABC \text{ と } AC \text{ は並列}) \rightarrow D$ が並列になる. $P(A \rightarrow D) = 1 - P(\overline{A(B) \rightarrow D})P(\overline{(A(B)C \text{ と } AC \text{ は並列}) \rightarrow D})$. ここで, $P(\overline{A(B) \rightarrow D}) = p$, また $P(\overline{(A(B)C \text{ と } AC \text{ は並列}) \rightarrow D}) = 1 - P(A(B)C \text{ と } AC \text{ は並列})P(C \rightarrow D) = 1 - (1 - p^2)(1 - p)$. BC のを除雪した場合: $B=C$ と考える. $P(A \rightarrow D) = (1 - p^2)^2$. 複数回余事象を使うのでややこしい.

83. 【未完】

◆4 分配則による漸化式

最初に事象 A が生じるまで, 繰り返して投げる実験を行う. 最初事象 A が生じるまでに必要な回数を n ($n = 1, 2, 3, \dots$) とし, n 回目で始めて A が生じる事象を A_n とし, その確率 $p_n \equiv P(A_n)$ とする. 最初 $\bar{B} = \emptyset$ ならば, $P(A_n) = P(A_n|B)P(B)$ であるが, $P(A_n|B) = P(A_{n-1})$ なので, $p_n = p_{n-1}P(B)$ 一般的には,

$$\begin{aligned} P(A_n) &= P(A_n|B)P(B) + P(A_n|\bar{B})P(\bar{B}) \\ &= P(A_{n-1})P(B) + P(A_{n-1})P(\bar{B}) \\ p_n &= p_{n-1}P(B) + (1 - p_{n-1})P(\bar{B}) \end{aligned}$$



大小2つのさいころを同時に投げる. 大小のさいころの目が同じとなる場合がちょうど3回になるまで繰り返し行い, 終了するまでの回数が n 回 ($n \geq 3$) とする確率を P_n とする. 次の問いに答えよ.

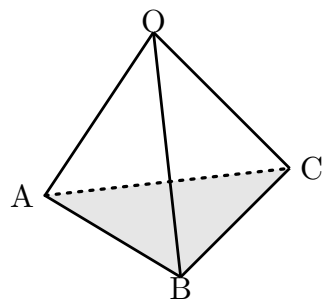
- P_5 を求めよ.
- $\frac{P_{n+1}}{P_n}$ を n を用いて表せ.
- P_n が最大となる n の値を求めよ.

◇【2020 愛知県中高他多数】 (a)1回の試行で, 大小2つのさいころの目が同じになる確率

は $\frac{6}{6^2} = \frac{1}{6}$ である. 5 回までに終了するのは, 4 回目までに 2 回同じ目が出て, 5 回目と同じ目が出て終了する. よって, ${}_4C_2(\frac{1}{6})^2(\frac{5}{6})^2\frac{1}{6} = \frac{25}{216}$. (b) P_n は $n-1$ 回までに 2 回同じ目が出て, n 回目に同じ目が出るので, $P_n = {}_{n-1}C_2(\frac{1}{6})^2(\frac{5}{6})^{n-3}\frac{1}{6}$. よって, $\frac{P_{n+1}}{P_n} = \frac{5n}{6n-2}$. (c) $\frac{P_{n+1}}{P_n} \geq 1$ のとき, $n \leq 12$, また, $\frac{P_{n+1}}{P_n} < 1$ のとき, $n > 12$, さらに, $n = 12$ では $\frac{P_{n+1}}{P_n} = 1$. 以上から, $P_1 < P_2 < \dots < P_{12} = P_{13} > P_{14} \dots$ となるので, $n = 12, 13$



駒 P は四面体 OABC の頂点の 1 つにある. 1 個のさいころを投げて, 出た目が 3 以上ならば駒 P を他の 3 頂点のいずれかに等しい確率で移し, 出た目が 2 以下ならば移さない. この試行を X とする. 初めは頂点 O に駒 P がある. 試行 X を n 回繰り返した後, 駒 P が頂点 A にある確率を P_n とする. このとき, 次の問いに答えよ.



(a) P_1 を求めよ.

(b) P_2 を求めよ.

(b) P_{n+1} と P_n の関係式を求めよ.

(c) P_n を n を用いて表せ. また, 極限值 $\lim_{n \rightarrow \infty} P_n$ を求めよ.

◆【2016, 2020 栃木県中高他多数】

(a) 出た目が 3 以上で, 他の 3 頂点のいずれかに等確率 $\frac{1}{3}$ で移るので, $P_1 = \frac{4}{6} \cdot \frac{1}{3} = \frac{1}{9}$

(b) 2 回の試行後に駒 P が頂点 A にあるのは $O \rightarrow O \rightarrow A$, $O \rightarrow A \rightarrow A$, $O \rightarrow B \rightarrow A$, $O \rightarrow C \rightarrow A$ の 4 通りの場合があり, 排反であるから,

$$P_2 = \frac{2}{6} \cdot \frac{4}{6} \cdot \frac{1}{3} + \frac{4}{6} \cdot \frac{1}{3} \cdot \frac{2}{6} + \frac{4}{6} \cdot \frac{1}{3} \cdot \frac{4}{6} \cdot \frac{1}{3} + \frac{4}{6} \cdot \frac{1}{3} \cdot \frac{4}{6} \cdot \frac{1}{3} = \frac{21}{80}$$

(c) $n+1$ 回の試行後に, 駒 P が頂点 A にあるのは次の 2 つの場合がある. ①駒 P が n 回後に A にあり, 2 以下の目が出る場合, ②駒 P が n 回後に A 以外 (確率は $1 - P_n$) にあり, 3 以上の目が出る場合. ①②から, $P_{n+1} = \frac{2}{6}P_n + \frac{4}{6}\frac{1}{3}(1 - P_n)$. まとめると,

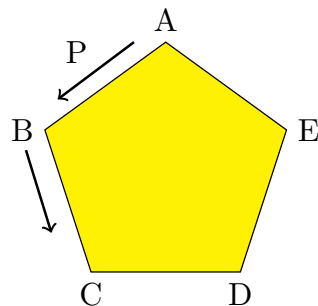
$$P_{n+1} = \frac{1}{9}P_n + \frac{2}{9}$$

(c) $P_{n+1} - \frac{1}{4} = \frac{1}{9}\left(P_n - \frac{1}{4}\right)$ と書き直して漸化式を解くと, $P_n = \left(P_1 - \frac{1}{4}\right)\left(\frac{1}{9}\right)^{n-1} + \frac{1}{4}$ となり, 初項を代入すると $P_n = -\frac{5}{36}\left(\frac{1}{9}\right)^{n-1} + \frac{1}{4}$. また, 極限値は $\lim_{n \rightarrow \infty} P_n = \frac{1}{4}$

84. ある実験を n 回繰り返す. ある事柄 A が起こる確率を p_n とする. 直前の実験で A が起こった時に続いて A が起こる確率を a , 直前で実験 A が起こらなかった時に続いて A が起こる確率を b とする. p_{n+1} を p_n , a , b を用いて表せ.

☞ 確率の漸化式はほとんどこの形になる.

85. 図のように, 1 辺の長さが 1 の正五角形 $ABCDE$ がある. この正五角形の辺上を反時計まわりに動く点 P はじめ頂点 A にある. P は, 1 回のさいころを 1 回投げるごとに自の数のだけ進む. たとえば, 最初に 2 の目が出たときは, P は頂点 A から頂点 C に移動して止まり, 次に 4 の目が出たときは, P は頂点 C から頂点 B に移動して止まる. サイコロを n 回投げ終わったとき, P が頂点 A , B で止まる確率をそれぞれ a_n , b_n とする, このとき, 次の問いに答えよ.



- (a) b_2 および b_3 を求めよ.
 (b) b_{n+1} と a_n の関係式を求めよ.
 (c) $\lim_{n \rightarrow \infty} b_n$ を求めよ.

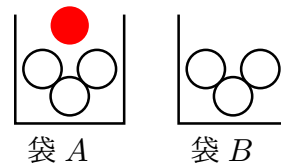
☞ (a) 2 回投げたとき, P が頂点 B で止まるのは, 1 回目の目を x , 2 回目の目を y とすると, $x + y = 1, 6, 11$ の場合である (2 回投げるので 1 はない). (b) 頂点 B で止まるのは A で 1 あるいは 6 の目が出た場合で, A 以外の頂点 (確率は $1 - a_n$) では 1 から 6 までのいずれかの目が出た場合. (c) $n \rightarrow \infty$ ではどの頂点でも同じ確率.

86. 十分な個数の当たり付きのお菓子が売られている. そのお菓子を 1 個購入して当たりが出れば同じお菓子をもう 1 個もらうことができ, もらったお菓子で当たりが出ても同じお菓子をもう 1 個もらうことができるという. 各お菓子で当たりが出る確率はすべて等しく p ($0 < p < 1$) であり, このお菓子を 1 個購入し, 獲得することができたお菓子の総数 (最初に購入した 1 個を含む) の期待値を K とする. ただし, 獲得したお菓子が n 個になったとき, 当たりが出てもそれ以上お菓子はもらえず, 獲得することができたお菓子の総数は n 個とする. このとき, 次の各問いに答えよ.

- (a) K_n を求めよ.
 (b) 極限 $\lim_{n \rightarrow \infty} K_n$ を求めよ.

☞ (a) 期待値を求める問題で, $K_n = 1 \times (1 - p) + 2 \times p(1 - p) + \dots$ (b) K_n から pK_n を引くと見通しがつく.

87. 袋 A には、赤玉 1 個、白玉 3 個、袋 B には、白玉 3 個が入っている。袋 A, B から同時に玉を無作為に 1 個ずつ取り出し、入れ替えて袋に戻すという作業を n 回行う。 n 回目の作業後、袋 A に赤玉が入っている確率を P_n とするとき、次の問いに答えよ。



- (a) P_1, P_2 を求めよ。
 (b) P_{n+1}, P_n の間に成り立つ関係式を求めよ。
 (c) P_n を求め、 $\lim_{n \rightarrow \infty} P_n$ を求めよ。
88. 箱 A に赤球が 2 個、箱 B に白球が 2 個入っている。「箱 A から球を 1 個取り出して箱 B に移したあと、箱 B から球を 1 個取り出して箱 A に移す」という試行を n 回繰り返すとする。ただし、どの球が取り出される確率も同様に確からしいとする。 n 回目の試行の後で箱 A に入っている赤球が 0 個、1 個、2 個である事象をそれぞれ S_0, S_1, S_2 とする。
- (a) n 回目の試行の後で事象 S_0, S_1, S_2 となる確率をそれぞれ p_n, q_n, r_n とする。 $p_n + q_n + r_n$ の値を求めよ。
 (b) r_{n+1} を q_n, r_n で表せ。
 (c) q_{n+1} を p_n, q_n, r_n で表せ。
 (d) r_n の漸化式とその一般解を求めよ。



最初に表が出るまで、コインを繰り返し投げ続ける実験を行う。最初の表が出るまでに必要なコインを投げる回数を n ($n = 1, 2, 3, \dots$) とする。 n 回目ですべて表が出る事象を A_n として、その確率 $p_n \equiv P(A_n)$ を、以下の方法で求めよ。ただし、 $P(H) = p, P(T) = q = 1 - p$ とする。

- (a) 単純に求める。
 (b) 分配則を用いて求めよ。
 (c) p_n の最大値を n をもちいて表せ。ただし、 $n \geq 2$ とする。

$$\diamond(a) P(\overbrace{T, \dots, T}^{n-1}, H) = P(T)^{n-1} P(H) = q^{n-1} p.$$

(b) 最初の投げ ($n = 1$) で表が出る事象を $B = H$ とすると、 $\bar{B} = T$ なので、 $A = A_n$ とおくと、分配則は $P(A_n) = P(A_n|H)P(H) + P(A_n|\bar{B})P(\bar{B}) = P(A_n|H)p + P(A_n|\bar{B})q$. 左辺は p_n である。 $P(A_n|H)$ は、最初の投げで表が出た場合に、 n 回目の投げで始めて表が出る事象の条件付確率である。ところが、 A_n は n 回目ですべて表が出る事象なので、それ以前の投げ $n-1, n-2, \dots$ では表が

出ていないはずである。したがって、この条件付き確率は $P(A_n|H) = 0$ でなければならない。これから、上式は $P(A_n) = P(A_n|T)q$ となる。つぎに、 $P(A_n|T)$ の意味について考えよう。 $P(A_n|\text{裏})$ は最初の投げで裏が出て、 n 回目の投げで始めて表が出る事象の条件付き確率である。最初の投げ以降の投げ $(2, 3, \dots, n)$ を考えれば、合計 $n-1$ 回投げる実験を行うことになる。この実験は、2 回目から最初に表が出るまで $n-1$ 回試行を行う事象 A_{n-1} に他ならない。したがって、 $P(A_n|T) = P(A_{n-1}) = p_{n-1}$ 。これから $p_n = q \cdot p_{n-1}$ が得られ、求める確率は $p_n = pq^{n-1}$ となる。

$$\begin{array}{c} p_n \\ \overbrace{T + \bigcirc + \bigcirc \cdots + \bigcirc} \\ \underbrace{\bigcirc + \bigcirc \cdots + \bigcirc}_{p_{n-1}} \end{array}$$

(c) $p_n = p(1-p)^{n-1}$ を p で微分して 0 とおくと、 $p = 1/n$ で最大になる。

89. 表が確率 p 、裏が確率 $q = 1-p$ で生起する硬貨で、最初に表が出るまで繰り返して振る実験を行う。偶数回 $(2, 4, 6, \dots)$ 振ったときに、最初に表が出る確率 $P(E)$ を求めよ。

☞ S を最初の振りで表が出る事象とすると、分配則から $P(A) = P(A|S)P(S) + P(A|\bar{S})P(\bar{S})$ 。ここで、 $P(A|S) = 1$ 、 $P(A|\bar{S}) = P(\bar{A}) = 1 - P(A)$

90. 表が確率 p 、裏が確率 $q = 1-p$ で生起する硬貨で、最初に表が出るまで繰り返して振る実験を行う。偶数回 $(2, 4, 6, \dots)$ 振ったときに、最初の表が出る確率 $P(E)$ を以下の方法で求めよ。

(a) 偶数回に始めて表の出る事象の確率をすべて調べる。

(b) 分配則を利用する。

☞ (a) $\{\text{裏, 表}\}, \{\text{裏, 裏, 裏, 表}\}$ などと続くので、独立を利用して級数の和を求める。

(b) 最初に投げた硬貨の表と裏の事象に関する分配則を利用。偶数回で表が出るので $P(E|\text{表}) = 0$ で、 $P(E) = P(E|\text{表})P(\text{表}) + P(E|\text{裏})P(\text{裏}) = P(E|\text{裏})q$ となる。

91. 【大相撲】大相撲で三力士が同じ勝ち星で 15 日間を終えると、巴戦で優勝を決める。巴戦とは勝ち残り 2 人ずつ対戦し、最初に 2 連勝した力士を優勝とする方式である。最初に対戦して勝った力士、負けた力士、最初は控えに回った力士のそれぞれの優勝確率を計算せよ。ただし、どの勝負も勝つ確率は五分五分とする。

☞ 理論的には無限回相撲を取る可能性があるのですが、級数の和で表せる事になるが、その必要はない。例として最初に勝った力士 A が優勝するという事象 A の確率を求める。次の勝負で優勝が決まるという事象を D とすると、まず、 $(\textcircled{A}, \textcircled{B})(\textcircled{A}, \textcircled{C})$ となれなよい。あと 1 回勝てば優勝するので $P(A) = \frac{1}{2}$ である。次に、 $(\textcircled{A}, \textcircled{B})(\textcircled{A}, \textcircled{C})(\textcircled{B}, \textcircled{C})$ ならば決まらないので一巡し、最初の状況から始めるので、 $P(A) = \frac{1}{2} \frac{1}{2} \frac{1}{2} P(A)$ 。こうして $P(A) = \frac{1}{2} + \frac{1}{2} \frac{1}{2} \frac{1}{2} P(A)$

が成り立つ。最初に勝った力士が有利になる。先の【先手必勝】の問題と原理は同じ。

以下は少々難しい問題



92. A 君, B 君がゲームをする。1 回ごとに負けた方が勝った方に 1 ポイントを渡す。 A 君が勝つ確率を p , B 君が勝つ確率を $q = 1 - p$ とする。はじめ, A 君は k ポイントを, B 君は $a - k$ ポイント持っているものとし, いずれかの持ち点が 0 あるいは a ポイントになればゲームは終了する。

(a) A 君が 0 ポイントとなる確率 q_k は $q_k = p \cdot q_{k+1} + q \cdot q_{k-1}$ を満たすことを証明せよ。

(b) 上記方程式の解を条件 $q_0 = 1$, $q_a = 0$ のもとに求めよ。

93. 2 人で同じ硬貨を投げ, 出る面によって得点が異なるゲームを行う。 A 君は表が出れば 1 点の得点を得, B 君は裏が出れば 1 点の得点を得るようなゲームである。ルールは少々複雑である。いま, 整数 $a \geq 0$, $b \geq 0$ を適当に定め, B 君が b 点得点する (b 回裏が出る) 前に A 君が a 点得点すれば (a 回表が出る) A が勝ち, そうでなければ B が勝つものとする。 A が勝つ事象を $A_{a,b}$ とし, 確率 $p(a, b) = P(A_{a,b})$ を以下の手順で求めよ。

(a) $p(a, b) = \frac{1}{2}p(a-1, b) + \frac{1}{2}p(a, b-1)$ が成り立つことを示せ。

(b) 初期条件は $p(0, b) = 1, p(a, 0) = 0$ となることを示せ。

(c) 初期条件を満たす解は $p(a, b) = \frac{1}{2^{a+b-1}} \sum_{k=0}^{b-1} {}_{a+b-1}C_k$ と表せることを示せ。

☞ (a) 最初硬貨を投げたときの表あるいは裏に対して分配則を適用。(c) これが初期条件 (b) を満たす解であることを元の方程式 (a) に代入することで確かめよ。その際,

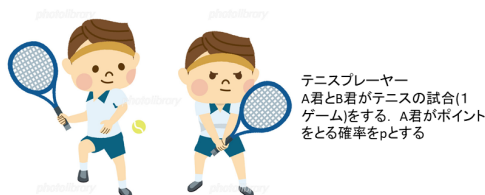
$$2 \sum_{k=0}^{b-1} {}_{a+b-2}C_k + 2 \sum_{k=0}^{b-2} {}_{a+b-2}C_k = 2 \sum_{k=0}^{b-1} {}_{a+b-2}C_k + 2 \sum_{k=1}^{b-1} {}_{a+b-2}C_{k-1} = 2 \sum_{k=0}^{b-1} {}_{a+b-2}C_k + 2 \sum_{k=0}^{b-1} {}_{a+b-2}C_{k-1} = 2 \sum_{k=0}^{b-1} {}_{a+b-1}C_k \text{ を用いよ。}$$

94. 【未完】

§ 3 応用問題



95. テニスの勝敗



プロテニスプレイヤーの統計 (勝敗とポイントをとる確率が詳細に載っている) を

ATP の HP (www.atpworldtour.com/en/performance-zone, 2020 年 2 月) から引用すると, ジョコビッチ選手と錦織選手の対戦成績は,

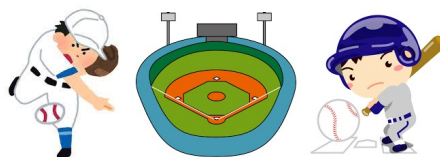
プレーヤー	ジョコビッチ	錦織
対戦成績	16 勝	2 勝
Service Record	67%	64%
Return Record	54%	52%

であった. この表を見て, ポイントをとる確率では 3 % 程度しか差がないが勝率は 89% にもなる. ジョコビッチ対錦織が 16 勝対 10 勝なら納得できだろうが, 力の差はほとんど無いにも関わらず, なせ, 錦織選手がジョコビッチ選手に勝てないのか不思議に思わないだろうか?

確率の問題として以下のように考えてみよう. A と B がテニスの試合 (1 ゲーム) をする. A がポイントをとる確率を p とする.

- A_k ($k = 0, 1, 2$) を B が 0, 15, 30 ポイントをとって, A が勝つ事象とする. 各事象の確率を求めよ.
- デュースとなる事象の確率 D を求めよ.
- 条件付確率 $P(A_d|D)$ を用いて, $P(A_d) = P(A_d|D)P(D)$ と表せる. $P(A_d|D)$ を求めよ.
- デュースになる場合も含めて, A がゲームを勝つ取る確率を求めよ
- $p = 0.64$ として, $g(0.64)$ を求めよ.
- 実際の対戦成績と比較して考察せよ.

96. プロ野球の日本シリーズの勝敗



日本シリーズでの勝敗に関して考察しよう. 通常 7 戦あり, 先に 4 勝した場合が優勝する. Wikipedia によると, 2019 年までの結果は下表のようになっている (引き分けを含む). 割合は回数の分布で, 最終戦までもつれこむ場合が意外にももっとも大きい.

試合数	回数	割合 (%)	$g_{4+k}(0.5)$
4	6	8.7	12.5
5	17	24.7	25.0
6	22	31.9	31.3
7	23	33.3	31.3
8	1	1.4	0.27

さて、チーム A, B が対戦したときの勝率をそれぞれ p , $(1-p)$ とする. まず A が $4+k$ 戦目で優勝する確率はその前までに 3 勝 k 敗して最後に 1 勝する確率だから ${}_{3+k}C_3 p^3 (1-p)^k p$. B が $4+k$ 戦目で優勝する確率も同様だから、これらを足せば第 $4+k$ 戦までいく確率を $g_{4+k}(p)$ とする.

(a) $g_{4+k}(p) = {}_{3+k}C_3 (p^3 (1-p)^k p + p^k (1-p)^3 (1-p))$ を証明せよ.

(b) 第 7 戦における, つまり, $k=3$ とおいた $g_7(p)$ を p の関数として表せ.

(c) $g_7(p)$ を最大にする p と最大値を求めよ.

97. 【SNA の拡散   】 ある人が n 人へ「このメールを受け取ったら, n 人へメールを送ってください」というメールを出したとしよう. メールを受け取った人が次の人に送らない確率を p とする. このメールを受け取ったら, n 人へメールを送ってください」というメールを出したとき, このチェーンメールはどこで止まるのだろうか. チェーンメールがどこかで止まる確率を $P_n(p)$ としよう. 次の人にメールを送らなければ ($p=1$), どこかで止まる ($P_n(p)=1$). $P_n(p=1)$ を以下の手順で求めよう.

(a) まず, この問題で必要になる n 次方程式 $x = p + (1-p)x^n$ の解について復習しておこう. ただし, 方程式の 1 つの解は $x=1$ であることを用いて, 1 以外で $x > 0$ を満たす解を求めよ.

(i) $x = p + (1-p)x^2$ の解は, $x =$

(ii) $x = p + (1-p)x^3$ の解は, $x =$

➡ 組立除法 (覚える?) を用いて因数分解. また, $n=5$ までは解析解はあるが, 手計算では困難.

(b) 送った n 人の内, 1 人が送らない事象を A とする (つまりここで止まる). いくつか止まる事象を B とすると, 分配則 $P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A})$ が成り立つ. これを利用して, $P_n(p) = p + (1-p)(P_n(p))^n$ が成り立つことを示せ.

(c) $P_2(p)$, $P_3(p)$ を求め, グラフを描け.

(d) n が大きいとき. 解を求めるのは困難である. しかし, メールが止まらず拡散するかどうかを知りたいので, $P_n(p)$ が 1 に近い時の近似解を $P_n(p) = 1 - \epsilon$

$0 < \epsilon \ll 1$) として調べよ.

- (e) 上記の結果から, メールが止まらず拡散する, つまり $P_n(p) = 1$ となる p を n を用いて表し, 考察せよ.

§4 確率変数

累積分布関数 (分布関数) : $F(x) = P(X \leq x)$

確率密度関数 (密度関数) : $f(x)$

累積分布関数と確率密度関数の関係 (積分と微分の関係と同類)

	連続型確率変数	離散型確率変数
累積分布関数 $F(x)$	$\int_{-\infty}^x f(x)dx$	$\sum_{k=1}^{[x]} P(X = k)$
確率密度関数 $f(x)$	$\frac{dF(x)}{dx}$	$P(X = k)$

$[x]$ はガウス記号で, x を超えない整数

確率密度関数 $f(x)$ が満たすべき条件 :

$$\begin{cases} f(x) \geq 0 \\ \int_{-\infty}^{\infty} f(x)dx = 1 \\ P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a) \quad (a \leq b) \end{cases}$$

累積分布関数 $F(x)$ が満たすべき条件 :

$F(x)$ は非減少関数, $P(\emptyset) = 0$ より $F(-\infty) = 0$, $P(\Omega) = 1$ より $F(\infty) = 1$

$$0 \leq F(x) \leq 1$$

平均値 (期待値) $m = E(X)$ と **分散** $\sigma^2 = V(X) = E((X - m)^2)$:

	連続型確率変数	離散型確率変数
平均 m	$\int_{-\infty}^{\infty} xf(x)dx$	$\sum_k kP(X = k)$
分散 σ^2	$\int_{-\infty}^{\infty} (x - m)^2 f(x)dx$	$\sum_k (k - m)^2 P(X = k)$

中央値 : $\int_{-\infty}^{x_0} f(x)dx = F(x_0) = \frac{1}{2}$ を満たす x_0

◆1 累積分布関数と確率密度関数



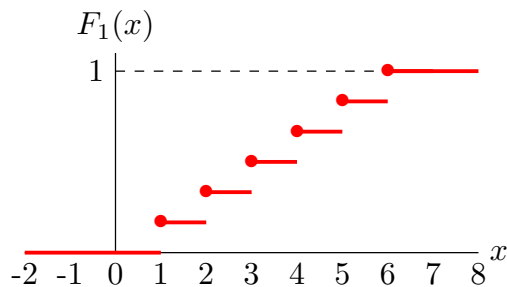
さいころ 1 個の目に関する離散型確率分布 $f_1(x) = P(x = k) = \frac{1}{6} (k = 1, 2, \dots, 6)$, また 2 回さいころを振ったときの目の和 $(x_1, x_2 = 1, 2, \dots, 6)$ に関する確率分布は $x = x_1 + x_2$ とおくと, $f_2(x) = P(X = x) = \frac{6 - |7 - x|}{36} (x = 2, 3, \dots, 12)$ は以下の表にまとめられる.

x	1	2	3	4	5	6
$f_1(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

x	2	3	4	5	6	7	8	9	10	11	12
$f_2(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

- (a) $f_1(x)$ の平均 m_1 , 分散 σ_1^2 , 累積分布関数 $F_1(x)$ を求めよ.
 (b) $f_2(x)$ の平均 m_2 , 分散 σ_2^2 , 累積分布関数 $F_2(x)$ を求めよ.

◇(a) $m_1 = \sum_{x=1}^6 x f_1(x) = \frac{1 + \dots + 6}{6} = \frac{21}{6} = \frac{7}{2}$, $\sigma_1^2 = \sum_{x=1}^6 (x - m_1)^2 f_1(x) = \frac{35}{2}$, 累積分布関数 $F_1(x) = P(X \leq x) = \sum_{k=1}^{[x]} P(X = k)$ は下図のようになる (累積分布関数は任意の x で定義され, 離散的確率変数では階段状のグラフになる). たとえば, $2 \leq x < 3$ (等号に注意) では $[x] = 2$ となるので, $F_1(x) = \sum_{k=1}^2 P(X = k) = P(X = 1) + P(X = 2) = 2 \times \frac{1}{6}$ である. $F_1(x)$ は累積分布関数の満たすべき条件を満たしていることを確認せよ. (b) $m_2 = \sum_{x=2}^{12} x f_2(x) = \frac{252}{36} = 7$, $\sigma_2^2 = \sum_{x=2}^{12} (x - m_2)^2 f_2(x) = \frac{35}{6}$. 累積分布関数 $F_2(x)$ は略



98. コインを投げる実験において, 表の出る確率を $P(H) = p$, 裏の出る確率を $P(T) = q = 1 - p$ とする. 離散型確率変数を $X(H) = 1$, $X(T) = 0$ と定める.

- (a) 累積分布関数を求め, 図示せよ.
 (b) 平均, 分散を求めよ.

👉 累積分布関数 $F(x) = P(X \leq x)$ を決める際, x の値によっては H, T が入るかどうかがチェックする必要がある.

99. 離散型確率分布が $P(X = k) = \frac{a}{k!} (k = 0, 1, 2, \dots)$ と与えられている.

- (a) a の値を求よ
- (b) 累積分布関数 $F(x)$ を求めよ.
- (c) $P(X > 2)$ を求めよ.

☞ (c) 離散型確率変数の場合, 等号に注意.

100. **プロ野球** 日本シリーズでは 7 戦で先に 4 勝したチームが優勝する. 各試合は独立で 2 チームは互角の力があると仮定して, 優勝が決まるのが第 X 戦としよう. (すでに述べた理論考察から) 確率分布は表のようになる.

x	4	5	6	7
$P(X = x)$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{5}{16}$

- (a) 平均, 分散を求めよ.
- (b) 累積分布関数を求め, 図示せよ.

101. **目の大小** さいころを 3 個投げて出た目のうち大きい方 (同じ目のときはその目) を X , 小さい方を Y とする.

- (a) X, Y の累積分布関数を求め, 図示せよ.
- (b) X, Y の平均, 分散を求めよ.
- (c) 投げる回数を n とし, $n \rightarrow \infty$ のときの X, Y の平均を求めよ.
- (d) $E(X) + E(Y)$ は n に依らず一定の値になる. その値を求めよ.

☞ 目の最大最小の確率の問題と同じように考えよう. たとえば, 大きい方が $X = k$ となるのは, $1 \sim k$ から $1 \sim (k-1)$ の場合を引けばいい.

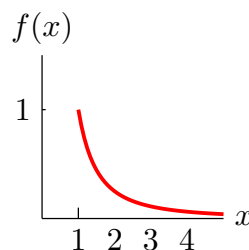
102. **確率変数の線型変換** 線型変換 $Y = aX + b$ の場合を考えよう. X の累積分布関数 $F_X(x)$, 確率密度関数 $f_X(x)$ は与えられとする.

- (a) 累積分布関数 $F_Y(y)$ を, $F_X(x)$ を用いて表せ.
- (b) 確率密度関数 $f_Y(y)$ を, $f_X(x)$ を用いて表せ.

☞ (a) a が正の場合と負の場合に分ける. 参考: $Y = \sigma X + m$ とすると, 標準正規分布 $X \sim N(0, 1)$ が一般の正規分布 $Y \sim N(m, \sigma^2)$ に変換される.

103. **へんな分布: べき分布の特徴** 確率変数 X の確率密度関数が $f(x) = x^{-2} (x \geq 1)$ で与えられたとする. 正規分布のような単峰性の分布とは著しく異なる性質を持っている. 分布は下のグラフを参照.

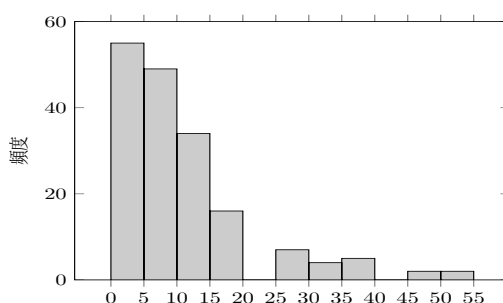
- (a) $f(x)$ が確率密度関数の条件をすべて満たしていることを示せ.
- (b) 累積分布関数を求めよ.
- (c) 平均, 分散, 中央値を求めよ.



平均や分散が存在しない分布は特殊なように思えるが、実際にはインターネット、所得分布、金融、地震（次の問題）など様々な分野で観測されている。べき分布はべき関数で表される。

104. 【余震の時間間隔分布】 下図は東日本大震災の本震後の余震の時間間隔のヒストグラムである（データの出典：www.iris.eduseismonlast30.html）。横軸 x の単位は分である。正確ではないが、おおよそべき関数 $x^{-1.5}$ で近似できるとして問いに答えよ。

- (a) べき関数を確認めよ。
(b) 平均、分散、中央値を求めよ。



105. ベンフォードの法則もへんな分布

分布は $f(X = d) = \log_{10} \frac{d+1}{d}$ であった。ここで、 $d = 1, 2, \dots, 9$ は任意の数字の最初の桁の整数である。

- (a) 平均と中央値を比べよ。
(b)

◆2 条件付き累積確率分布と条件付き確率密度関数



確率密度関数が $f(x) = ce^{-cx} (x \geq 0)$ で与えられたとする。ここで、 c は正の定数である。

- (a) 累積分布関数 $F(x)$ を求めよ。
(b) 条件付き累積分布関数 $F(x|X > t)$, 条件付き確率密度関数 $f(x|X > t)$ を求め、 $f(x)$ を用いて表せ。

◇(a) $x < 0$ で $F(x) = 0$, $x \geq 0$ で $F(x) = 1 - e^{-cx}$. (b) $0 < t < x$ とする。

条件付き確率から $P(X \leq x|X > t) = \frac{P(X \leq x \cap X > t)}{P(X > t)} = \frac{P(t < X \leq x)}{P(X > t)}$

となる。ここで、 $P(t < X \leq x) = \int_t^x ce^{-cx} dx = e^{-ct} - e^{-cx}$, $P(X > t) =$

$\int_t^\infty ce^{-cx}dx = e^{-ct}$. よって, 条件付き累積分布関数は $F(x|X > t) = P(X \leq x|X > t) = 1 - e^{-c(x-t)} = 1 - f(x-t)$. 条件付き確率密度関数は $f(x|X > t) = \frac{dF(x|X > t)}{dx} = e^{-c(x-t)} = f(x-t)$.

106. ある製品の寿命が x である確率が $P(X \geq x) = x^{-1}$ ($x = 1, 2, 3, \dots$) と表せたでしょう. いま, B を期間 a では故障しない事象とする. B が与えられた場合, 製品の寿命が x となる条件付き確率密度関数を求めよ.

☞ 題意から $P(B) = P(X \geq a) = a^{-1}$. $P(X \geq x|B) = P(X \geq x \cap B)/P(B) = P(X \geq x)/P(B)$,

107. 硬貨を 3 回投げ, X を表の出る回数とする. 2 回以上表が出たことが分かっている場合の条件付き平均 $E(X|X \geq 2)$ を求めよ. また, 条件をつけない場合の平均 $E(X)$ と比較せよ.

☞ 2 回以上表がでるので, $E(X|X \geq 2) > E(X)$ は明らかであろう.

108. クイズをする. 正解すると賞金 a 円を受け取り, 2 回連続して誤答すると, クイズを終えなければならない. 正答する確率を p , 誤答する確率を $q = 1 - p$ とする.

- (a) クイズをする平均回数を求めよ.
 (b) 期待される賞金 Y を求めよ.
 (c) 賞金を受け取る ($E(Y) > 0$) ための p の範囲を求めよ.

☞ X をクイズを X 回する事象, C を正答する事象とすると, $P(X) = P(X|C)P(C) + P(X|\bar{C}C)P(\bar{C}C) + P(X|\bar{C}\bar{C})P(\bar{C}\bar{C})$. ここで, $P(AB)$ は最初に A , 次に B が生じる事象の確率とする.

以下は少々難しい問題



109. 平均値, 中央値の重要な意味 以下のことを証明せよ.

- (a) 平均値は, 平均 2 乗誤差 (mean square error)

$$C_1 = \int_{-\infty}^{\infty} (x - m)^2 f(x) dx \text{ を最小にする } m \text{ である.}$$

- (b) 中央値は, 絶対誤差 (mean absolute error)

$$C_2 = \int_{-\infty}^{\infty} |x - c| f(x) dx \text{ を最小にする } c \text{ である.}$$

☞ 平均値や中央値はその周りのばらつき具合が最小であることを示している. 両者の代表値としての違いはばらつきの評価である.

110. 確率変数 X の分散 $V(X)$ を, Y に関して条件をつけた平均 $E(X|Y)$, 分散 $V(X|Y)$ を利用して算出できる

$$V(X) = E(V(X|Y)) + V(E(X|Y))$$

が成り立つことを証明せよ. ここで, $E(V(X|Y))$ の E は Y についての期待値,

$V(E(X|Y))$ の V は Y についての分散である。難しそうな式であるが、「 X の分散は、条件付き分散 $V(X|Y)$ の Y についての平均と、条件付き平均 $E(X|Y)$ の Y についての分散の和である」

☞ $E(X) = E(E(X|Y))$, $E(X^2) = E(E(X^2|Y))$ である。これらの式に $V(X) = E(X^2) - E(X)^2$ を使う。

111. ここで改めて、問題 33 を見直して、ベイズの定理の意義を考えよ。

§ 5 いろいろな確率分布

一様分布 (離散型, 連続型): $P(X = k) = \text{一定}$, $f(x) = \text{一定}$

二項分布 $B(n, p)$ (離散型): ある試行の結果の生起する事象 A の確率 $P(A) = p$ とし, この試行を独立に n 回繰り返したときの事象 A が生じた回数を X とする。

$$P(X = k) = {}_n C_k p^k (1-p)^{n-k}$$

$$E(X) = np, V(X) = np(1-p)$$

正規分布 $N(m, \sigma^2)$ (連続型): $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$

$$E(X) = m, V(X) = \sigma^2$$

ポアソン分布 (離散型): 稀にしか起らない現象の生起する事象の数を X とする。

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

$$E(X) = \lambda, V(X) = \lambda$$

◆1 いろいろな分布

112. 日常よく用いられる分布に標準的な一様分布, 二項分布, 正規分布, ポアソン分布がある。どのようなもの(現象)に用いられているか調べよ。

一様分布 _____

二項分布 _____

正規分布 _____

ポアソン分布 _____

113. 同じ対象であっても, その時代で分布が大きく変化する場合もある。たとえば, インターネットの通信量(パケット)はかつてはポアソン分布が適用できたが, 今では当てはまらなく別な分布(べき分布など)が必要になってきた。このように標準的な分布以外に数多くの分布がある。興味があれば調べてみよう。

◆1 一様分布

114. **離散型一様分布** 離散型確率変数 X が $\{1, 2, \dots, N\}$ で一様分布するとしよう (下図も参照). このとき, 確率分布は $k = 1, 2, \dots, N$ に対して,

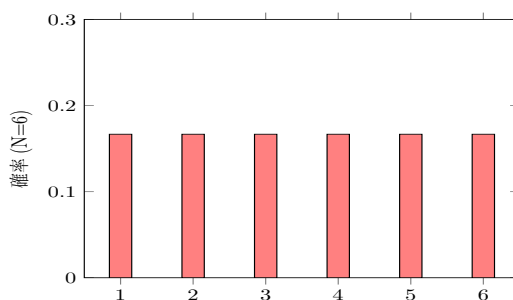
$$P(X = k) = \boxed{}$$

である. 平均は $m = E(X) = \sum_{k=1}^N kP(X = k)$ から, $m = \boxed{}$ である. また, 分散は $\sigma^2 = V(X) = E(X^2) - m^2$ から求める.

$$\begin{aligned} \text{理由: } V(X) &= \sum_{k=1}^N (k - m)^2 P(X = k) = \sum_{k=1}^N (k^2 - 2mk + m^2) P(X = k) \\ &= \sum_{k=1}^N (k^2 - m^2) P(X = k) = \sum_{k=1}^N k^2 P(X = k) - m^2 \\ &= E(X^2) - m^2 \end{aligned}$$

ここで, $E(X^2) = \boxed{}$ を用いると, $\sigma^2 = \boxed{}$ である.

👉 数列の和 $\sum n$, $\sum n^2$ の公式覚えてる?

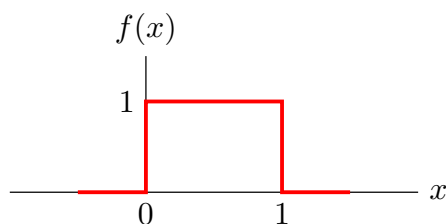


115. **連続型一様分布** 連続型確率変数 X の取る値の範囲を $0 \leq X \leq 1$ とする. 連続型一様分布とは, 離散型と同様に, その確率密度関数 $f(x)$ が一定の場合である (下図も参照). $f(x) = a$ のおくと, 確率密度関数が満たすべき条件 $\int_0^1 f(x)dx = 1$ から, $a = \boxed{}$ であり,

$$f(x) = \begin{cases} 0 & x < 0 \\ \boxed{} & 0 \leq x \leq 1 \\ 0 & 1 < x \end{cases}$$

となる. 平均は $E(X) = \int_0^1 xf(x)dx$ から, $m = \boxed{}$ であり. 分散は

$V(X) = \int_0^1 (x - m)^2 f(x)dx$ から, $\sigma^2 = \boxed{}$ である.



116. $a, b (a < b)$ を定数とする. 確率変数 X の取る値の範囲が $a \leq X \leq b$ で, その確率密度関数が $f(x) = k(a \leq x \leq b)$ と表されたとしよう. このとき, 定数 k の値および平均, 分散を求めよ.

☞ 前問の区間 $[0, 1]$ での一様分布の一般化.

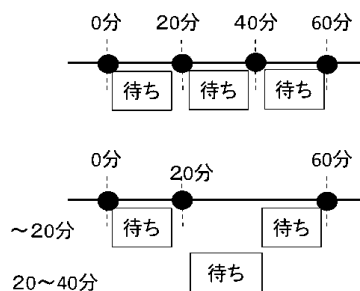
117. 一様分布の累積確率分布 確率変数 X が区間 $[a, b]$ で一様な確率密度関数の累積確率分布を求め, 図示せよ.

118. ある駅に到着する電車  の運行予定が次の (a), (b) の 2 つの場合について, 行き当たりばったり (無作為) に駅に到着するとする. 次の電車を待つ時間の確率分布 (確率密度関数) を求め, 12 分以上次の電車を待つ確率を求めよ.

(a) 毎時 0 分, 20 分, 40 分に電車が到着する.

(b) 毎時 0 分, 20 分に電車が到着する.

☞ (a) 待ち時間が 0 から 10 分間での一様分布, (b) 待ち時間が 0 から 20 分の場合と, 20 分から 40 分の場合があり, それぞれ一様分布 (下図を参考).



119. A 駅を発車した特急列車  は 20 分, 急行列車  は 25 分, 普通列車  は 30 分で B 駅に到着する. 特急列車は毎時 0 分, 急行列車は毎時 20 分と 40 分, 普通列車は毎時 x 分, $x + 20$ 分および $x + 40$ 分に A 駅を出発する. 乗客は時刻表を知らないものとし, A 駅に到着後, 最も早く発車する列車に乗るものとする. このとき, 次の問に答えよ.

(a) 乗客が特急列車, 急行列車に乗る確率をそれぞれ求めよ.

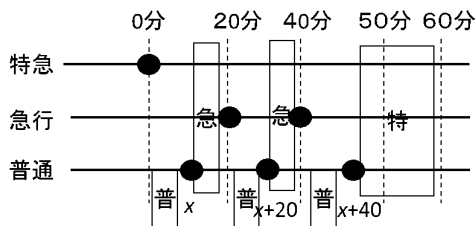
(b) 乗客が A 駅に到着してから, 普通列車に乗車し B 駅に到着するまでの平均時間を求めよ.

(c) 乗客が A 駅に到着してから, B 駅に到着するまでの時間の期待値を最小にする整数 x を求めよ.

☞ 前問の応用で, 普通列車, 急行列車, 特急列車で待ち時間が異なる (下図を参考).

(b) 一様分布の平均を利用する. (c) は 2 次関数の最小値を求める問題になるが, x は整数であることに注意.

120. 長さ L の棒を無作為に 2 分割したとき, 短いほうの断片の長さを X とする.



(a) X の確率密度関数を求めよ.

(b) X の平均, 分散を求めよ.

☞ X のとりうる範囲は $0 \leq X \leq L/2$ であることに注意.

121. 中心を O とする半径 R の円の内部に無作為に 1 点を選び, その点を通る中心を O とする円の面積を X とする.

(a) X の密度関数を求めよ.

(b) 平均, 分散を求めよ.

☞ 無作為の意味は, 1 点を通る中心を O とする円の面積が一様になるとする. 半径の 1 点が一様になる場合だと難しくなるので後で扱うことにする.

122. 長さ a の線分上で無作為に 1 点を選ぶ. 短かい方の線分と長い方の線分の長さの比 R が $\frac{1}{3}$ より小さい確率を求めよ.

☞ 選んだ点 X が $a/2$ より小さいか大きいかにによって場合分け. たとえば, $X < a/2$ のとき, $X/(a - X) < 1/3$ から, $X < a/4$ となるので, 求めたい確率への寄与は $P(X < a/2)P(X < a/4|X < a/2) = P(X < a/4)$ と表せる.

◆2 二項分布

123. 1 個のさいころを 3 回投げるとき, 1 の目が出る回数を確率変数 X とする. X の確率分布の特徴を以下の手順で調べよう.

(a) 確率 $P(X = k) (k = 0, 1, 2, 3)$ の値を求めよ.

$$P(X = 0) = \boxed{}, P(X = 1) = \boxed{}$$

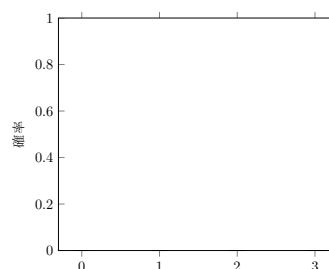
$$P(X = 2) = \boxed{}, P(X = 3) = \boxed{}$$

(b) 以下の表をうめ, また, ヒストグラム (棒グラフ) を描け.

(c) 平均値 (期待値), 分散を求めよ. また, 二項分布 $B(n, p)$ の公式 $m = E(X) = np$, $V(X) = \sigma^2 = np(1 - p)$ に一致することを確認めよ.

☞ 二項分布の平均 np , 分散 $np(1 - p)$ は大変重要であるが, この例で確かに成り立っていることを確かめよう.

X	0	1	2	3	合計
P					1



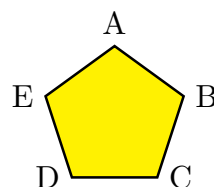
124. 次の問いに答えよ.

- (a) $B\left(12, \frac{1}{4}\right)$ の平均 (期待値), 分散を求めよ.
- (b) 1 枚の硬貨を 5 回投げるとき, 表が出る回数を X とする. $P(X \leq 1)$, 平均, 分散を求めよ.
- (c) 不良品 3% の製品の山から 50 個取りだしたときの不良品の個数を確率変数 X とするとき, X の平均, 分散, 標準偏差を求めよ.
- (d) ある試験で問題が 5 問あり, それぞれの問題の解答が 4 択になっている. 全ての問題に無作為に解答したとき, 2 問だけ正解する確率を求めよ.

125. 以下の問いに答えよ.

- (a) さいころを 5 回投げたとき, 6 の目がちょうど 2 回出る確率を求めよ.
- (b) 袋の中に, 白玉 6 個, 赤玉 4 個が入っています. この袋から無作為に玉を 1 個取り出し, 色を調べてからもとに戻します. この試行を 4 回続けて行うとき, 白玉がちょうど 2 回出る確率を求めよ.
- ➡ 白玉がちょうど 2 回出るのは ○●○●● のような場合である.
- (c) A, B の 2 チームが試合をする. 先に 4 勝したチームが優勝となり, その後の試合は行わない. 全ての対戦で A が勝つ確率と B が勝つ確率はともに $\frac{1}{2}$ で, 引き分けはないものとするとき, 6 試合目で A の優勝する確率を求めよ.
- ➡ 6 試合目で A が優勝するのは BAABAA のような場合である.

126. 正五角形 ABCDE を考える. A, B, C, D, E はこの順番に時計回りに並んでいる. A から出発して, さいころを振って出た目が奇数なら右隣の点へ移り, 偶数なら左隣の点へ移る. 7 回さいころを振るとき, 点 $X(X = A, B, C, D, E)$ にいる確率 P_X を求めよ.



- ➡ 奇数が x 回, 偶数が y 回出たとすると $x + y = 7$. たとえば, B にいるとすると, $x - y = 1, -4, 6$ で, 1 周する場合も含む.

127. n を 3 以上の自然数とする. 当たりくじ 2 本を含む n 本のくじがある. くじを引いて, 当たりなら持ち点に 1 を加算し, はずれなら持ち点は変わらないとする. 最初の持ち点は 0 とし, くじを引いてはもどすという試行を n 回繰り返す. k を 0 以上の整数とする. n 回の試行が終了した時点の持ち点が k となる確率を $p_n(k)$ と

する.

(a) 確率 $p_n(k)$ を求めよ.

(b) $p(k) = \lim_{n \rightarrow \infty} p_n(k)$ を求めよ. ただし, e を自然対数の底とすると,

$$\lim_{n \rightarrow \infty} \left(1 - \frac{2}{n}\right) = e^{-2} \text{ であることを用いてもよい.}$$

(c) 値 $p(k)$ が最大となるような k の値を求めよ.

☞ (a) 二項分布, (c) $p(k+1)/p(k)$ の値を調べよう.

128. 【負の二項分布】 途中でゲームが終わった. 賞金の配分どうしよう

(a) ある事象の生起する確率が $P(X=1)=p$, $P(X=0)=1-p$ のベルヌイ試行の列 $X_1, X_2, \dots$ において初めて r 回 1 が出る間に出る 0 の回数を N とする. ただし, $r \geq 1$ は整数の定数である. N の分布を求めよ.

(b) A, B の 2 人がコイン投げによるゲームをする. どちらかが先に 5 勝した段階でゲームは終わり, 賞金 1600 円を受け取ることになっていたが, A が 3 勝, B が 2 勝した段階でゲームを中止することとなった. 賞金はどのように配分するのが公平だろうか, 上問 (a) をもとに計算せよ.

☞ (b) すでに A が 3 勝, B が 2 勝している. この後, A が勝つ場合は, A はあと 2 勝し (最後は A が勝つ), B が勝つ回数は 0, 1, 2 までである.

☞ 二項分布 (試行回数は固定, 事象の起こる回数は確率変数). 一方, 負の二項分布 (事象の起こる回数は固定, 試行回数は確率変数).

129. コインを 2 人が独立に n 回投げるとき, 以下の問いに答えよ.

(a) $\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}$ を証明せよ.

(b) 2 人が同じ回数だけ H を出す確率を求めよ.

(c) n が大きいとき, この確率は $\frac{1}{\sqrt{\pi n}}$ と近似できることを示せ.

☞ (a) $(1+t)^n(t+1)^n = (1+t)^{2n}$ の両辺に 2 項定理を適用, (b) 独立な事象の同時分布は各事象の確率の積, (c) スターリングの近似式 $n! \cong \sqrt{2\pi n} n^n e^{-n}$ を用いよ.

130. 二項分布の母関数とモーメント母関数 二項分布に関して以下の問いに答えよ.

(a) 母関数は $G(s) = (q + ps)^n$ と表せることを示せ.

(b) モーメント母関数は $G(s) = (q + pe^t)^n$ と表せることを示せ.

☞ 母関数とモーメント母関数に関しては後で説明.

131. 二項分布の平均と分散 二項分布 $B(n, p)$ の平均値 (期待値) が $E(X) = np$, 分散が $V(X) = np(1-p)$ であることを以下の考え方で証明せよ.

(a) k 回目試行で事象が生起すれば 1, 生起しなければ 0 となるような確率変数を X_k とすると, $X = X_1 + X_2 + \dots + X_n$ と表せる. 各 X_k は独立であることを利用する.

- (b) 二項定理 $(a+b)^n = \sum_{k=0}^n {}_nC_k a^k b^{n-k}$ を用いる. ここに, $a = p$, $b = 1 - p$ に代入する.
- (c) 定義通りに, $m = E(X) = \sum_{k=0}^n kP(X = k)$, $V(X) = \sum_{k=0}^n k^2 P(X = k) - m^2$ を計算する. ここで, $P(X = k) = {}_nC_k p^k (1-p)^{n-k}$.
- (d) 二項分布の母関数 $G(s) = (q + ps)^n$ を用いる.
- 👉 (d) の母関数に関しては後で説明.

◆ 3 正規分布

標準正規分布表を用いた演習は統計の方で行う

132. 正規分布とは

- (a) 平均 m , 分散 σ^2 の正規分布の確率密度関数 $N(m, \sigma^2)$ は

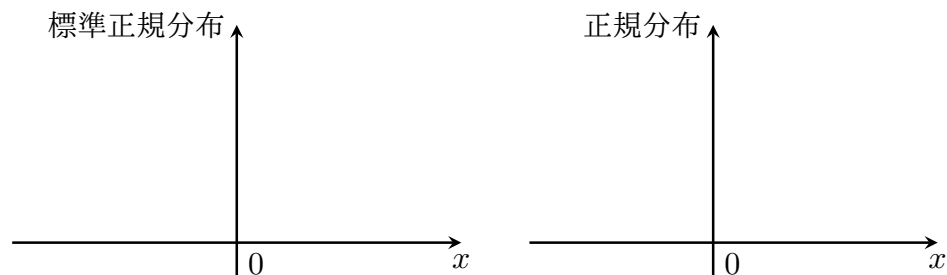
$$f(x,) = \frac{1}{\sqrt{2\pi \boxed{}}} \exp\left(-\frac{\boxed{}}{2\boxed{}}\right)$$

である. 特に, $m = \boxed{}$, $\sigma = \boxed{}$ の正規分布を標準正規分布といい, その確率密度関数は

$$f(x) = \frac{1}{\sqrt{\boxed{}}} \exp\left(-\frac{\boxed{}}{2\boxed{}}\right)$$

である. したがって, $N(0, 1)$ と表せる.

- (b) グラフを書いて, 標準正規分布と正規分布の違いを指摘せよ.



133. $g(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$ (ガウス関数とも言う) と表される標準正規分布の確率密度関数は,

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

を満たすことを証明せよ.

👉 簡単そうに見えて結構難しい. 積分を I とおき, $I = 1$ ではなく $I^2 = 1$ を証明する.

134. 正規分布 $N(m, \sigma^2)$ の確率密度関数 $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$ の変曲点は $m \pm \sigma$ であることを示せ.

135. 標準正規分布の累積分布関数は $G(x) = \int_{-\infty}^x g(z) dz$ は前問から $G(\infty) = 1$ を満たす. 以下の等式を証明せよ.

(a) $G(-x) = 1 - G(x)$

(b) $|x|$ が十分大きいと, 近似的に $G(x) \cong 1 - \frac{1}{x}g(x)$ と表せる.

☞ (a) $g(x)$ の縦軸に関する対称性を利用. (b) この場合, 累積分布 $G(x)$ は積分をしなくても $g(x)$ から直接求められる.

136. ある航空会社が, 予約をした人で実際には飛行機に乗らない可能性が 10% あることを経験的に知っていたとする. このとき, 420 の座席に 441 の予約をとったとすると, 正規分布を使って, 少なくとも 1 人をオーバーブックする確率を求めよ.

☞ 格安航空券で搭乗する場合は早めに空港に行こう.

137. X が標準正規分布 $N(0, 1)$ にしたがう,

$$Y = \sigma X + m$$

とする, このとき, Y は平均 m , 分散 σ^2 の正規分布にしたがうことを証明せよ.

☞ これは $N(m, \sigma^2)$ は $N(0, 1)$ から簡単な変換で求められることを示す. これを利用したのが標準正規分布表であるが, これに関しては統計の方で調べる.

138. 正規分布 $N(m, \sigma^2)$ にしたがう確率変数 X に対して, 偏差値

$$Y = 50 + 10 \frac{X - m}{\sigma}$$

はどのような分布にしたがうか.

☞ 標準化: $N(m, \sigma^2) \rightarrow N(0, 1)$, 偏差値: $N(m, \sigma^2) \rightarrow N(50, 10^2)$

139. 正規分布の平均, 分散, 中央値 正規分布に関して以下のことを証明せよ.

(a) 平均 $E(X) = \int_{-\infty}^{\infty} xf(x)dx$ は m を与える.

(b) 分散 $V(X) = \int_{-\infty}^{\infty} (x - m)^2 f(x)dx$ は σ^2 を与える.

(c) 中央値 $\int_{-\infty}^{x_0} f(x)dx = \frac{1}{2}$ となる x_0 は m を与える.

◆4 ポアソン分布

140. ポアソン分布の平均と分散 ポアソン分布にしたがう確率変数 X の確率分布は

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

と表せる. ここで $(k = 0, 1, 2, \dots)$ ある. λ はパラメータあるいは母数と呼ばれている.

(a) $\sum_{k=0}^{\infty} P(X = k) = 1$ 満たすことを示せ.

(b) 平均は $E(X) = \lambda$, 分散も $V(X) = \lambda$ となることを示せ.

☞ $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} \dots = \sum_{k=1}^{\infty} \frac{x^k}{k!}$ を利用しよう.

141. 以下の問いに答えよ.

(a) ジャンボジェットには 400 名の乗客が乗れる. 予約した各乗客が独立に確率

0.01 でキャンセルするものとする．もし，404 名の予約を受け付けば，少なくとも 1 名の乗客が乗れなくなる確率はいくらか．

(b) 3000 個の部品で，1 個の部品が不良である確率は 10^{-3} であった．この中に 5 個より多い不良品がある確率を求めよ．

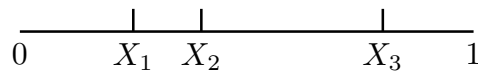
(c) サッカーのワールドカップ決勝戦のチケットがキャンセルされる割合は 0.01% である．1 万枚のチケットを販売したとき，キャンセルされるチケットの数 X はポアソン分布にしたがうとする．ただしチケット購入者がキャンセルする場合，各々の行動は独立とする．決勝戦でキャンセルが 2 つ以上ある確率を求めよ．

以下は少々難しい問題



142. U_1, U_2 および U_3 を互いに独立に区間 $(0, 1)$ 上の一様分布にしたがう確率変数とし， X_1 をそれらのうち最も小さいもの， X_2 を 2 番目に小さいもの，そして X_3 を最も大きいものとする．このとき， $j = 1, 2, 3$ に対し， X_j の確率密度関数 $f_j(x)$ を求め，それぞれの期待値を求めよ．

👉 U_1, U_2, U_3 の最小値が x となる確率は， U_j の中のいずれか（3 通り）が x の値を取り，それ以外の二つの U_j が x 以上であればよい．一様分布なのでそれらの確率は $1 - x$ で， $f_1(x) = 3(1 - x)^2$ ．

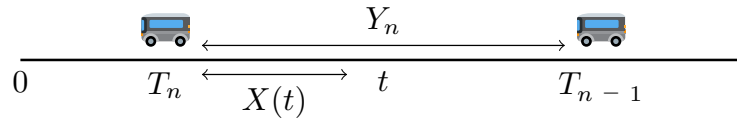


143. 【バスの待ち時間にいらいら】 大学前の停留所で時刻通りに来ないバスにいつもいらいらする．そこで以下のような問題を考えよう．停留所にバスとそれに乗りたい客が次々に到着する T_n を n 台目のバスの到着（＝出発）時刻とする．ここに， $T_0 = 0$ とする．バスは到着するとその時刻に待っていた客を瞬時に全員乗せて出発する．また，その時刻にバス停に来た客は待ち時間なしでそのバスに乗るものとする． $X(t)$ を時刻 t にバス停に来た客がバスに乗るまでに待つ時間とすると， $X(t)$ は以下のようなになる．

$$X(t) = \begin{cases} 0 & t = 0 \\ T_n - t & T_{n-1} < t \leq T_n \end{cases}$$

待ち時間 $\{X(t)\}_{t \in [0, \infty)}$ の平均を以下にしたがって計算してみよう．（最初のバスの出発 ($t = 0$) から十分時間が経過した後は，この過程は定常過程と仮定する．この部分は気にしないで進めよう）．定常状態で無作為な時刻にバス停に来る客の待ち時間の期待値を求める．簡単のために，バスの到着時間間隔 $Y_n = T_n - T_{n-1} (n = 1, 2, \dots)$ は互いに独立で期待値，分散が有限な同一分布に

したがうとし、1次モーメントを $m_1 = E[Y_n]$ 、2次モーメントを $m_2 = E[Y_n^2]$ とおく。



- (a) 平均待ち時間 $\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t X(s) ds$ は、 $\frac{m_2}{2m_1}$ と表せることを示せ。
- (b) $\frac{m_2}{2m_1} \geq \frac{m_1}{2}$ を示せ。
- (c) バスが一定間隔でくる場合、 m_1 は一定で平均待ち時間が $\frac{m_1}{2}$ である。以上の結果はこれより大きい事を示しているが、考察せよ。長い時間待たされることがあると、平均はその値に大きく依存する。

☞ m_1 は平均 m 、分散は $\sigma^2 = m_2 - m^2$ であった。(a) 直感では、平均待ち時間はバスの平均到着間隔の半分 $m_1/2$ と考えそうであるが実は違う。

144. 二項分布 $B(n, p)$ において p は未知として、 n 回の試行のうち、 X 回事象が生じたとする。このとき、実験結果として得られる n, X を用いて、最も適切な p の値を決めよ。「同様の確からしさ」の問題で扱ったように、実社会の問題では初めから p が与えられている状況にあるとは限らない。この問題は、実験をして得られる結果から、逆に p を推定する問題である。もっとも、推定するための基準が必要になる。ここでは、確率 ${}_nC_x p^x (1-p)^{n-x}$ を p について最大にするという基準を用いる。

☞ このように確率を最大にする方法を最尤推定法と呼び、情報科学ではよく用いる。

145. **重要な不等式と応用** 有限な期待値、分散をもつどのような分布にしたがう確率変数 X であっても、以下の不等式が成り立つことを証明せよ。

- (a) 【マルコフの不等式】 a を正の実数とすると、 $P(|X| \geq a) \leq \frac{E[|X|]}{a}$ が成り立つ。

- (b) 【チェビシェフの不等式】 k を正の実数とすると、 $P(|X - m| \geq k) \leq \frac{\sigma^2}{k^2}$

☞ 期待値のまわりの確率のおおまかな広がりを与えるという意味で重要。たとえば $k = 2\sigma$ とおくと、平均から標準偏差の2倍以上離れた値は全体の1/4を超えないことがわかる。

- (c) マルコフの不等式の応用

有価証券報告書(2014年)によると、上場企業3660社に従業員数の平均はおおよそ1400人である。このとき、マルコフの不等式を用いて、10000人以上の企業数はどの程度か推定せよ。

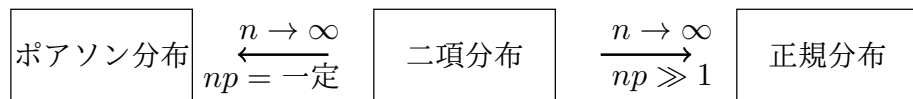
- (d) チェビシェフの不等式の応用

上記の問題同様，上場企業 3660 社に従業員数の平均はおよそ 1400 人，標準偏差は 3680 人である．このとき，チェビシェフの不等式を用いて，10000 人以上の企業数はどの程度か推定せよ．

146. **二項分布，正規分布，ポアソン分布の関係** 以下のことを証明せよ．

(a) 正規分布 $N(m, \sigma^2)$ は，二項分布 $B(n, p)$ で n が大きく， np が小さくない場合の近似になっている．

(b) ポアソン分布は，二項分布 $B(n, p)$ で n が大きく， np が小さい場合の近似になっている．



147. 【指数分布の無記憶性】 確率変数 X がパラメタ λ の指数分布にしたがう（確率密度関数 $f(x) = \lambda e^{-\lambda x} (x \geq 0)$ ）とき， $P(X > t) = P(X > s + t | X > s)$ が成り立つことを「指数分布の無記憶性」と呼ぶ．実際，

$$\text{左辺: } P(X > t) = 1 - P(X \leq t) = 1 - \int_0^t \lambda e^{-\lambda x} dx = e^{-\lambda t}$$

$$\text{右辺: } P(X > s + t | X > s) = P(X > s + t)P(X > s) = e^{-\lambda(s+t)}e^{-\lambda s} = e^{-\lambda t}$$

例えば，「（ある時点から）ある町において交通事故が発生するまでの時間」は，指数分布で近似できる．今日事故が起きるかどうかの予測は昨日事故が起きたかどうかの影響を受けない，と仮定できるならば，「今から t 日間，無事故が続く確率」と「過去 s 日間無事故であったとして，さらに今日からあと t 日間無事故が続く確率」（これは「残余寿命が t 日以上」と解釈できる）が等しくなるのは，自然である．なお，連続分布で無記憶性を持つのは指数分布だけである．

§ 6 確率変数の関数

同時累積分布関数 : $F_{XY}(x, y) = P(X \leq x, Y \leq y)$

$$F_{XY}(x, y) = \sum_{x \in X} \sum_{y \in Y} P(x, y)$$

同時確率密度関数 : $f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$

$$\begin{cases} f_{XY}(x, y) \geq 0 \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y) dx dy = 1 \\ P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x, y) dx dy \quad (a < b, c < d) \end{cases}$$

周辺確率密度関数 :

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y) dx$$

X と Y が独立 : $F_{XY}(x, y) = F_X(x)F_Y(y)$, $f_{XY}(x, y) = f_X(x)f_Y(y)$

共分散 : $\text{Cov}(X, Y) = E((X - m_X)(Y - m_Y))$

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

X と Y が独立 : $\text{Cov}(X, Y) = 0$

相関係数 (ピアソンの積率相関係数) : $\rho = \frac{E((X - m_X)(Y - m_Y))}{\sqrt{V(X)}\sqrt{V(Y)}} \quad -1 \leq \rho \leq 1$

注意 : 相関は関連を示すだけで, 因果関係ではない

◆ 1 2 つの確率変数の線形結合

148. 大学入学前と入学後の成績のように, 相関関係が重要な場合が実社会では多く見られる. 具体例をいくつか挙げよ.

_____ と _____
 _____ と _____
 _____ と _____



確率変数 X, Y に対して, a, b を定数として, 以下のことを証明せ.

(a) $E(aX + b) = aE(X) + b$

(b) $V(aX + b) = a^2V(X)$

(c) $E(X + Y) = E(X) + E(Y)$

確率変数 X と Y が独立ならば, 以下のことを証明せよ.

(d) $E(XY) = E(X)E(Y)$

(e) $\text{Cov}(X, Y) = 0 \rightarrow V(X + Y) = V(X) + V(Y) \Leftarrow$ 以下の問題

◇連続型確率変数の場合について, いろいろな方法で証明する.

(a) $E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy = \frac{1}{|a|} \int_{-\infty}^{\infty} y f_X\left(\frac{y-b}{a}\right) dy$ と変形する. ここで,
 $a \geq 0$ の場合, $\int_{-\infty}^{\infty} (ax+b) f_X(x) dx = a \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} b f_X(x) dx$ となる.

また, $a < 0$ の場合, $-\int_{\infty}^{-\infty} (ax+b) f_X(x) dx = a \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} b f_X(x) dx$
 となる. いずれの場合でも, $E(Y) = aE(X) + b$ となる.

(b) $V(Y) = E((Y - m_Y)^2) = E((aX + b - am_X - b)^2) = a^2 E(X^2 - 2m_X X +$

$$m_X^2) = a^2(E(X^2) - 2m_X E(X) + m_X^2) = a^2 E(X^2 - m_X^2) = a^2 V(X)$$

(c) 同時確率密度関数を $f(x, y)$ とすると,

$$\begin{aligned} E(X + Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x + y) f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x, y) dx dy + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} x f_X(x) dx + \int_{-\infty}^{\infty} y f_Y(y) dy = E(X) + E(Y) \end{aligned}$$

(d) X と Y が独立なので, 周辺確率密度関数をそれぞれ $f_X(x), f_Y(y)$ とすると, $f(x, y) = f_X(x)f_Y(y)$ とかける. よって,

$$\begin{aligned} E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_X(x) f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} f_X(x) dx \int_{-\infty}^{\infty} y f_Y(y) dy = E(X)E(Y) \end{aligned}$$

149. 独立な場合の例 袋 A には赤玉 2 個, 白玉 3 個, 袋 B には赤玉 3 個, 白玉 2 個が入っている. それぞれの袋から 2 個の玉を同時に取り出すとき, 取り出した計 4 個の中の赤玉の個数を Z とする.

(a) 袋 A から 2 個取り出したときの赤玉の個数を確率変数 X , 袋 B から 2 個取り出したときの赤玉の個数を確率変数 Y とする. X, Y の確率分布をそれぞれ求めよ.

$$\begin{aligned} P(X = 0) &= \boxed{}, P(X = 1) = \boxed{}, P(X = 2) = \boxed{} \\ P(Y = 0) &= \boxed{}, P(Y = 1) = \boxed{}, P(Y = 2) = \boxed{} \end{aligned}$$

(b) X, Y の平均および分散をそれぞれ求めよ.

(c) Z の平均および分散を求めよ.

150. 独立でない場合の例 2 本の当たりくじを含む 10 本のくじがある. まず A 君がくじを 1 本引き, 残りのくじから B 君が同時に 2 本引くとく.

(a) A 君, B 君の当たりくじの数をそれぞれ X, Y とする. X と Y の同時分布を求めよ.

$X \setminus Y$	0	1	2	合計
0				
1				
合計				1

- (b) $E(X), E(Y)$ を求めよ.
- (c) $E(XY) \neq E(X)E(Y)$ を確かめよ.
151. コインを2回投げ, X を1回目と2回目の出た面が同じとき10, そうでないとき0である確率変数, Y を1回目と2回目が両方表のとき20, それ以外のとき0である確率変数とする.
- (a) 期待値 $E(X), E(Y)$, 分散 $V(X), V(Y)$ を求めよ.
- (b) 同時確率密度関数 $f_{XY}(x, y)$ を求め, 以下の表を埋めよ.

$X \setminus Y$	20	0	合計
10			$\frac{1}{2}$
0			$\frac{1}{2}$
合計	$\frac{1}{4}$	$\frac{3}{4}$	1

- (c) 共分散 $Cov(X, Y)$ を計算し, X, Y が独立かどうか調べよ.
- (d) $E(X + Y) = E(X) + E(Y)$, $V(X + Y) = V(X) + V(Y) + 2Cov(X, Y)$ が成り立つことを確かめよ.
152. 下表で与えられる同時分布について, 確率変数 X と Y は互いに独立として空欄をうめよ.

$X \setminus Y$	1	2	合計
1	$\frac{2}{5}$		
2	$\frac{1}{5}$		
合計			1

👉 周辺確率分布 $P(Y = k) = \sum_{j=1}^2 P(X = j, Y = k)$, 独立性 $P(X = j, Y = k) = P(X = j)P(Y = k)$ などを使う.

153. 2つの確率変数の分散 確率変数 X と Y に対して, 以下のことを証明せよ.
- (a) $V(X + Y) = V(X) + 2Cov(X, Y) + V(Y)$. ここで, 共分散は $Cov(X, Y) = E((X - m_X)(Y - m_Y))$ で定義する.
- (b) 特に X と Y が独立なとき, $V(X + Y) = V(X) + V(Y)$
154. $E(X) = 10/3, V(X) = 20/9$ のとき, 次の確率変数 Y の期待値, 分散を求めよ.
- (a) $Y = 3X - 2$
- (b) $Y = -2X$
155. 確率変数 X の期待値が2で分散が5, 確率変数 Y の期待値が-1で分散が3であり, X と Y が互いに独立であるとする. 次の確率変数の期待値と分散を求めよ.
- (a) $X + Y$

(b) $X - 2Y$

156. 硬貨を2枚投げて、表の出る硬貨の枚数を X 、さいころを1個投げて、出る目の数を Y とする。このとき、 $Z = 3X - Y$ で表される確率変数 Z の期待値を求めよ。ただし、投げた硬貨は表か裏のどちらか一方が出るものとし、その出方は同様に確からしいものとする。また、さいころは、1 から 6 のどの目が出ることも同様に確からしいものとする。

157. 期待値 5、標準偏差 2 の確率変数 X から、変換 $Y = aX + b$ によって、期待値 0、標準偏差 1 の確率変数 Y をつくりたい。定数 a, b の値を求めよ。ただし、 $a > 0$ とする。

158. 区間 $[0, 1]$ で一様分布する連続型確率変数を X 、また区間 $[a, b]$ 上で連続一様分布する確率変数を Y とする。

(a) 確率変数 X の平均、分散を求めよ。

(b) X と Y の関係式を求めよ。

(c) Y の平均、分散を計算せよ。

☞ $[0, 1]$ 上の一様分布では $E(X) = \frac{1}{2}, V(X) = \frac{1}{12}$ 。「いろいろな分布…」の問題を参照。

159. 2つの確率変数 X と Y に対して、

(a) 相関係数が 1 になるのは X と Y がどのような関係にある場合か。

(b) -1 の場合はどのような場合か。

☞ 任意の z に対して、 $V(zX + Y) \geq 0$ であるが、これに判別式を利用。

160. 【コーシー・シュワルツの不等式】 2つの確率変数 X と Y に対して、 $|E(XY)| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}$ が成り立つことを証明せよ。等号は $Y = kX$ となる定数 k が存在するときである。

☞ $E(|Y - kX|^2) \geq 0$ に判別式を利用。

161. 【さいころ何回投げれば㊦がでるかな】 さいころの㊦がでる確率は $\frac{1}{6}$ であるが、6回振れば㊦は1回は出てくるだろうと、なんとなく思う。このことを確率的にとらえる以下の問いに答えよ。

(a) 確率 p で起きる事象が生起するまでの試行回数の期待値を求めよ。

(b) 確率 p で起きる事象が k 回生起するまでの試行回数の期待値を求めよ。



2つの確率変数の和, $Z = X + Y$ がどのような分布にしたがうか調べよう.

(a) $x+y \leq z$ の領域での確率が Z の累積確率分布 $F_Z(z) = P(X+Y \leq z)$ である. X と Y の同時確率密度関数 $f_{XY}(x, y)$ を用いて表せ.

(b) 独立なとき, どのように表される.

具体例: 保険会社が独立に多数の保険請求を受けた時の請求の合計を求める場合, 多数の工程で発生する欠陥の和として考える工場の品質管理など

$$\blacktriangleright (a) F_Z(z) = \iint_{x+y \leq z} f_{XY}(x, y) dx dy = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{z-x} f_{XY}(x, y) dx dy$$

である. 2重積分の順序を変えて, $\int_{y=-\infty}^{\infty} \int_{x=-\infty}^{z-y} f_{XY}(x, y) dx dy$ としてもよい.

このとき, 確率密度関数は $f_Z(z) = \frac{dF_Z(z)}{dz}$ は

$$f_Z(z) = \int_{-\infty}^{\infty} f_{XY}(z-y, y) dy = \int_{-\infty}^{\infty} f_{XY}(x, z-x) dx$$

と表される

(b) 特に, X と Y が独立なら, 周辺確率密度関数を用い $f_{XY}(x, y) = f_X(x)f_Y(y)$ が成り立つので,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y)f_Y(y) dy = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x) dx$$

と表せる^{*3}. (離散: $p_z = \sum_x p_x p_{z-x}$)

162. X と Y が区間 $[0, 1]$ で一様で独立とする.

(a) $Z = X + Y$ の累積分布関数 $F_Z(z)$, 確率密度関数 $f_Z(z)$ を求めよ. また, $f_Z(z)$ を図示せよ.

(b) $f_Z(z)$ は $z = 1$ に最大値を持つ単峰性の分布であるが, 一様分布の和からなぜそのような分布になるのは. その理由を考えよ.

(c) $f_Z(z)$ を用いて, 平均 $E(Z)$, 分散 $V(Z)$ を求めよ. また, これからの値は, $E(X+Y) = E(X) + E(Y)$, $V(X+Y) = V(X) + V(Y)$ を用いて計算した値に等しいことを示せ.

👉 (b) 解答には3つの一様分布する変数の和の確率密度関数も示しておいた. なめらかで釣鐘型の分布になる. さらに多くの変数の和をとると, 中心極限定理から正規分布に近づく.

^{*3} 任意の関数 $g(x)$ に対して, $\frac{d}{dx} \int^{g(x)} f(u, v) du = \frac{dg(x)}{dx} f(g(x), v)$ が成り立つ. 特に, $g(x) = x$ とすれば,

$$\frac{d}{dx} \int^x f(u, v) du = f(x, v)$$

163. X と Y が区間 $[-a, a]$ で一様で独立とする. ただし, $a > 0$ とする. $X + Y$ の場合を参考にすると, $Z = X - Y$ の確率密度関数は, X と Y が独立ならば, $f_Z(z) = \int_{-\infty}^{\infty} f_X(z+y)f_Y(y)dy$ と表せる.

(a) Z の確率密度関数 $f_Z(z)$ を求めよ.

(b) $f_Z(z)$ は $z = 0$ に最大値を持つ単峰性の分布であるが, 一様分布の差からなぜそのような分布になるのは. その理由を考えよ.

(c) $f_Z(z)$ を用いて, 平均 Z , 分散 $V(Z)$ を求めよ. また, これからの値は, $E(X + Y) = E(X) + E(Y)$, $V(X - Y) = V(X) + V(Y) = V(X)V(Y)$ を用いて計算した値に等しいことを示せ.

164. 互いに独立な連続型確率変数 X, Y に対して, $Z = \min(X, Y)$ の確率密度関数 $f_Z(z)$ を $f_X(x)$ と $f_Y(y)$ で表せ.

☞ 累積分布関数は定義から $F_Z(z) = P(\min(X, Y) \leq z) = P(X \leq z \cup Y \leq z) = P(X \leq z) + P(Y \leq z) - P(X \leq z \cap Y \leq z) = F_X(z) + F_Y(z) - F_X(z)F_Y(z)$.

165. 独立な確率変数 X, Y は同じ累積分布関数 $F(t)$ を持つとする. すなわち $P(X \leq t) = P(Y \leq t) = F(t)$. とする.

(a) $P(X \leq t, Y \leq t)$ を $F(t)$ で表わせ.

(b) 確率変数 $Z = \max\{X, Y\}$ の累積分布関数 $F_Z(z)$ を求めよ.

☞ (b) $P(\max\{X, Y\} \leq z) = P(X \leq z \cap Y \leq z)$.

166. X, Y の同時確率密度関数を $f(x, y) = c(x + y)$ と表せる. ただし, x と y の範囲は $0 \leq x + y \leq 1, x, y \geq 0$ とする.

(a) c の値を求めよ.

(b) 周辺確率密度関数 $f_X(x)$ を求めよ.

☞ 積分範囲に注意.

167. 確率変数 X, Y を原点から半径 1 の範囲で一様に分布するとき, 単位円 (半径が 1) の面積は π なので, 同時確率密度関数は $f(x, y) = \frac{1}{\pi} (x^2 + y^2 \leq 1)$ と表される.

(a) X および Y の周辺確率密度関数を求めよ.

(b) X と Y が独立かどうか調べよ.

168. 会計報告など, 膨大な数字が掲載されている冊子の中から, ランダムに取り上げた数字の上 2 桁の数字をそれぞれ X, Y ($X = 1, 2, \dots, 9, Y = 0, 1, 2, \dots, 9$) とする. このとき, X と Y の同時確率分布が

$$p_{x,y} = P(X = x, Y = y) = \log_{10}(1 + (10x + y)^{-1})$$

と表せることが経験的に知られている. このとき, X の周辺確率分布 $p_x = P(X = x)$ がベンフォードの法則を与えることを示せ.

☞ X, Y のとる数の範囲に注意.

以下は少々難しい問題



169. 【バスと電車の待ち時間にいらいら】

- (a) A 駅を 15 分毎に発車するバスがある. 無作為に A 駅に着くとき, 乗車するまでの待ち時間を X 分とする. 平均 $E[X]$ および $P(X < 3)$ を求めよ.
- (b) B 駅でバスを下車し, 無作為に電車に乗り継ぐ. 電車は B 駅を 10 分ごとに発車するが, 電車に乗るまでの待ち時間を Y とする. X と Y が独立のとき, $E(X + Y)$ および $P(X + Y < 6)$ を求めよ.



以下の 3 題は確率分布の再生性に関する. 再生性とは同じ種類の独立な確率変数の和はその種類の確率変数になるという性質で, 以下の問題で取り上げる二項分布, 正規分布, ポアソン分布以外に, コーシー分布, カイ 2 乗分布, 負の二項分布などが知られている.

170. 【二項分布の和は二項分布】 二項分布にしたがう独立な確率変数 $X_1 \sim B(n, p), X_2 \sim B(m, p)$ の和 $X_1 + X_2$ は二項分布 $B(n + m, p)$ にしたがうことを以下の手順で示せ.

- (a) $\sum_{x=0}^n {}_n C_x \times {}_m C_{z-x} = {}_{n+m} C_z$ を証明せよ.
- (b) これを用いて, $X_1 + X_2$ は $B(n + m, p)$ にしたがうことを示せ.

171. 【正規分布の和は正規分布】 $\sigma_1, \sigma_2 > 0$ とする. X_1, X_2 をそれぞれ $N(m_1, \sigma_1), N(m_2, \sigma_2)$ にしたがう独立な確率変数とすると, $X_1 + X_2$ は $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ にしたがうことを示せ.

172. 【ポアソン分布の和はポアソン分布】 確率変数 X_1 がパラメータ λ_1 の, X_2 がパラメータ λ_2 のポアソン分布にしたがうとする. このとき, X_1, X_2 が独立ならば, $X_1 + X_2$ はパラメータ $\lambda_1 + \lambda_2$ のポアソン分布にしたがうことを示せ.

◆2 確率変数の非線形関数



$Y = X^2$ のとき, Y の累積分布関数 $F_Y(y)$, 確率密度関数 $f_Y(y)$ と, X の累積分布関数 $F_X(x)$, 確率密度関数 $f_X(x)$ に関係について, 以下の問いに答えよ.

- (a) $F_Y(y)$ を $F_X(x)$ を用いて表せ.
- (b) $f_Y(y)$ を $f_X(x)$ を用いて表せ.
- (c) X が標準正規分布 $N(0, 1)$ にしたがう場合に Y の累積分布関数と確率密度関数を求めよ.
- (d) このとき, 平均 $E(Y)$ と分散 $V(Y)$ を求めよ.

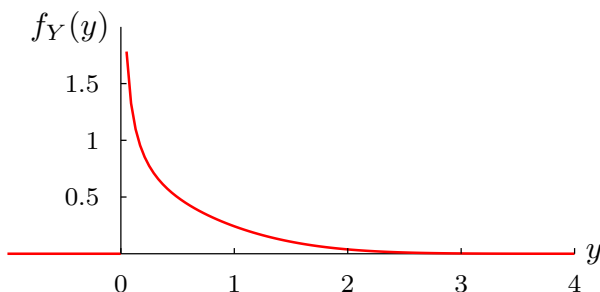
👉 $E(X^2) = E(X)^2$ とならないことを確認せよ.

◇(a) $y \geq 0$ ならば, $x^2 \leq y$ を満たす x の領域は $-\sqrt{y} \leq x \leq \sqrt{y}$ である. これから, $F_Y(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$ と表せる. 一方, $y < 0$ ならば, $x^2 \leq y$ を満たす x は存在しないので, $F_Y(y) = 0$ である.

$$(b) f_Y(y) = \frac{d}{dy} (F_X(\sqrt{y}) - F_X(-\sqrt{y})) = \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y}))$$

(c) $F_Y(y) = 2P(0 \leq X \leq \sqrt{y}) = \frac{2}{\sqrt{2\pi}} \int_0^{\sqrt{y}} e^{-t^2/2} dt$. $F_Y(y)$ を y で微分することで得られる. $y > 0$ では, $f_Y(y) = \frac{2}{\sqrt{2\pi}} \frac{d}{dy} \int_0^{\sqrt{y}} e^{-t^2/2} dt = \frac{2}{\sqrt{2\pi}} \times \frac{1}{2\sqrt{y}} e^{-y/2} = \frac{e^{-y/2}}{\sqrt{2\pi y}}$, $y \leq 0$ では, $f_Y(y) = 0$ である. $f_Y(y)$ のグラフは以下のようなになる.

$x = 0$ で無限になるが, 当然 $\int_0^\infty f_Y(y) dy = 1$ である (不定積分は初等関数では表せないので, sagemath 等のソフトを用いて確かめよう).



(d) 平均 $E(Y) = 1$, 分散 $V(Y) = 2$ (これらも初等関数では表せないのでソフトを用いよう). 一方, 標準正規分布から $E(X) = 0$, $V(X) = 1$ なので, $E(Y) = E(X^2) \neq E(X)^2$ である.

173. 以下の場合に, Y の累積分布関数 $F_Y(y)$ と確率密度関数 $f_Y(y)$ を, X の累積分布

関数 $F_X(x)$ と確率密度関数 $f_X(x)$ を用いて表せ.

(a) $Y = |X|$

(b) $Y = e^X$

(c) $Y = -\log X$ (ただし, $X > 0$)

👉 $F_Y(y) = P(Y \leq y)$ に各小問の Y を代入して, X で表せばよい.

174. $Y = g(X)$ として, 以下の各 g に関して, Y の累積分布関数 $F_Y(y)$ を X の累積分布関数 $F_X(x)$ で表せ.

(a) X が指数分布 $e^{-\lambda x}$ ($\lambda > 0$) にしたがうとし, Y を X の整数部, つまり $g = [x]$

(b) $x \leq 0$ で $g(x) = -1$, $x > 0$ で $g(x) = 1$

(c) $x \geq 0$ で $g(x) = x + c$, $x < 0$ で $g(x) = x - c$. ただし, $c > 0$ とする. $g(x)$ は $x = 0$ で不連続である.

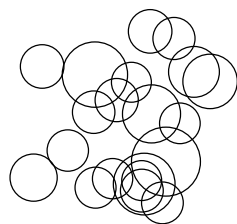
175. **円の面積の分布** 半径 X の円の面積 $Y = \pi X^2$ において, 半径 X が区間 $[0, 1]$ で一様に分布しているとする. 半径が一様分布したとしても, 下図から分かるように面積が小さい円が多いように見える?. このことを確かめるため, 面積 Y の分布に関して以下の問いに答えよ.

(a) 累積分布関数 $F_Y(y)$, 確率密度関数 $f_Y(y)$ を求めよ.

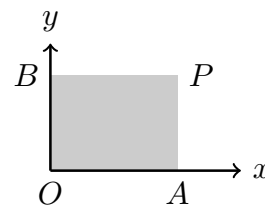
(b) 平均 $E(Y)$, 分散 $V(Y)$, 中央値を求めよ.

(c) $E(Y) = \pi E(X^2) \neq \pi E(X)^2$ を確認せよ.

👉 基本的には上の例題と同じ.



$L > 0$ とする. 正方形の領域 $\Omega = \{(x, y); 0 \leq x, y \leq L\}$ から, 無作為に 1 点 P を選び, x 軸と y 軸に垂線を下ろし, その足を A, B とする. 長方形 $OAPB$ の面積 S の平均値を求めよ.



◇ A の x 座標, B の y 座標を表す確率変数を X, Y とする. 長方形 $OAPB$ の面積 S は $S = XY$. P は Ω から無作為に選ばれるので, 事象 $E \in \Omega$ に対して, $P(E) = \frac{|E|}{|\Omega|} = \frac{|E|}{L^2}$. よって, $P(X \leq x, Y \leq y) = \frac{xy}{L^2}$ となる. 一方, $P(X \leq x) = \frac{xL}{L^2} = \frac{x}{L}$, $P(Y \leq y) = \frac{yL}{L^2} = \frac{y}{L}$ であるから, X, Y の独立性から, $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$ が成り立つ. よって, $E(S) = E(XY) = E(X)E(Y)$. X, Y は $[0, L]$ 上で一様なので, $E(X) = E(Y) = \frac{L}{2}$. し

$$\text{たがって, } E(S) = \frac{L^2}{4}$$

176. 整数 $1, 2, \dots, 49$ の中から 2 つ選び, それぞれ, 離散型確率変数 X と Y とする. ただし, 取り出した数は戻さないとする. このとき, $Z = \max\{X, Y\}$ の累積確率分布を求めよ.

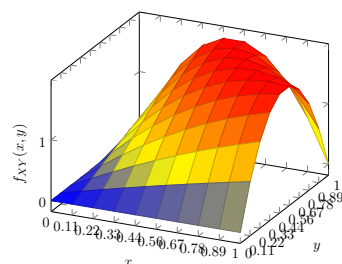
☞ 確率分布は $P(X = x, Y = y) = \frac{1}{49} \frac{1}{48} (x \neq y)$. 同じ数字は選ばれないので $P(X = x, Y = x) = 0$.

177. 同時確率密度関数が図のように $f_{XY}(x, y) = \frac{36}{5}xy(1 - xy) (0 \leq x, y \leq 1)$ と与えられている.

(a) X の周辺密度関数 $f_X(x)$ を求め, $E(X)$ を求めよ.

(b) X と Y の共分散 $\text{Cov}(X, Y)$ を求めよ.

☞ (a) $f_X(x)$ は $f_{XY}(x, y)$ を y について積分すればよい. (b) Cov のように x, y について積分するときは積分範囲に注意すること.



178. a, b ($0 < a, b < 1$) を定数とする. 離散型確率変数 X, Y の同時確率分布を $f_{XY}(x, y) = P(X = x, Y = y) = a^x b^y$ (x, y は自然数) とする.

(a) a, b の満たすべき条件を求めよ.

(b) $f_{XY}(x, y)$ の周辺確率分布から X, Y の独立性を調べよ.

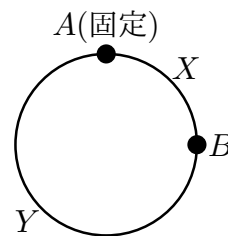
(c) $E(X), E(Y), E(XY)$ を計算して, X, Y の独立性を調べよ.

179. 周の長さ 1 の円の円周上に 1 点 A を固定し, 次に同じ円周上で一様に分布するように別の点 B を取る. それら 2 点で区切られてできる 2 つの弧 AB のうち, 短い方の長さを X とし, 長い方の長さを Y とする.

(a) X および Y の期待値と標準偏差をそれぞれ求めよ. また, X と Y の間の相関係数を求めよ.

(b) X と Y の比 $W = X/Y$ の累積分布関数 $F(w)$ および確率密度関数 $f(w)$ を求め, $F(w)$ と $f(w)$ の概形を描け.

(c) 比 W の期待値, 中央値を求めよ.



☞ 題意から X は $[0, 1/2]$ 上の一様分布. $X + Y = 1$

180. 区間 $[0, 1]$ 上の一様分布に従う互いに独立な確率変数 U, V, W に対し, α, β, γ を正の定数として, $X = U^\alpha, Y = V^\beta, Z = W^\gamma$ を考える.

(a) $\alpha = 2, \beta = 3$ のとき, $U = u$ が与えられた下での条件付き確率 $P(X > Y | U = u)$ を求めよ. これから $X > Y$ となる確率を求めよ.

(b) 任意の α, β, γ に対し, X が X, Y, Z の中で最大となる確率 $P(X = \max(X, Y, Z))$ を求めよ.

👉 (b) $X = \max(X, Y, Z)$ となるためには, $X > Y$ かつ $X > Z$ が成り立てばよい.

181. $X^2 + Y^2$ の分布 確率変数 X と Y が独立で, ともに $N(0, \sigma^2)$ にしたがうとき, 確率変数を $Z = X^2 + Y^2$ とする.

(a) $\frac{1}{2\pi\sigma^2} \iint_{x^2+y^2 \leq z} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) dx dy = 1 - \exp\left(-\frac{z}{2\sigma^2}\right)$ を証明せよ.

(b) 累積分布関数, 確率密度関数を求めよ.

(c) 平均, 分散を求めよ.

👉 これを n 個の変数に一般化したものがカイ 2 乗分布. (a) 極座標変換.

◆3 複雑な確率変数 (統計の基礎を与えるいろいろな分布)



182. 確率変数 $X_1, X_2, \dots, X_n$ が互いに独立で同じ標準正規分布 $N(0, 1)$ にしたがうとき, $X_1^2 + X_2^2 + \dots + X_n^2$ は自由度 n のカイ 2 乗分布にしたがう.

$$f(x) = \frac{1}{2^n \Gamma(n/2)} e^{-\frac{x}{2}} x^{\frac{n}{2}-1} (x \geq 0)$$

$x < 0$ では $f(x) = 0$ である.

確率変数 X が自由度 n のカイ 2 乗分布, 確率変数 Y が自由度 m のカイ 2 乗分布に従っているとする. また, X, Y は独立であるとする. このとき, 確率変数 $X + Y$ は自由度 $n + m$ のカイ 2 乗分布にしたがう.

183. 確率変数 X は標準正規分布 $N(0, 1)$ にしたがう, 確率変数 Y は自由度 n のカイ 2 乗分布に従うとする. また, X, Y は独立であるとする. このとき, $T = \frac{X}{\sqrt{Y/n}}$ は自由度 n の t 分布にしたがうことを照明せよ.

184. 確率変数 $X_1, X_2, \dots, X_n$ が互いに独立で同じ正規分布 $N(m, \sigma^2)$ にしたがうとき,

$$\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k, S = \sum_{k=1}^n (X_k - \bar{X})^2$$

とおくと, $\bar{X}$ と S は独立で, また, S/σ^2 は自由度 $n - 1$ のカイ 2 乗分布にしたがうことを照明せよ.

§7 モーメント, 母関数

モーメント: $m_k = E[X^k] = \int_b^a x^k f(x) dx$

$m = m_1$ (平均は 1 次モーメント), $\sigma^2 = m_2 - m^2$

母関数: $G(s) = E[s^X] = \int_{-\infty}^{\infty} f(x) s^x dx$

モーメント (積率) 母関数 : $M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} f(x)e^{sx} dx$

母関数とモーメント母関数の関係 : $M(t) = G(e^t)$

モーメント母関数によるモーメント : $m_k = M^{(k)}(0)$

母関数の独立性 : $E[s^{X+Y}] = E[s^X]E[s^Y]$, $E[e^{t(X+Y)}] = E[e^{tX}]E[e^{tY}]$

185. 母関数, モーメント母関数の基本的な使い方 母関数 $G(s)$, モーメント母関数 $M(t)$ が与えられたときに, 平均 $m = E[X]$, 分散 $\sigma^2 = \text{Var}(X)$ を求めるための以下の式を証明せよ.

(a) 母関数 : $m = E[X]$, $\sigma^2 = G^{(2)}(1) - m^2 + m$

(b) モーメント母関数 : $m = M'(0)$, $\sigma^2 = E[X^2] - E[X]^2 = M^{(2)}(0) - m^2 = m_2 - m^2$

186. 以下の場合の母関数, モーメント母関数を求めよ.

(a) 離散型一様分布 $p_n = \frac{1}{N}$ ($1 \leq n \leq N$) にしたがうときの母関数

(b) X が連続型一様分布 $f(x) = \frac{1}{b-a}$ ($a \leq x \leq b$) にしたがうときのモーメント母関数

(c) X が標準正規分布 $N(0, 1)$ にしたがうときのモーメント母関数

(d) X が 2 項分布 $B(n, p)$ にしたがうときのモーメント母関数

(e) X がパラメータ λ のポアソン分布にしたがうときのモーメント母関数

187. 0 か 1 の値をとる離散型確率変数 X において, それぞれの確率が $P(X = 1) = p$, $P(X = 0) = q = 1 - p$ と与えられている. モーメント母関数を導き, 平均と分散を求めよ.

188. X_k が 2 項分布に従うベルヌイ試行 $X_1, X_2, \dots, X_n$ で, $S_n = \sum_{k=1}^n X_k$ の母関数を導き, 平均と分散を求めよ. ただし, $P(X_k = 1) = p$, $P(X_k = -1) = q = 1 - p$ とする.

189. 正規分布の和は正規分布 X_1 が正規分布 $N(m_1, \sigma_1^2)$ にしたがうとき, X_2 が正規分布 $N(m_2, \sigma_2^2)$ にしたがうとき, X_1 と X_2 が独立であるとき, $X_1 + X_2$ は正規分布 $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ にしたがうことをモーメント母関数を用いて示せ.

☞ 正規分布のモーメント母関数 $M(t) = \exp\left(mt + \frac{1}{2}\sigma^2 t^2\right)$ に母関数の独立性を使う.

以下は少々難しい問題



190. 始めて続けて 2 回裏が出るまで, 硬貨を繰り返して投げる. 始めて続けて $n - 1$ 回表が出るまでに必要な硬貨投げの回数を X とする.

(a) 母関数 $G(s) = E[s^X]$ を n を用いて表せ.

(b) 平均を n を用いて表せ.

191. **【保険の請求】** 保険会社が、独立な N 個の保険請求 X_n ($n = 1, 2, \dots, N$) を受けると、請求の合計は $T = \sum_{n=1}^N X_n$ で表される. X_k は同じ分布にしたがい、その母関数を $G_X(s)$ 、また N に関する母関数を $G_N(s)$ とする. X_n と N がそれぞれ独立な確率変数として、保険請求の合計である確率変数 T の母関数 $E[s^T]$ を以下の手順にしたがって求めよ.

(a) N を固定した場合の T の母関数を $E[s^T|N]$ とすると、 $E[s^T|N] = E[s^X|N]^N = (G_X(s))^N$ と表せることを示せ.

(b) N が確率変数となると、 N についてさらに期待値をとり、 $E[s^T] = E[E[s^T|N]]$ とする. $E[s^T] = E[\{G_X(s)\}^N] = G_N(G_X(s))$ となることを示せ.

☞ (a) 独立性から $E[s^T|N] = E[s^{X_1+X_2+\dots+X_N}|N] = (E[s^{X_i}|N])^N = (G_X(s))^N$.

(b) N が確率変数になると、 N について期待値をとり $E[s^T] = E[E[s^T|N]]$ とする. 外側の括弧は N に関する期待値. なぜなら、 $G_N(s) = E[s^N]$ で $s = G_X(s)$ とおけるからである.

§ 8 応用問題



192. **【少子化 🧑 🧒 🧒 🧒 を考える】** 最近のデータ（政府統計 国民生活基礎調査「グラフでみる世帯の状況、児童の有無及び児童数別にみた世帯数の構成割合」）によると、各世帯の児童数の割合 (%) は以下のようになっている.

年	1 人	人 2	3 人	なし
2010	11.3	10.7	3.3	74.7
2013	10.9	10.1	3.2	75.9
2016	10.9	9.4	3.1	76.6
平均	11.0	10.1	3.2	75.7

以下の計算を簡単にするため、平均をとった値を用いる.

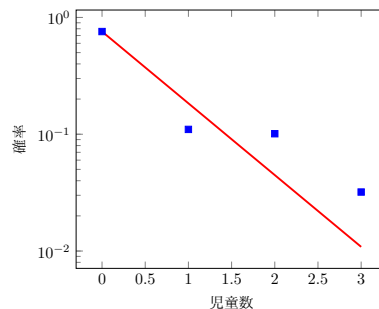
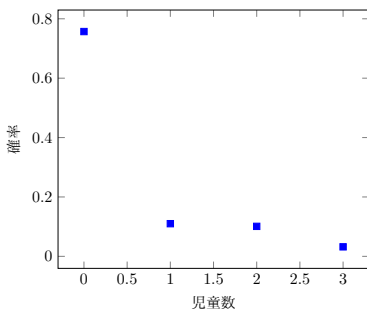
A_n : 児童数が $n(0, 1, 2, 3)$ の世帯という事象

としよう. 表の平均から、 $P(A_0) = 0.757, P(A_1) = 0.11, P(A_2) = 0.101, P(A_3) = 0.032$ である. 児童（何人でもいい）のいる世帯の確率を p とすると、

$$P(A_1) + P(A_2) + P(A_3) = p, P(A_0) = 1 - p$$

である. 下左図には $P(A_k)$ を、右図には片対数グラフ（縦軸は確率の対数 $\log P(A_k)$ ）を載せた.

(a) 右図の方対数グラフから、大雑把であるが、確率の対数は児童数に線形に依



存している様子が見えるので, $\log P(A_n) = \log a + bn$ と表せる. よって, $P(A_n) = ab^n$ と書ける事になる. ただし, $a > 0, 0 < b < 1$. a, b を用いて表せば, 1 世帯に子供が n 人いる確率 $P(A_n) = (1-p)p^n (n > 0, 0 < p < 1)$ となること示せ. またこの式の確率的な意味を与えよ.

子供が男である確率および女である確率を $1/2$ とする. ある世帯に男の子供が k 人いる確率 $Q(n, k)$ とすると, 男の子供が k 人いる確率は求めたい. これは $\sum_{n=k}^{\infty} Q(n, k)$ と表せる. 以下の問いに答えよ.

(b) $Q(n, k) = (1-p) \left(\frac{1}{2}\right)^n {}_n C_k$ となることを示せ.

(c) 母関数を $G(s) = \sum_{k=0}^n s^k (\sum_{n=k}^{\infty} Q(n, k))$ と定義する. このとき, $\sum_{n=k}^{\infty} Q(n, k) = \left(\frac{1}{k!} \frac{d^k G(s)}{ds^k} \right)_{s=0}$ となることを示せ. なお, $\sum_{k=0}^n \sum_{n=k}^{\infty} \rightarrow \sum_{n=0}^{\infty} \sum_{k=0}^n$ のように和を交換できることと, 二項定理 $\sum_{k=0}^n {}_n C_k s^k = (1+s)^n$ に注意せよ.

(d) $\sum_{n=k}^{\infty} Q(n, k) = \frac{(1-p)(p/2)^k}{(1-p/2)^{k+1}}$ となることを示せ.

(e) 上記の確率を最大にする p を k を用いて表せ. また, この結果について考察せよ.

☞ 確率の和を求めることになるが, (c) のように母関数が役立つ.

193. 【工場  の機械管理】 機械のある部品は故障した場合, すぐに新品と取り替えることにしていた. しかし, ある程度使った部品は壊れていなくても強制的に取り替えることによって, 管理が容易になり, コストも抑えられるのではないかと考えた. つまり, 取り付けてから t 経過した部品は壊れていなくても取り替える. その場合のコスト B は, 故障による取り替えのコスト A より, 少なくともすむという. この取り替え方法が良いか悪いか, 評価をしたい部品の寿命を X , その密度関数を $f(x)$ とする. 二つの確率変数 Z (部品の使用期間), (部品交換のコスト) W を下

のように定義する.

$$Z = \min\{t, X\}, \quad W = \begin{cases} A & (Z < t) \\ B & (Z = t) \end{cases}$$

- (a) Z の期待値を求めよ.
- (b) W の期待値を求めよ.
- (c) X がパラメータ λ の指数分布 $e^{-\lambda t}$ に従うと仮定し, $E(W)/E(Z)$ を最小にする t があるかどうか調べ, その意味を考察せよ.
194. 【コンビニ²⁴の店長】 売店であるスイーツを販売している. そのスイーツの流行が変わる前に売れそうな数を X とし, その確率分布を $F(x)$ とする. そのスイーツが1個売れると a 円の利益になるが, 流行が変わり売れ残ると b 円の損になる. X が大きいとし, 流行前後で同じ確率分布 $F(x)$, 確率密度関数 $f(x)$ で近似できるものと仮定する.
- (a) ネットの収益の期待値を $a, b, f(x)$ を用いて表せ.
- (b) 売店の店長は, ネットの収益を最大にするためにはどの程度仕入れればよいだろうか.
- (c) 結果に関して考察せよ.
195. 【くじを考案しよう⁷⁷⁷】 少し複雑なくじを考えよう. 各くじには 000000 から 999999 までの番号が付けてある. ある特定の番号を当選番号とすると, 当たりくじを引く確率は単純で, 10^{-6} である. しかし, 次のような当たりくじを考えると, どうなるだろうか. 上3桁の数字の和が, 下3桁の数字の和に等しいときに当たりくじとする. たとえば, 番号が 351108 のくじは, $3+5+1=1+0+8$ となるので, 当たりである. 当たりくじが, 結構多くでてきそうである. このように当たりくじを決めると, 当たりくじを引く確率がどの程度になるのか, 簡単には見積れないように思える. 当たりくじの確率がわからないと賞金も決められない. しかし, 母関数を用いると, 以下に示すように, 案外簡単に決められる.

§9 確率過程

簡単な例を通じて確率過程の概念を学ぼう. 確率変数が時間に依存する場合で, 離散時刻 n では X_n , あるいは連続時刻 t では X_t と表す. これらの時間的に変動する変数のことを サンプルパスと呼ぶ. 以下では, マルコフ連鎖, ポアソン過程, ランダムウォークの演習問題を扱う.

◆ 1 マルコフ過程

時刻 n で状態 X_n が j である確率: $p_n(j) = P(X_n = j)$

状態確率分布: $\mathbf{p}_n = (p_n(1), p_n(2), \dots, p_n(N))$

初期分布: $\mathbf{p}_0 = (p_0(1), p_0(2), \dots, p_0(N))$

時刻 $n-1$ で状態 i を時刻 n で状態 j に移す推移確率行列: $p_{ij} = P(X_n = j | X_{n-1} = i)$

$$p_n(j) = \sum_{i=1}^N P(X_{n-1} = i) P(X_n = j | X_{n-1} = i) = \sum_{i=1}^N p_{n-1}(i) p_{ij}$$

ここで、ベクトルである状態確率分布でまとめると、

$$\mathbf{p}_n(j) = \mathbf{p}_{n-1}(i) P$$

初期分布から始めると、

$$\mathbf{p}_n(j) = \mathbf{p}_0(i) P^n$$

$n \rightarrow \infty$ では、初期分布には無関係は定常分布（不変分布）になる、

$$\pi = \pi P$$

これまで扱ってきた漸化式の一種の応用と見なせなくはないが、多変数になっている点
が異なる。行列の復習から始めよう。

196. 2×2 の行列 $P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ に関して以下の問いに答えよ。

(a) 逆行列が存在するための条件を示し、逆行列を求めよ。

(b) 2つの固有値 λ_1, λ_2 , 固有ベクトル $\mathbf{x}_1, \mathbf{x}_2$ を求めよ。

(c) 固有ベクトルを並べた行列 $Q = (\mathbf{x}_1, \mathbf{x}_2)$ を用いて、 $PQ = Q \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ となることを示せ。

(d) $P^n = Q \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} Q^{-1}$ と表せることを示し、 P^n を求めよ。

👉 (a) 行列式, 逆行列, 固有値など行列の基本。

197. 上記問題と同じようにして、行列 $P = \begin{pmatrix} 3 & 1 \\ 2 & 2 \end{pmatrix}$ に関して以下の問いに答えよ。

(a) P の固有値, 固有ベクトルを求めよ。

(b) 固有ベクトルを並べた行列を Q とすれば、 $D = Q^{-1}PQ$ が対角行列になることを示せ。

(c) $P^n = QD^nQ^{-1}$ と表せることを利用し、 P^n を求めよ*1。

198. θ を任意の定数とし、行列 $A = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$ に関して以下の問いに答えよ。

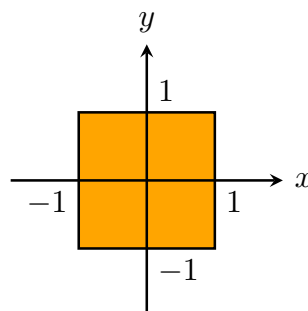
*1 Python では以下のコードで計算できる

- (a) A の固有値, 固有ベクトルを求めよ.
- (b) 固有ベクトルからなる行列を P とし, $D = P^{-1}AP$ が対角行列になることを示せ.
- (c) A^n を求め, その幾何学的な意味を述べよ.

👉 $\cos \theta + i \sin \theta = e^{i\theta}, (\cos \theta + i \sin \theta)^n = e^{ni\theta}$

199. 行列 $A = \begin{pmatrix} 3 & 1 \\ 2 & 2 \end{pmatrix}$ に関して以下の問いに答えよ.

- (a) $\mathbf{e}_1 = (2, 0), \mathbf{e}_2 = (0, 2)$ を 2 辺とする正方形 (面積=4) を線形写像 $f(\mathbf{x}) = A\mathbf{x}$ で写した先の平行四辺形の面積が, $4 \times |\det(A)|$ で与えられることを示せ.
- (c) 固有値, 固有ベクトルの幾何学的意味を考えるため, 図のような正方形の領域が A によってどのように変換させるか考察せよ. また, (d) の結果が正しいことを確認せよ.



👉 (a) 外積と平行四辺形の面積の関係.



推移確率行列の作成. ある部品の寿命の確率分布が $(0.1, 0.2, 0.3, 0.4)$ と与えられているとき, 推移確率行列, 状態遷移図がどのように表せるか. ここで, 0.1 とは取り替えて次の点検時に不良となる確率が 0.1, 0.2 とはさらに次の点検時に不良となる確率が 0.2 であることを意味する. 0.3, 0.4 も同様である,

取り替えてから 4 回目の点検までには必ず取り替える ($0.1+0.2+0.3+0.4 = 1$ から 4 回目の点検までには必ず故障する), 取り得る状態は 4 通り (ここでは, 状態 0,1,2,3 としよう) で, 推移確率は 4×4 の行列で表されることに注意.

◀点検直後の部品の年齢 (取り替えてからの経過日数) を X_n ($n = 0, 1, 2, 3$) とする. 推移確率は, 点検時に取り替えなかったとすると, その経過日数まで故障がなかったという条件の下で, 次の取り替えまでに故障が起きる確率なので,

```
from sympy import *
n=symbols('n')
P=Matrix([[3,1],[2,2]])
print(P**n)
```

条件付き確率を使わなければならない,

取替直後→1回点検: $p_{01} = P(1 = \text{良好} | 0 = \text{良好}) = 1 - 0.1 = 0.9$

1回点検→2回点検: $p_{12} = P(2 = \text{良好} | 1 = \text{良好}) = \frac{1 - (0.1 + 0.2)}{0.9} = \frac{0.7}{0.9}$

2回点検→3回点検: $p_{23} = P(3 = \text{良好} | 2 = \text{良好}) = \frac{1 - (0.1 + 0.2 + 0.3)}{0.7} = \frac{0.4}{0.7}$

(3回点検→4回点検: $p_{34} = P(4 = \text{良好} | 3 = \text{良好}) = \frac{1 - (0.1 + 0.2 + 0.3 + 0.4)}{0.4} =$

0) ←必要なし

$$p_{00} = P(0 = \text{良好} | 0 = \text{良好}) = 1 - p_{01}$$

$$p_{10} = P(0 = \text{良好} | 1 = \text{良好}) = 1 - p_{12}$$

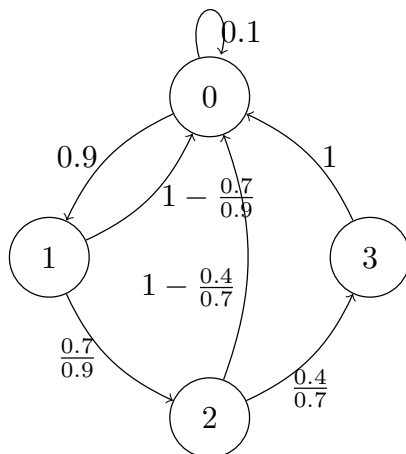
$$p_{20} = P(0 = \text{良好} | 2 = \text{良好}) = 1 - p_{23}$$

$$p_{30} = P(1 = \text{良好} | 3 = \text{良好}) = 1$$

以上の結果をまとめると, 推移確率行列は以下ようになる.

$$\begin{pmatrix} p_{00} & p_{01} & p_{02} & p_{03} \\ p_{10} & p_{11} & p_{12} & p_{13} \\ p_{20} & p_{21} & p_{22} & p_{23} \\ p_{30} & p_{31} & p_{32} & p_{33} \end{pmatrix} = \begin{pmatrix} 1 - 0.9 & 0.9 & 0 & 0 \\ 1 - \frac{0.7}{0.9} & 0 & \frac{0.7}{0.9} & 0 \\ 1 - \frac{0.4}{0.7} & 0 & 0 & \frac{0.4}{0.7} \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

状態遷移図は以下に示す. ただし, 0 となる推移確率は示していない.

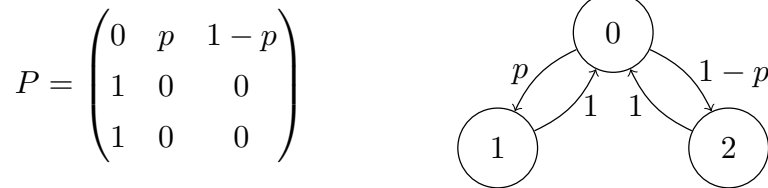


200. 先生が生徒に定期的に 3 種類の試験を課している. その評価は「できた」か「できなかった」かのいずれかであるものとする. 3 種類の試験ができたと評価される確率はそれぞれ, $p_1 = 0.3, p_2 = 0.6, p_3 = 0.9$ であるという. 前回の試験ができたとなると, 次の試験は 3 種類から等確率で選ばれた問題が出題される. もしできなかったとすると問題 1 が出題される. 問題 1, 2, 3 が出題される比率を求めよ.

$P(X_{n+1} = 1 | X_n = i)$ は「問題 i が不正解か ($1 - p_i$), 正解で次に問題 1 を選ぶ」確率 ($p_i/3$) に等しい. $P(X_{n+1} = j (\neq 1) | X_n = i)$ は「問題 i に正解し, 次に問

題 j を選ぶ確率に等しい ($p_i/3$)。従って
$$P = \begin{pmatrix} 0.8 & 0.1 & 0.1 \\ 0.6 & 0.2 & 0.2 \\ 0.4 & 0.3 & 0.3 \end{pmatrix}$$

201. $0 < p < 1$ として、次の推移確率行列で定義される 3 つの状態に関するマルコフ連鎖に関して以下の問いに答えよ。



- (a) 推移確率行列からわかる推移の特徴を述べよ。
 (b) 周期 2 を持つ周期的連鎖であることを示せ。また、状態 0 の再帰時間の分布を計算せよ。
 (c) 状態 1 および 2 の再帰時間の平均を p を用いて表せ。

➡ (a) 状態遷移図を書いてみよ。(b) 状態 0 にある初期状態を $X_0 = (1, 0, 0)$ において、 $X_0 P$, $X_0 P^2$ を計算せよ。周期 2 を持つというのは $X_0 P^2 = X_0$ を意味する。(c) 2 時刻後に戻る確率、4 時刻後に戻る確率など考える。

202. 以下の各小問はいずれも安定分布を求める問題である。

- (a) 従業員が N (十分多いとする) 人いる企業で、個々の従業員は 3 つのジョブクラスに分かれていて、クラス間をマルコフ連鎖にしたがってたがいに独立に移動しているものとする。その推移確率行列は

$$P = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.4 & 0.5 \end{pmatrix}$$

と与えられている。各クラスの従業員数の比率を計算せよ。

- (b) 道路を走っているトラックの後ろを走っている車は 4 台中 3 台は普通の乗用車、乗用車の後ろを走っている車は 5 台中 1 台がトラックであった。走っている乗用車とトラックの割合はどれくらいと考えればよいか。
 (c) ある町では 2 日続けて晴れたことがない。天気の種類は晴れ、曇り、雨の 3 通りとする。晴れの日次は曇りと雨が等確率で起き、曇り、雨の次の日は $1/2$ の確率で前日と同じ天気となり、天気が変わるとすれば残りの二つの可能性が等確率で起きるといふ、曇りの比率はどれくらいか。

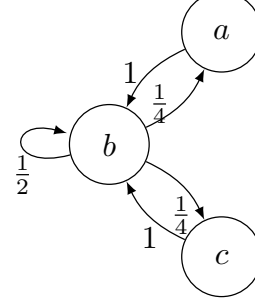
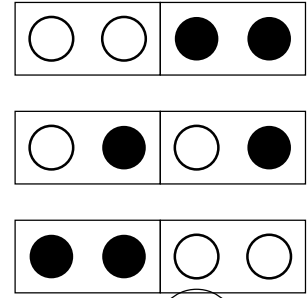
203. m 個の遺伝子グループを考える。この遺伝子は二つのタイプのいずれかである。ある代で j 個の遺伝子が I 型だったとして、次の世代で I 型遺伝子の個数が k 個になる確率は以下の式で与えられるものとする。、 $k = 0, 1, \dots, m$ 確率変数 X_n を n 世代目の I 型遺伝子の個数とし、 $X_0 = j$ とする。このとき (1) $E(X_n)$ を求めな

さい。(2) すべての遺伝子が I 型になってしまう確率はいくつですか。



白玉 2 個, 黒玉 2 個を 2 つの壺に 2 個ずつ入れる.

入れ方は, 図のように, 状態 a (白白)(黒黒), 状態 b (白黒)(白黒), 状態 c (黒黒)(白白) の 3 通りある. いま, 無作為にそれぞれの壺から玉 1 個を同時に取り出し他方の壺に入れる. この動作を n 回繰り返すとき, それぞれの壺の玉の状態の確率 $P_n(n = 1, 2, \dots)$ はどのように変化するであろうか. 最初の状態を (a, b, c) として以下の問いに答えよ.



- (a) 各状態の変化の仕方は右図の遷移図で表されることを示せ.
- (b) 遷移行列 P を求めよ.
- (c) $P_n = P^n$ を求めよ.
- (d) $\lim_{n \rightarrow \infty} P_n$ を求めよ.

◆(a) 省略

$$(b) P = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 \end{pmatrix}$$

(c) $P_n = P^n = QD^nQ^{-1}$. ここで, D は P を対角化した行列, Q は固有ベクトルを用いた行列である.

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}, Q = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -\frac{1}{2} & 1 \\ -1 & 0 & 1 \end{pmatrix}, Q^{-1} = \begin{pmatrix} \frac{1}{6} & \frac{1}{3} & -\frac{1}{2} \\ \frac{2}{3} & -\frac{2}{3} & 0 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} \end{pmatrix}$$

$$\begin{aligned} P_n &= P^n = QD^nQ^{-1} = Q \begin{pmatrix} 1 & 0 & 0 \\ 0 & (-\frac{1}{2})^n & 0 \\ 0 & 0 & 0 \end{pmatrix} Q^{-1} \\ &= \begin{pmatrix} \frac{1}{6} + \frac{1}{3}(-\frac{1}{2})^n & \frac{2}{3}(1 - (-\frac{1}{2})^n) & \frac{1}{6} + \frac{1}{3}(-\frac{1}{2})^n \\ \frac{1}{6} + \frac{1}{3}(-\frac{1}{2})^{n+1} & \frac{1}{3}(2 + (-\frac{1}{2})^n) & \frac{1}{6} + \frac{1}{3}(-\frac{1}{2})^{n+1} \\ \frac{1}{6} + \frac{1}{3}(-\frac{1}{2})^n & \frac{2}{3}(1 - (-\frac{1}{2})^n) & \frac{1}{6} + \frac{1}{3}(-\frac{1}{2})^n \end{pmatrix} \end{aligned}$$

$$(d) \lim_{n \rightarrow \infty} P_n = \begin{pmatrix} \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{pmatrix}$$

204. p を定数とし、2 状態の推移確率行列が $P = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}$ が与えられていると

する。以下の問いに答えよ。

(a) P の固有値、2 つの固有ベクトル $\mathbf{x}_1, \mathbf{x}_2$ を求めよ。

(b) 固有ベクトルを用いて、 $P = (\mathbf{x}_1, \mathbf{x}_2)Q(\mathbf{x}_1, \mathbf{x}_2)^{-1}$ のように書き換えよ、ここで、 Q は対角化行列である。

(c) n ステップ後の推移確率 P^n を p で表せ..

205. 【人の噂はあてにならない】 10% の割合で誤ったことを伝えるときに間違い可能性を取り入れて以下の行列を考える。

$$P = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$$

n 番目の人にのどのように伝わっているだろうか、以下の問いに答えよ。

(a) 各状態の変化の仕方は右図の遷移図で表されることを示せ。

(b) 遷移行列 P を求めよ。

(c) $P_n = P^n$ を求めよ。

(d) $\lim_{n \rightarrow \infty} P_n$ を求めよ。

206. 【未完】

◆2 ランダムウォーク (乱歩)

独立同分布する確率変数列 $X_1, X_2, \dots$ の部分和

$$S_n = X_1 + X_2 + \dots + X_n = S_{n-1} + X_n$$

$n = 1, 2, \dots$ $S_0 = 0$ をランダムウォークという。

(狭義の) ランダムウォーク: $X_1, X_2, \dots$ が 2 値の分布 (ベルヌイ分布) で、 $P(X = 1) = 1 - P(X = -1) = p$

$X_i = +1$ となる確率変数が $n - k$ 個、 $X_i = -1$ となる確率変数が k 個あるとき、 $S_n = n - 2k$ ($k = 1, 2, \dots, n$) となる。確率は $P(S_n = n - 2k) = \frac{nC_k}{2^n}$

197. ランダムウォークは日本語では乱歩と呼ばれているが、酔っ払いが左右にふらふら歩く様子にたとえられるように、予測がつかない様子を表す。つまり、予測が不能ということである。特に重要なのは株価の変動で、ランダムウォークであることが市場の公平性 (効率的と言う) の理論的な背景になっている。金融以外にも、いろ

いろな分野でランダムウォークが見られる。調べよ。

簡単な例から始めよう。

- (a) 数直線上を動く点 P が原点の位置にある。1 個のさいころを投げて、4 以下の目が出たときには P は正の向きに 1 だけ進み、他の目が出たときには P は負の向きに 1 だけ進む。さいころを 5 回続けて投げたとき、点 P の座標が 3 である確率を求めよ。
- (b) 点 P は初め数直線上の原点 O にあり、さいころを 1 回投げるごとに、3 の倍数の目が出たら数直線上を正の方向に 3、それ以外の目が出たら負の方向に 1 だけ進む。8 回さいころを投げるとき、点 P が原点 O にある確率を求めよ。

(a) 4 以下の目が x 回とすると、他の目は $5-x$ 回。 $x - (5-x) = 3$ より、 $x = 4$ になる。たとえば、点 P の座標が 3 になるのは $(1, 1, 1, 1, -1)$ のような場合。つまり 4 以下の目が 4 回、5, 6 が 1 回出ることになる。よって、 ${}_5C_4 \left(\frac{4}{6}\right)^4 \left(\frac{2}{6}\right)^{5-4} = \frac{80}{243}$ 。

(b) 同様にする。 $3+3-1-1-1-1-1-1=0$ 。 ${}_8C_2 \left(\frac{2}{6}\right)^2 \left(\frac{4}{6}\right)^{8-2} = \frac{1792}{6561}$ 。

198. 数直線上を動く点 P が原点の位置にある。1 個のさいころを投げて、6 の約数の目が出たときには点 P は正の向きに 2 だけ進み、それ以外の目が出たときには点 P は負の向きに 1 だけ進む。さいころを 6 回続けて投げたとき、点 P が原点の位置にある確率を求めよ。【2017 度山口県中】6 の約数の目は 1, 2, 3, 6 なので、正の向きに進む確率は $4/6$ 。さいころを 6 回続けて投げた原点に戻る $(+2, -1)$ の組合せの回数は $(2, 4)$ のみ。よって、 ${}_6C_2 \left(\frac{4}{6}\right)^2 \left(\frac{2}{6}\right)^{6-2} = \frac{20}{243}$ 。

p の（単純あるいは狭義の）ランダムウォーク $\{S_n, n = 0, 1, 2, \dots\}$ について、以下に答えよ。

- (a) $P(S_{2n+1} = 2k + 1)$ および $P(S_{2n+1} = k)$ を求めよ。
- (b) 期待値、分散を求めよ。

$2n+1$ 回の推移の内, 正の方向に移動するのが u 回, 負の方向に移動するのが d 回あったとすると, $u+d=2n+1$, $u-d=2k+1$ が成り立つので, これを解くと $u=n+k+1$ となるので, $P(S_{2n+1}=2k+1)=\frac{2n+1}{n+k+1}p^{n+k+1}(1-p^{n-k})$.

また, 偶数回の場合だけを取り出した $k=n, n-2, \dots, -n$ の場合は, $P(S_{2n+1}=k)=\frac{n}{(n+k)/2}p^{(n+k)/2}(1-p^{(n-k)/2})$

$S_n = S_{n-1} + X_n$ と, $E[X_n]=0$, $V[X_n]=1$ から,

$$E[S_n] = E[S_{n-1}] = \dots = E[S_0] = 0$$

$$V[S_n] = V[S_{n-1}] + 1 = \dots = V[S_0] + n = n$$

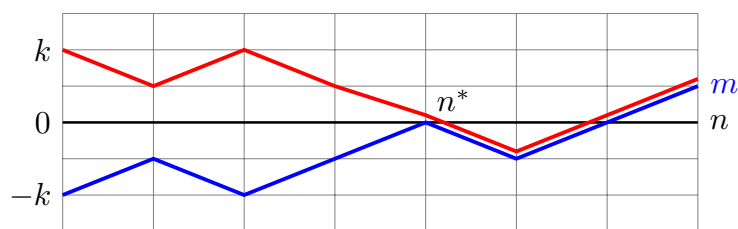
199. 【鏡像の原理】 対称な単純ランダムウォーク ($P(X_j=1)=P(X_j=-1)=\frac{1}{2}$) に関するいろいろな確率を計算する上で, 鏡像の原理という, 簡単で便利な方法がある. 使いこなせるようにしよう.

整数 k, m に対して, $(0, k)$ から (n, m) へ至るランダムウォークのパス (下図の赤線) の内で, x 軸を訪問する (途中で 0 になる) パスの本数は, $(0, -k)$ から (n, m) へ至るパス (青線) の本数は

$$nC_{(n+k+m)/2}$$

に等しい.

☞ 最初に原点を訪問するのが n^* 回目の推移とすると, $(0, k)$ から $(n^*, 0)$ へ至るパス (赤線) を x 軸に対称な位置に置き換えると, $(0, k)$ から (n, m) までのパスは $(0, -k)$ から (n, m) までのパス (青線) に一対一に対応するので, 両パスの数は等しい.



- (a) $(0, -k)$ から (n, m) までのパス (青線) は $nC_{(n+k+m)/2}$ 通りあることを証明せよ.
- (b) $n=7, k=2, m=1$ の場合はパスは何通りにあるか.
200. ランダムウォークのプログラムを作成してシミュレーションを実行しよう.

R では,

```
plot(cumsum(sample(c(-1,1),1000,replace=TRUE)),xlab="n",ylab="x(n)",type="S")
```

を実行すれば b グラフが作成できる,

201. 【原点回帰】 対称な単純ランダムウォーク ($P(X_j=1)=P(X_j=-1)=\frac{1}{2}$) において, $S_{2n} = X_1 + X_2 + \dots + X_{2n}$ とおく. ここで, 偶数回移動するときのみ

S_{2n} が 0 になる. このことを原点に回帰するという.

(a) $2n$ ステップ後に原点に回帰する ($S_{2n} = 0$) となる確率が $P(S_{2n} = 0) = \frac{2n C_n}{2^{2n}}$ (後で使うので, これを v_{2n} と定義する) となることを示せ.

(b) $P(S_{2n} = 2k) = \frac{2n C_{n+k}}{2^{2n}} - \frac{2n C_{n+k+1}}{2^{2n}}$ を示せ.

💡 鏡像の原理を適用しよう.

(c) $2n$ 回の推移の間, $2n$ も含めて, 負にならない (0 または正) 確率は, $2n$ 回目で原点を訪問する確率 v_{2n} と等しいことを証明せよ.

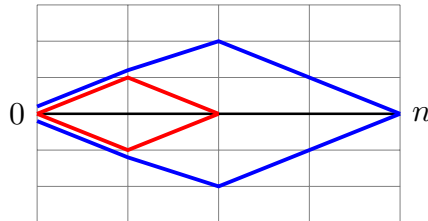
💡 負にならない事象を A とすると, これは S_{2n} は 0 か正の偶数 (最大 $2n$) で事象ある. 分配則 (加法定理) から $P(A) = \sum_{j=0}^n P(A|S_{2n} = 2j)P(S_{2n} = 2j) = \sum_{j=0}^n P(S_{2n} = 2j) = v_{2n}$ であることを示せばよい.

(d) n をいろいろ変えて, v_{2n} の値を調べよ.

202. 【再帰性】 対称な単純ランダムウォークに関して, 時刻 $2n$ に初めて原点へ戻る確率 f_{2n} を以下にしたがって求めよう.

(a) $n = 1$ (赤線), 2 (青線) の場合は簡単であるが, 初めて原点へ戻る確率を樹形図を描いて求めよ. $n = 3$ の場合はどうなるか.

$$f_2 = \boxed{} \quad f_4 = \boxed{} \quad f_6 = \boxed{}$$



(b) 時点 n で初めて状態 s に到達する確率は

$$\phi_s(n) = \frac{n-1 C_{(n+s-2)/2} - n-1 C_{(n+s)/2}}{2^n} = \frac{s}{n} \frac{n C_{(n+s)/2}}{2^n}$$

と表せることを示せ.

💡 時点 n で初めて状態 s に到達するパスは, 時点 $n-1$ では状態 $s-1$ にいなければならない. $(n-1, s-1) \rightarrow (n, s)$. 次に, $(0, 0)$ から $(n-1, s-1)$ へ至るパスの中で, 状態 s を訪問しないパスの本数を考える. ここでも, 鏡像の原理を適用しよう.

(c) 原点に初めて戻る時刻を T (偶数のみ) とし, $P(T = 2n)$ を f_{2n} と記すと

$$f_{2n} = \frac{1}{2}(\phi_1(2n-1) + \phi_1(2n-1)) = \phi_1(2n-1) \text{ と表せることを示せ.}$$

(d) $f_{2n} = \frac{1}{2^{2n}} \frac{(2n)!}{n!n!} \frac{1}{2n-1}$ となることを示せ.

(e) sagemath などを用いて, $\sum_{n=1}^{\infty} f_{2n} = 1$ となることを示せ*4. このことを再帰的という.

(f) n が大きいとき, スターリングの公式を使って, $f_{2n} = \frac{1}{\sqrt{\pi n}} \frac{1}{2n-1}$ と近似できることを示せ. また, この式の特徴を述べよ

203. 【母関数を用いた再帰性の証明】 母関数を用いると, 前問の再帰性の結果は案外簡単に導ける. ここでは対称なランダムウォークが特別な場合にあたる, 非対称なランダムウォーク ($P(X_j = 1) = p, P(X_j = -1) = q = 1 - p$) を考える.

(a) 各ステップでの距離を $X_1, X_2, X_3, \dots$ とすると, n ステップ後には原点 ($X_0 = 0$) から $S_n = \sum_{j=0}^n X_j$ に移動する. T_r を特定の位置 r に初めて到達するステップ数, 特に, T_0 は始めて原点に戻るステップ数 (偶数である) とする. このとき, ステップ数 r に関する母関数は,

$$E[s^{T_r}] = (E[s^{T_1}])^r$$

と表せることを示せ.

(b) 前問 (a) で,

$$E[s^{T_1}] = \frac{1 - \sqrt{1 - 4pqs^2}}{2qs}, \quad E[s^{T_0}] = 1 - \sqrt{1 - 4pqs^2}$$

となることを示せ.

(c) 対称なランダムウォーク ($p = \frac{1}{2}$) では原点に回帰する場合の母関数は $E[s^{T_0}] = 1 - \sqrt{1 - s^2}$ となる. これを s で展開すると, 原点に初めて戻る時刻 $T = 2n$ (偶数のみ) での確率 $f_{2n} = \frac{1}{2^{2n}} \frac{(2n)!}{n!n!} \frac{1}{2n-1}$ を与えることを示せ.

(d) 平均を求めよ.

204. リードを続ける確率 【逆正弦法則】

長さ $2n$ のランダムウォーク ($2n$ に原点に戻る) が, $2m$ 回正の領域に, 残りの $2n - 2m$ 回負の領域に存在する確率 $p_m(n)$ は, $p_m(n) = v_{2m}v_{2n-2m}$ と表せる. ここで, 正の領域にいる頻度と負の領域にいる頻度がどの程度違うか考えよう. 対称なランダムウォークでは普通, 正の領域正のにも負の領域のにも同程度にあるものと思いきや, 結果は意外なもので, パラドクスともいえるものである.

(a) n が大きいとき (十分長いランダムウォークということがある), スターリングの公式を使って, $v_{2m} \approx \frac{1}{\sqrt{\pi m}}$ と近似できることを示し, $p_m(n) = v_{2m}v_{2n-2m} \approx \frac{1}{\pi} \frac{1}{\sqrt{k(n-k)}}$ を導け.

*4 sagemath では以下のように入力すれば 1 になることが示せる.

```
var('n')
```

```
sum(2^(-2*n)*factorial(2*n)/(factorial(n)^2*(2*n-1)),n,1,oo)
```

(b) 正の領域にいる頻度を $x = \frac{m}{n}$ とする. したがって, 区間は $[0,1]$ となる. 十分長いランダムウォークにおいて, 正の領域に留まる相対頻度が x の確率密度関数 $P(x) = \frac{1}{\pi\sqrt{x(1-x)}}$ と表せる.

$$\frac{1}{\pi} \frac{1}{n} \frac{1}{\sqrt{\frac{k}{n}(1-\frac{k}{n})}} = \frac{1}{\pi\sqrt{x(1-x)}}$$

(c) $P(x < a) = \int_0^a \frac{1}{\pi\sqrt{x(1-x)}} dx = \frac{2}{\pi} \arcsin \sqrt{a}$ (逆正弦法則という)

(d) $\int_0^1 P(x) dx = 1$ を証明し, 平均, 分散を求めよ.

205. 🎲 【投票問題】 市長の信任投票を行ったら a 票対 b 票で信任された, つまり $a > b$ という前提で考える. このとき, 1 票ずつ開票するものとして, 信任票が常にリードを保ち続けて全投票の開票を終える確率を求めよ.

(a-i) (下図を参照): $a = 2, b = 1$ とする. A : 信任するという事象とする. 開票順は $AA\bar{A}, A\bar{A}A, \bar{A}AA$ の 3 通り (${}_{2+1}C_2 = 3$) であるが, 条件を満たすのは $AA\bar{A}$ の 1 通りだから, 確率は $\frac{1}{3}$ である.

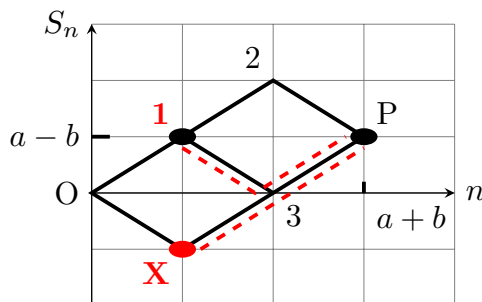
$${}_{2+1-1}C_{2-1} - {}_{2+1-1}C_2 = 2 - 1 = 1$$

$$1- > P : {}_{2+1-1}C_{2-1} \text{ 通り}$$

$X- > O- > 1- > 3- > P$ と $O- > 1- > 3- > P$ は同じ場合の数で,

$${}_{2+1-1}C_{2-1} = 1 \text{ 通り}$$

(a-ii) 一般の場合の確率を求めよ.



$A : \nearrow, \bar{A} : \searrow$

$AA\bar{A} : O12P$

$A\bar{A}A : O13P$

$\bar{A}AA : OX3P$

(b) $2n$ 回勝負して途中, 勝ち数が常に負け数を上回っていながら最後に引き分けで終わる確率は, $w_{2n} = \frac{1}{n} \frac{{}_{2n-2}C_{n-1}}{2^{2n-1}}$ で与えられることを証明しなさい.

ヒント 条件に当てはまるパスは, 点 $(1,1)$ から点 $(2n-1,1)$ を経由し, 時点 2 から時点 $2n-2$ までは 0 という状態を取ることはできません. 状態 0 を通るパスの本数は鏡像の原理によって数えることが出来ませんか.

206.

207. 【ギャンブラーの破産問題 🎲】

ギャンブラー X 氏は次のルールで賭を行っています:

(a) X 氏は賭け金 1, つまり, 勝てば +1, 負ければ -1 の損得' というゲームを繰り返す

(b) 各々のゲームで, X 氏の勝つ確率は $p(0 < p < 1)$

(c) X 氏が目標金額 a を獲得するか, 所持金が 0 になる (つまり破産する) 時点で賭は終了する

ギャンブラーの破産問題: (ギャンブラーの破産問題). X 氏は所持金 $n(0 < n < a)$ でゲームを開始した. X 氏が, 目標金額 a を獲得する前に破産する確率 $Q(n)$ を求めよ.

2 回目以降は確率 p で所持金 $n + 1$, 確率 $1 - p$ で所持金 $n - 1$ の状態 から賭を続けるので,

$$Q(n) = pQ(n + 1) + (1 - p)Q(n - 1)$$

となります. 結局 問題この式を満たす数列 $Q(n)$ を見つければ解ける.

- 208.** インフルエンザの感染モデル. n_0 個の赤い玉と, $N - n_0$ 個の白い玉を用意します. ランダムに二つの玉を取り出し, 二つの玉の色が同じならばそのまま元に戻し, さもなければ, 確率 q で白い玉を赤く塗り替えて元に戻すものとします. 言うまでもなく, 赤い玉が感染者を表し, 二つの玉を取り出すことはランダムに 2 人が出会うことを表し, 色を塗り替えるということ はある確率で感染するということを表しています. n_0 は最初の保菌者の数を表します. X_n を n 回の試行後の赤い玉の個数を表すものとする, $X_n; n = 0, 1, 2, \dots$ は吸収的マルコフ連鎖になることを示し, $X_0 = 1$ から始めて全員が感染するまでに何ステップかかるか計算しなさい.

第2章

統計

§ 1 統計の基本

◆ 1 度数分布，ヒストグラム，平均，分散，標準偏差，相関係数

平均（値）： $\bar{x} = \sum_{k=1}^n x_k = \frac{x_1 + x_2 + \dots + x_n}{n}$

分散： $s^2 = \sum_{k=1}^n (x_k - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}$

標準偏差： $s = \sqrt{s^2}$

重要な関係式： $s^2 = \bar{x}^2 - \bar{x}^2$

👉 R やエクセルでは，平均は変わらないが，不偏分散（共分散も同様）を用いているので，分母は n ではなく $n - 1$ になっている。

以下の**度数分布表（相対度数分布表）**は平成 22 年の国民・健康栄養調査（政府統計）に掲載されている収縮期（最高）血圧の分布である．度数分布（frequency distribution）に関連する用語はきちっと覚えておこう．

	人数	%
総数	1744	100
90mmHg 未満	1	0.1
90～99	16	0.9
100～109	100	5.7
110～119	219	12.6
120～129	351	20.1
130～139	380	21.8
140～149	327	18.8
150～159	197	11.3
160～169	77	4.4
170～179	45	2.6
180mmHg 以上	31	1.8
平均値 (mmHg)	135.3	
標準偏差	18.2	

階級，階級幅，階級数

例：階級幅 10(mmHg)，階級数 11

階級値：各階級を代表する値（真ん中の値）

例：120～130 の階級の階級値は $(120 + 130)/2 = 125$

度数：各階級に入っている個数

例：120～130 の階級の度数は 78（人）

ヒストグラム，度数折れ線

階級幅を横，度数を縦として得られる棒グラフ，棒の midpoint を順に結んで得られる折れ線



次のデータは8人の生徒に100点満点のテストを行った結果である。このデータの分散を求めよ。

75, 71, 79, 63, 83, 71, 87, 55 (点)

- (a) 平均を求めよ。
(b) 分散, 標準偏差を求めよ。

◇【2020 静岡県高より】(a) 仮平均を71とすると, 平均は $71 + \frac{4+0+8-8+12+0+16-16}{8} = 71 + \frac{16}{8} = 73$. (b) 分散は $\frac{(75-73)^2 + (71-73)^2 + \dots + (55-73)^2}{8} = 96$, 標準偏差は $\sqrt{96} = 4\sqrt{6}$

👉 大きい数字の場合は仮平均を使おう。

151. 平均, 標準偏差, 分散

- (a) 10人の生徒の右手の握力 $x(kg)$ のデータを次の通りとする。この時, 分散および標準偏差をそれぞれ求めよ。ただし, 小数第2位を四捨五入して答えよ。

22, 22, 23, 24, 25, 25, 26, 27, 27, 29

- (b) 次のデータは5人の生徒の計算テストを行った結果である。このデータの標準偏差を求めよ。ただし, $\sqrt{\quad}$ の中の数をもっと小さい自然数にせよ。

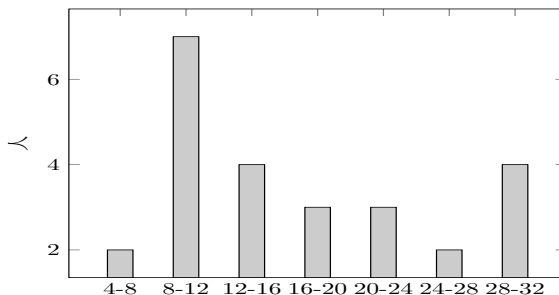
90点, 80点, 50点, 30点, 0点

- (c) 次の度数分布表について, 平均値, 分散, 標準偏差を求めよ。

階級 (以上未満)	40-50	30-40	20-30	10-20	0-10	計
度数	3	3	7	5	2	20

👉 度数分布の場合は階級値を用いる。

- (d) 下のヒストグラムは, ある中学校の生徒25人に実施したハンドボール投げの記録を表したものである。このヒストグラムから, 生徒25人のハンドボール投げの記録の平均値を求めよ。



- (e) 以下の49個のデータ x の標準偏差を以下の手順したがって答えよ。

階級 (以上未満)	55-60	60-65	65-70	70-75	75-80	80-85
度数	2	10	23	9	4	1

手順1 : $x = y + 67.5$ (仮平均を 67.5 とする)

手順2 : $y = 5z$ とおくと, 階級値 z は $-2, -1, 0, 1, 2, 3$ となる

手順3 : z の平均 $\bar{z}$ を求める

手順4 : x の分散 $= y$ の分散 $= 25 \times (z$ の分散) から, x の分散は $25(\overline{z^2} - \bar{z}^2)$

☞ ちょっと時間がかかるだろうが最後までやろう.

152. a は 0 以上の整数とする. データ 43, 53, 45, 49, 57, 41, 55, 51, 47, a について, 次の問いに答えよ.

(a) 平均値が 50 であるとき a の値を求めよ.

(b) a の値がわからないとき, 中央値をすべて求めよ.

☞ (b) a がどの位置になるかによって場合分けしよう.

153. 分散の関係式

(a) 20 個の僅かなるデータがある. そのうちの 5 個の値の平均値は 5, 分散は 15 であり, 残りの 15 個の値の平均値は 9, 分散は 19 である. このデータの平均値と分散を求めよ.

(b) 20 個の値からなるデータがあり, そのうち 12 個の値の平均値は 5, 分散は 3, 残りの 8 個の平均値は 7.5, 分散は 4 である. データの平均と分散を求めよ.

154. 与えられた平均, 分散などからデータ値を決める

(a) 4 個の値からなるデータがあり, このデータの最小値は 4, 最大値は 14, 平均値は 8 である. このとき, このデータの中央値を求めよ.

(b) 次の 10 個の値からなるデータの平均値が 5, 分散が 5 であるとき, a, b の値を求めよ. ただし, $a \leq b$ とする.

5, 2, 3, 8, 6, 3, 7, 6, a, b

(c) 5 個の値 1, 5, $a, 13, 19$ からなるデータの平均値が $a + 2$ であるとき, このデータの分散を求めなさい.

(d) 7 人の生徒の 1 学期間の読書冊数を調べたところ, $a, b, c, 1, 9, 11, 12$ (冊) で, 7 人の平均は 9 冊, 標準偏差は $\sqrt{7}$ 冊であった. 7 人の読書冊数を多い順に並べたとき, ちょうど真ん中は 10 冊であった. このとき, a の値はいくらか. ただし, $a < c$ とする.

☞ (a)(b) 平均が与えられているので $s^2 = \bar{x^2} - \bar{x}^2$ を用いるとよい. (d) 中央値の位置に注意すると b が決まる.

◆ 2 相関

散布図 : x と y の値の組を座標とする点を平面上にとったグラフ

共分散 : $s_{xy} = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{n}$

相関係数 : $r = \frac{s_{xy}}{s_x s_y}, -1 \leq r \leq 1$

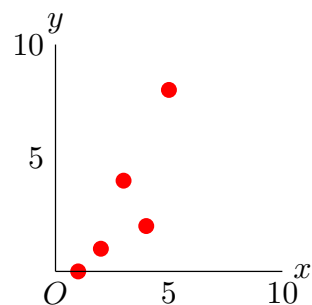
$r > 0$: 正の相関, $r < 0$: 負の相関, $|r|$ が小さい : 無相関



表は5人の生徒に10点満点の2種類のテスト A, B を行った得点の結果である. テスト A とテスト B の得点の相関係数を求めよ.

A の得点 x	5	1	3	4	2
B の得点 y	8	0	4	2	1

- (a) 散布図をかき, 図から何が言えるだろう.
 (b) 共分散, 相関係数を求めよ.



◇【2017 鳥取県高】(a) このような初歩的な問題であっても, ただ計算するだけでなく考察してほしい. 多量のデータを扱う実問題でも視覚的に特徴をとらえることは重要である. 傾きは正でかつ直線上にデータが並んでいるので, $+1$ に近い相関係数と思われる. (b) まず必要なのが平均と分散である. A の平均 3, 分散 2, B の平均 3, 分散 8 である. 共分散は $\frac{(5-3)(8-2)+(1-3)(0-2)+(3-3)(4-2)+(4-3)(2-2)+(2-3)(1-2)}{5} = \frac{17}{5}$ となるので, 相関係数は $\frac{17/5}{\sqrt{2}\sqrt{8}} = \frac{17}{20} = 0.85$

155. 以下の事項に関し, 横軸, 縦軸を明確にして, 正の相関, 負の相関, それとも無相関か, (理由を述べて) 判断せよ.

- (a) スマホの使用時間と数学の成績
 (b) 年齢と総コレステロール値の関係
 (c) 年代別の映画鑑賞率
 (d) パソコンの使用時間と視力

156. 相関係数の計算手順 2 つの変量 x, y が次の表で与えられるとき, 相関係数を求め, 相関の有無を答えよ.

x	1	2	3	4	5
y	0	4	1	2	3

数値計算する問題 (多量のデータをコンピュータで処理するのが統計学であるため大学入試ではほとんど見られないが, 高校や教員採用試験では簡単に計算できることを前提にした問題が課される) では, 丁寧に計算することが求められる. そのために, 以下の表を作成して数値を埋めるとよい. 表の一番下の行は合計を示すが, まず, x, y の合計から, 平均 $\bar{x}, \bar{y}$ を求めておく. 赤字の部分は常に 0 となるので, 計算のミスを防げる.

x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
1	0					
2	4					
3	1					
4	2					
5	8					
		0	0			

157. 以下の問いに答えよ.

- (a) 次の表は 2 つの変量 x, y についてのデータである. x と y の相関係数を小数第 3 位を四捨五入して答えよ.

x	4	2	5	1	8
y	7	3	9	6	10

- (b) 次の表は, あるクラスにおける生徒 5 人の英語と国語のテストの点数をまとめたものである. 英語と国語のテストの点数の相関係数を求めよ.

英語	78	79	81	77	80
数学	77	79	85	81	83

- (c) 5 人の生徒に数学と英語の小テストを行った. とともに 10 点満点である数学, 英語の得点をそれぞれ x, y で表すと, 以下の表のようになった. x と y の相関係数が 0.2 のとき, a の値を求めよ.

x	8	5	4	7	6
y	6	7	4	a	3

158. 次のデータの相関係数は -1 である. 表の一部を除き Y の値が虫食いになっている. 表の中の a, b の値はいくらか,

x	10	1	3	8	5	9	4	2	7	9
y		91		49	67			85	a	b

👉 相関係数は -1 から, Y と X の関係式がわかる.

◆ 3 四分位数と箱ひげ図

分布の範囲 (レンジ): 最大値-最小値

四分位数 (しぶんいすう): 小さい方から第1四分位数 Q_1 , 第2四分位数 (中央値) Q_2 , 第3四分位数 Q_3

四分位範囲: $Q_3 - Q_1$

四分位偏差: 四分位範囲を2で割った $\frac{Q_3 - Q_1}{2}$

159. 10個のデータ 45, 63, 23, 74, 76, 39, 33, 27, 55, 26 に関して以下の空欄を埋めよ.

最小値= , 最大値= , 分布の範囲=

第1四分位数 Q_1 = , 第2四分位数 (中央値) Q_2 =

第3四分位数 Q_3 =

四分位範囲= , 四分位偏差=

👉 まずは昇順にデータを並び替えよう.

160. **R と違う** 上問のデータを x とし, 世界で最もよく用いられ, 大学の講義科目にもある統計用のフリーソフト **R** を用いて計算すると以下ようになる. `summary(x)` あるいは `quantile(x)` と入力すると, 四分位数はすぐ分かる. 上問で求めた値と比べると, 中央値は同じだが他の四分位数は違う. なぜだろうか.

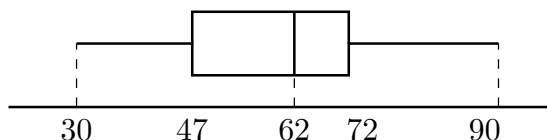
```
> x<-c(45, 63, 23, 74, 76, 39, 33, 27, 55, 26)
```

```
> summary(x)
```

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
23.0   28.5   42.0   46.1   61.0   76.0
```

161. 下の表は生徒8人の数学の得点である. ただし, $a < b < c$ を満たす. 下図は表をもとに作成した箱ひげ図である. このとき, a, c を求めよ. また, 平均値, 標準偏差を求めよ.

生徒	1	2	3	4	5	6	7	8
得点	b	43	92	a	57	68	51	c



162. 次のデータに関して問いに答えよ. ただし, x の値は0以上の整数とする.

68, 80, 74, 43, 56, 93, 64, x

このとき, 8個のデータの中央値のとり得る値は何通りあるか.

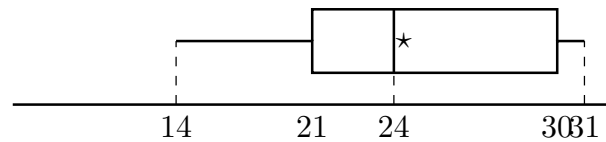
以下は少々難しい問題 R, エクセルを用いよう.



163. 代表値としての平均と中央値のどちらが適切であろうか. R あるいはエクセルを用いて, 以下の問題で検討せよ.

- (a) 以下のデータは 1990 年から 2016 年までの東京での桜🌸の開花時期を示す（出典：気象庁）。数値は 3 月の日付（1～31）である。

20, 30, 24, 24, 31, 31, 31, 21, 27, 24, 30, 23, 16, 27, 18, 31, 21, 20, 22, 21, 22, 28, 31, 16, 25, 23, 21



平均値（大まかな位置は箱ひげ図で星印★で示した）と標準偏差を求め，中央値と比較して分布の特徴を述べよ。

- (b) 以下のデータは 2017 年度の各都道府県の成人 1 人当たりのビール🍺の消費量（ L ）を多い順並べたものである（出典：国税庁課税 HP）。第 1 位は東京都，第 2 位は大阪府などである。箱ひげ図を書き。平均値と中央値を比較せよ。

3639.4, 2499.8, 2326.2, 2189.5, 2119.4, 2080.5, 2053.8, 1992.1, 1978.2, 1977.7, 1972.6, 1959.6, 1933.6, 1890.2, 1888.6, 1887.6, 1883.9, 1878.7, 1808.2, 1791.4, 1790.4, 1783.2, 1773.5, 1772.2, 1772, 1734.3, 1713.4, 1712.9, 1699.7, 1685.5, 1647.4, 1646.9, 1627.4, 1626.8, 1617.3, 1592.1, 1588.8, 1528.8, 1514.1, 1447.5, 1426.8, 1374.6, 1293.2, 1290.3, 1229.3

- (c) 以下のデータは 1959 年から 1980 年まで 22 間の元巨人軍の王貞治氏のホームラン数の記録である。箱ひげ図を書き。平均値と中央値を比較せよ。

7, 17, 13, 38, 40, 55, 42, 48, 47, 49, 44, 47, 39, 48, 51, 49, 33, 49, 50, 39, 33, 30

164. 実データの相関関係は複雑である

以下のデータは 2015 年の月別の気温とビール出荷量である（出典，気温は気象庁，ビールは Asahi スーパードライ月別出荷量（万箱））。

- (a) 正の相関，負の相関，無相関のどれか。
 (b) 相関係数を求めよ。
 (c) 暑いとビールの出荷量が増えると普通は思うが，相関係数は小さく。相関が認められない。理由を考え，その考えに基づいて改めと相関係数を求めよ。

月	1	2	3	4	5	6	7	8	9	10	11	12
気温	5.8	5.7	10.3	14.5	21.1	22.1	26.2	26.7	22.6	18.4	13.9	9.3
出荷量	462	645	776	843	853	958	1109	1037	768	810	766	1351

165. 疑似相関

給料が高いと血圧も高い

アイスクリームの売り上げが多い年ほど、溺死者が増える
 富裕層ほど朝型である
 この宝くじ売り場でよくあたりがでる
 ある高校が「東大合格者何名！」

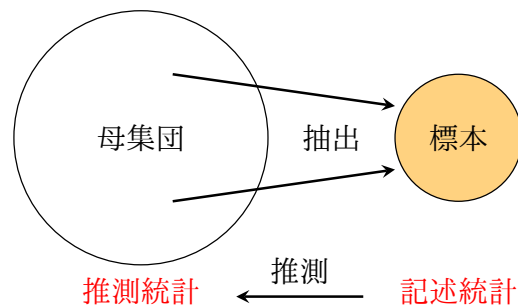
§2 標本と母集団

◆1 記述統計と推測統計（推計統計）

記述統計：標本の標本平均，標本分散など

推測統計：母集団の母平均，母分散などの区間推定と検定

171. 記述統計と推測統計の概要



172. まずは例から推測統計の流れを把握しよう。

Aさん 女優さんはスクリーンで見るとすごく大きく見えるし。実際背が高いのでしょうね。

Bさん でも一人一人の身長を調べると，確かに背が高い人もいるけど，案外そうでもない人もあるよね。

Aさん それじゃ。調べよう。といっても，アメリカに限っても数万人いるようだし，すべての女優さん【母集団】は調べられないよ。どうしたらいいのかな。

Bさん では，何人かの女優さんのデータを Wikipedia で調べ，女優全体の身長がどのくらいか推測してみよう。調べた女優さんは26人【標本】で，身長 (cm) は 161, 168, 164, 157, 175, 169, 168, 171.5, 175, 160, 160, 170, 167, 160, 160, 160, 173, 171, 170, 157, 170, 163, 155, 170, 159, 169



A さん 数万人いるのだから，ちょっと少ないような気がするし，どのように選んだのかな．

B さん これ以上調べるのは大変だし，自分が知っているか好きな女優さんだね．
この意味では偏った取り方をしているが，データ自体は偏りが無いような気がするね【無作為抽出（ランダムサンプリング）】．

では，標準的な手順で推測統計をおこなおう．

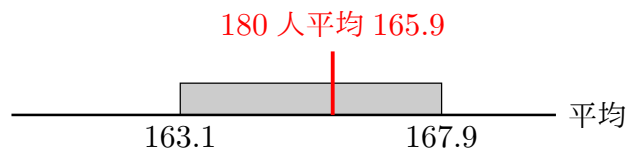
問題 1：26 名は全女優を代表できるか？

記述統計（標本）

26 名の身長平均 $\bar{x} = 165.5(\text{cm})$ ，標準偏差 $s = 6.00$

推測統計（母集団）：区間推定

母集団の母平均（未知である真の平均）の 95% 信頼区間は $\left[\bar{x} - 1.96 \frac{s}{\sqrt{26}}, \bar{x} + 1.96 \frac{s}{\sqrt{26}} \right] = [163.1, 167.9]$ で，95% の信頼度で母平均はこの区間内にある．



推測結果の妥当性：多くのデータが掲載されている「外国映画女優名鑑，共同通信社，2008」によると，180 人の平均（母平均ではないが）は 165.9 で，この区間内にある．

米疾病対策センターによると米国の成人女性の平均身長は 163cm だそうである．
ここで発想をかえて，女優の母平均を 163 と見なせるかどうか，考えてみよう．
 $\sqrt{2}$ が無理数かどうかを調べる背理法を用いる．

仮定： $\sqrt{2}$ は有理数 $\rightarrow$ 結果 $\rightarrow$ 矛盾 $\rightarrow \sqrt{2}$ は有理数でない

これを，統計では確率的に実行する．

問題 2：女優は成人女性より背が高いか？

記述統計（標本）

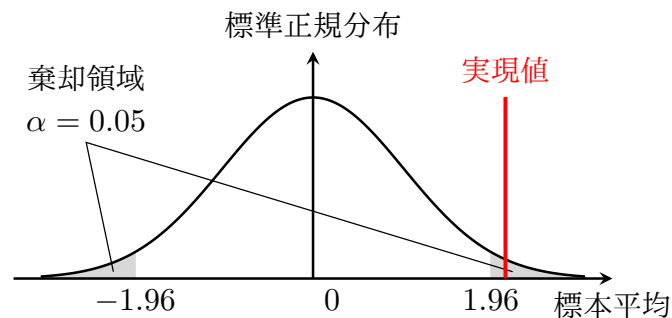
前と同じ

推測統計（母集団）：検定

帰無仮説：女優の平均身長は成人女性の平均身長と同じと仮定

帰無仮説からの結果：標本平均 $\bar{x}$ は $\frac{\bar{x} - 163}{s/\sqrt{26}}$ は標準正規分布 $N(0, 1)$ にした

がう．つまりは標本平均の 95% は $[-1.96, 1.96]$ に入る．



検定

検定統計量： $\frac{165.5 - 163}{6.0/\sqrt{26}} = 2.12$ 棄却域 $(-\infty, -1.96]$ あるいは $[1.96, \infty)$

にあり，**有意水準 5%** で帰無仮説を**棄却する**．女優の平均身長は成人女性の平均身長と同じでない．当然高い．

173. 以下の言葉を説明せよ．

- (a) 信頼度（信頼係数）
- (b) 信頼区間
- (c) 検定統計量
- (d) 有意水準
- (e) 帰無仮説と対立仮説
- (f) 棄却域と採択域

◆2 正規分布を用いた確率の計算に慣れよう．インターネット等で調べ，標準正規分布表を別途用意しておくこと．

174. 標準正規分布に従う確率変数を Z で表すと，標準正規分布 $N(0, 1)$ の確率密度関数は

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

と表せる．標準正規分布のグラフを描こう．図を参照．

(a) $f(z)$ は確率密度関数の条件, 特に $P(-\infty \leq Z \leq \infty) = \int_{-\infty}^{\infty} f(z)dz = 1$ を満たすことを証明せよ.

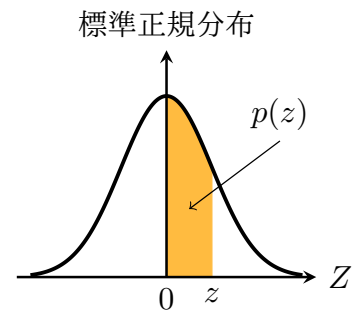
(b) グラフから, $z > 0$ のとき,

$$P(-z \leq Z \leq 0) = P(0 \leq Z \leq z)$$

$$P(Z \leq 0) = P(Z \geq 0) = \frac{1}{2}$$

となることが読み取れる. なぜか.

(c) 標準正規分布表は $p(z) = P(0 \leq Z \leq z)$ の値を載せている. いろいろな確率が必要になっても, これだけで十分である. なぜか.



➡ (a) そのままでは積分は初等関数で表せいが, 2 乗した関数の積分を考えよう. 大学院入試によく出る問題

175. 標準正規分布表の使い方に慣れよう 次の値を求めよ.

(a) $P(0 \leq Z \leq 1)$

(b) $P(1.41 \leq Z \leq 1.78)$

(c) $P(-2 \leq Z \leq 1.23)$

176. X を正規分布 $N(20, 4^2)$ に従う確率変数とする.

(a) $P(17.8 \leq X < 21.0)$ を求めよ.

(b) $P(X \leq c) = 0.8980$ となる c を求めよ.

177. ある企業の入社試験は採用枠 300 名のところ 500 名の応募があった. 試験の結果は 500 点満点の試験に対し平均点 245 点, 標準偏差 50 点であった. 得点の分布が正規分布であるとみなされるとき, 合格最低点はおよそ何点であるか. 小数点以下を切り上げて答えよ. ただし, 確率変数 Z が標準正規分布に従うとき, $P(Z > 0.25) = 0.4$, $P(Z > 0.5) = 0.3$, $P(Z > 0.54) = 0.2$ とする.

178. 以下の問いに答えよ.

(a) 得点が正規分布に従うある試験の平均点は 60 点, 標準偏差は 11 点であった. 38 点以上 93 点以下の人数が 404 人であったとき, 受験者の総数はいくらか.

(b) 確率変数 X が正規分布 $N(50, 10^2)$ に従うとき, $P(X \geq \alpha) = 0.025$ が成り立つような α を求めよ.

179. 近似に伴う補正 確率変数 X が二項分布に従うとき, その期待値を m , 分散を σ^2 とするとき, n が十分大きいときは近似的に正規分布 $N(m, \sigma^2)$ にしたがう. この近似を用いるときは補正が必要になる場合もある.

(a) 公平なコインを 100 回投げるとき, 表が 62 回以上出る確率を求めよ.

(b) 1 枚の硬貨を 400 回投げるときの表の出る回数を X とするとき, $P(|X - 200| \leq 20)$ を求めよ.

➡ 半目 (半整数) 補正が必要. たとえば, 5 以上の確率を計算としたいとすると, 正規分布

で近似する場合、4と5の間の4.5を用いることでより正確な値を求める。

180. ある大学では過去のデータによると入学試験の合格者の内、入学を辞退するものが4%いるという。1000人の定員のところ1050人合格としたとき、定員割れを起こす確率を二項分布の正規分布近似を用いて求めよ。

☞ 半目補正を忘れないように。

以下は少々難しい問題



181. 高校の3年生の10%はその身長が176cmを超え、15%は165cm以下である。身長分布が正規分布に従うとして男子高校3年生の身長の平均と標準偏差を求めよ。ただし、答えは小数第2位を四捨五入せよ。

☞ 平均と標準偏差が未知数。

182. 1000人がある大学の入学試験に挑んだ。300点満点で平均点は147点で分散は2025点であった。得点分布は正規分布にしたがうとして次の問いに答えよ。

(a) 192点を取ると偏差値はいくらか？ただし、

$$\text{偏差値} = 50 + 10 \frac{\text{得点} - \text{平均点}}{\text{標準偏差}} \text{ で与えられる。}$$

(b) 300人が合格するとして、次の4人の合否を判定せよ。

A : 124点, B : 147点, C : 169点, D : 194点

ただし、平均0、分散1の標準正規分布にしたがう確率変数 X に関して $P(0 \leq X \leq 0.5) = 0.1915$ を用いよ。

183. ある走り幅跳び選手の飛距離 X は平均6.5m、標準偏差0.2mの正規分布に従う。

(a) この選手の1回の飛距離が6.9mを超える確率を求めよ。

(b) この選手が3回跳ぶとき、3回中1回だけ飛距離が6.9mを超える確率を求めよ。

(c) 100回に1回、この選手が超えると期待される飛距離はいくらか。

☞ (a) 正規分布, (b) 二項分布, (c) 正規分布と補間。

§3 推定と検定

◆1 標本の統計量が母集団の母数の良い推定量であるための条件

【標本の統計量】 n 個のデータに関して、与えられたデータ $X_1, X_2, \dots, X_n$ から作られる関数 $T(X_1, X_2, \dots, X_n)$ を統計量という。また、推定したい量のことを母集団の母数という。

もっとも簡単で重要な例は標本平均、標本分散。

【統計量の一致性】 統計量 $T_n = T(X_1, X_2, \dots, X_n)$ が母集団の母数 θ の**一致推定量**であるとは、任意の $\epsilon > 0$ に対して、 $\lim_{n \rightarrow \infty} P(|T_n - \theta| < \epsilon) = 1$ となることである。

大きな標本での統計量（例：標本平均）は母集団の母数（例：母平均）に近づくこと。

【統計量の不偏性】 統計量 $T = T(X_1, X_2, \dots, X_n)$ が母集団の母数 θ の**不偏推定量**であるとは、 $E[T] = \theta$ となることである。

標本によって統計量（例：標本平均）は異なるが、あらゆるとり方が可能ならその平均（例：標本平均の期待値）は母集団の母数（例：母平均）に一致すること。

$$\text{標本平均 } \bar{X} = \frac{1}{n} \sum_{k=1}^n X_k : \text{一致性, 不偏性をもつ}$$

$$\text{標本分散 } S^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X})^2 : \text{一致性をもつが, 不偏性をもたない}$$

$$\text{不偏分散 } U^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2 : \text{一致性, 不偏性をもつ}$$

以上から、統計では標本平均と不偏分散を用いる。ただし、標本サイズ n が大きい場合は標本分散と不偏分散はほぼ等しい。これら以外に確率を最大にするという**最尤性**も有している。

184. 【平均の一致性と不偏性の証明】 $X_1, X_2, \dots, X_n$ は独立で同一の分布にしたがう確率変数で、平均と分散が有限、つまり、 $m = E(X_k)$, $V(X_k) \leq \sigma^2$, ($k = 1, 2, \dots, n$) と仮定する。標本平均を

$$T_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

とおき、この分布にしたがう母集団から無作為に抽出された多きさ n の標本の標本平均が一致性を有することを以下の手順で証明せよ。

- (a) 【大数の（弱）法則⇒一致性】 チェビシエフの不等式を利用して、任意の $\epsilon > 0$ に対して $P(|T_n - m| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$ を示せ。これから、 $\lim_{n \rightarrow \infty} P(|T_n - m| > \epsilon) = 0$ （大数の弱法則）が成り立ち、 T_n は一致性をもつことを示せ。
- (b) 平均の不偏性 $E(T_n) = m$ を示せ。

☞ (a) チェビシエフの不等式は、確率変数 Z に対して $P(|Z| > \epsilon) \leq \frac{E(|Z|^2)}{\epsilon^2}$ と書ける。大数の（強）法則 $\lim_{n \rightarrow \infty} P(T_n = m) = 1$ が成り立つためには、 T_n にさらなる条件が必要である（省略）。(b) 独立な X, Y に対して、 $E(aX + bY) = aE(X) + bE(Y)$ 。

185. さいころを用いて、平均の一致性と不偏性を視覚化して、理解を深めよう。

186. 【不偏分散の一致性と不偏性の証明】 独立な同一な分布にしたがう n 個の確率変数 $X_1, X_2, \dots, X_n$ において、標本平均を $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ とおく。不偏分散を

$$U_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2$$

とおく.

(a) チェビシエフの不等式を利用して, 任意の $\epsilon > 0$ に対して $P(|U_n^2 - \sigma^2| > \epsilon) \leq \frac{\sigma^2}{n\epsilon}$ を示せ. これから, $\lim_{n \rightarrow \infty} P(|U_n^2 - \sigma^2| > \epsilon) = 0$ が成り立ち, 不偏分散 U_n^2 は一貫性をもつことを示せ.

(b) 不偏分散の不偏性 $E(U_n^2) = \sigma^2$ を示せ.

☞ (a) n が大きい場合のことなので標本分散も不偏分散も一貫性をもつ. (b) 独立な X, Y に対して, $V(aX + bY) = a^2V(X) + b^2V(Y)$.

187. さいころを用いて, 不偏分散の一貫性と不偏性を視覚化して, 理解を深めよう.

188. 以下の場合の統計量はなにか.

(a) n を既知として, 二項分布 $B(n, p)$ から抽出した標本を X とする. 母数 p を推定するための統計量

(b) 医薬品 (経口製剤) の効き目は血液中最大になる濃度で決まるという. 先発医薬品と後発 (ジェネリック) 医薬品の性能を比較するための統計量

189. $a > 0$ とする. $[0, a]$ 上の一様分布にしたがう母集団から独立に抽出した大きさ n の標本 $X_1, X_2, \dots, X_n$ の標本平均 $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ は $\frac{a}{2}$ の不偏推定量であることを示せ.

190. X が母数 λ のポワソン分布にしたがうとする.

(a) X の不偏推定量を求めよ.

(b) どのような統計量がその不偏推定量として λ^k (k は正の整数) を与えるか.

☞ ポワソン分布は $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ で, 母関数 $G(S) = \sum_{k=0}^{\infty} P(X = k) s^k = e^{\lambda(s-1)}$ が役立つ.

§ 4 区間推定

◆ 1 区間推定の基礎

母平均 m , 母比率 p の推定. 標本平均を $\bar{x}$, 標本比率を $\bar{p}$ とする

(1) 正規母集団の母分散が既知か大標本の場合

標準化した変数 $Z = \frac{\bar{x} - m}{s/\sqrt{n}}$ は標準正規分布にしたがう.

$$\text{95\% 信頼区間} : \left[\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right]$$

$$\text{95\% 信頼区間} : \left[\bar{p} - 1.96 \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}, \bar{p} + 1.96 \sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \right]$$

90% の場合は 1.96 を 1.64 に置き換える.

(2) 正規母集団の母分散が未知の場合

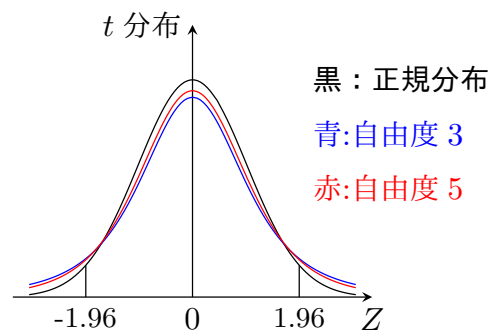
標準化した変数 $Z = \frac{\bar{x} - m}{U/\sqrt{n}}$ は自由度が $n-1$ の t 分布にしたがう.

$$\text{95\% 信頼区間} : \left[\bar{x} - t_{n-1}(0.05) \frac{U}{\sqrt{n}}, \bar{x} + t_{n-1}(0.05) \frac{U}{\sqrt{n}} \right]$$

小標本の例 : $t_{4-1}(0.05) = 3.18$, $t_{6-1}(0.05) = 2.57$, $t_{21-1}(0.05) = 2.09$

大標本 : $t_{n-1}(0.05) = 1.96$ としてよい.

標準正規分布, 自由度をいろいろ変えて t 分布を図示して, 比較しよう. 下図は自由度が 3 (青), 5 (赤) と標準正規分布との比較であるが, グラフとして正規分布 (黒) とあまり変わらない. つまり, 標本サイズが 20, 30 以上の実用的な場合にはさほど差はないのである.



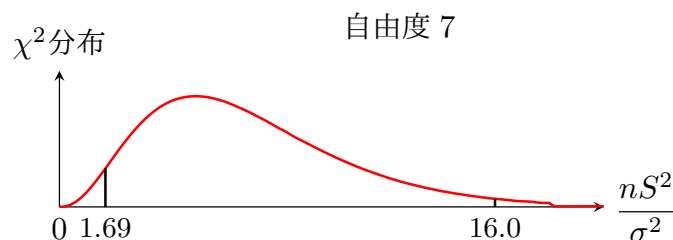
母分散 σ の推定. 標本平均を $\bar{x}$ とする

(1) 正規母集団の場合

変数 $\frac{nS^2}{\sigma^2} = \frac{(n-1)U^2}{\sigma^2}$ は自由度が $n-1$ の χ^2 分布にしたがう.

$$\text{95\% 信頼区間} : \left[\frac{nS^2}{\chi_{n-1}^2(0.05/2)}, \frac{nS^2}{\chi_{n-1}^2(1-0.05/2)} \right]$$

例 (下図): 自由度 7 の χ^2 分布, $\chi_7^2(0.05/2) = 1.69$, $\chi_7^2(1-0.05/2) = 16.0$



190. 【標準誤差】(Standard error) はエクセルで統計の計算する時 (データ分析で基本統計量を計算), 結果の項目を見ると, 平均の次に

$$\text{標準誤差} \frac{\sigma}{\sqrt{n}}$$

がある. 大標本では $U/\sqrt{n}$ となる. これが重要な意味を持つのは, 信頼区間の計

算にこの値が使われているからである.

(a) 標本平均 $\bar{X}$ の標準偏差は標準誤差 $\sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}}$ を与えることを示せ.

(b) 母標準偏差 $\sigma = 5$ の母集団から大きさ 25 の標本を無作為抽出するとき, 標準誤差は である.

(c) ある大学の男子学生の平均身長は 171cm, 標準偏差は 6cm である. この大学の男子学生 36 人を無作為抽出するとき, 標準誤差を求めよ.

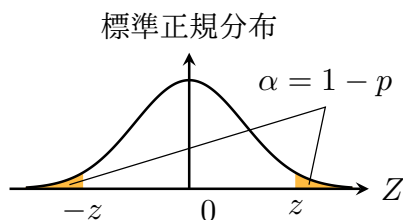
(d) エクセルを用いて, 1, 2, 3, 4, 5, 6 の標準誤差は である.

191. ある母集団の分散 σ^2 は既知であるとする. この母集団から無作為復元抽出で作った大きさ n の標本の平均値が $\bar{x}$ となったとする. このとき, 母平均 m の信頼区間は

$$\left[\bar{x} - z \frac{\sigma}{\sqrt{n}}, \bar{x} + z \frac{\sigma}{\sqrt{n}} \right]$$

と書ける. ここで, 信頼係数 95% の場合は, 信頼区間は $z = 1.96$ である. $P(|Z| \geq 1.96) = 1 - 0.95$. これを一般化して, $P(|Z| \geq z) = 1 - p$. p を与えたときに, これを満たすような z の値を計算用ソフトを用いて計算しよう.

$$\int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = \frac{1-p}{2}$$



p	0.2	0.15	0.1	0.05	0.01
$100(1-p)\%$	80	85	90	95	99
z の値				1.96	



【しばしば間違った捉え方をする重要な点】 母集団の母平均の区間推定を行なう問題で, 抽出した標本を使って計算した結果をレポートにまとめた.

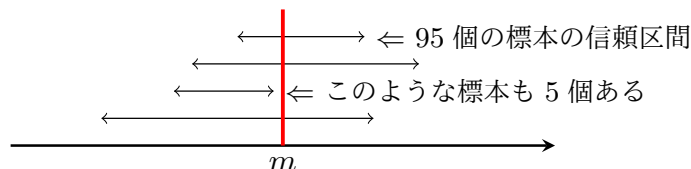
(1) 母平均の 95% 信頼区間は $[5, 8]$

(2) したがって, 母平均が $[5, 8]$ にある確率は 95% である

このレポートは区間推定の結果を正確に表現しているだろうか.

◇(1) に関しては計算が正確に行われていれば正しい. (2) これが問題であるが, なんとなく正しく聞こえてしまうので, ここで注意しておきたい. たとえば, 明日雨になる確率は 80% と言ったときには, 明日雨になるかどうか分からないので雨の可能性を確率として表している. ところが, 母平均を確率で表現することは正しくない. なぜなら, 母平均は未知ではあるが定まった値であり, 確率の対象ではな

い. 正しくは, 標本は抽出したデータに依存するので, 標本から計算した信頼区間も変わる. (2) は, たとえば無作為抽出を 100 回繰り返したとき, 母平均を含む信頼区間が 95 個ぐらいあることを表す (下図を参照). (1) だけをレポートにすればよい.



入荷した塩の袋のうちから, 25 個を無作為に抽出して重さを量ったところ, 標本平均が 297.4(g) で, 標本不偏分散は $7.5^2(g^2)$ であった. 塩の 1 袋の重さの平均を信頼度 95% で推定せよ.

(a) 母集団は未知のままを進める.

(1) 母集団は未知であるが, 大標本 ($n = 25$ は大きい) と見なす.

⇒ 標本平均: 中心極限定理から正規分布 $N(m, \sigma^2/n)$ で近似

⇒ 母分散 σ : 標本分散 $S \simeq U$ で置き換える

(b) 母集団は正規分布と仮定して進める.

(2) 正確に計算する (小標本でも OK).

⇒ 標本平均: $t(n-1)$ 分布にしたがう.

⇒ 母分散 σ : 標本分散 U で置き換える.

(3) 大標本と見なる.

⇒ 標本平均: 正規分布 $N(m, \sigma^2/n)$ で近似

⇒ 母分散 σ : 標本分散 $S \simeq U$ で置き換える

結局, 正規分布 (1)(3) を用いるか, t 分布 (2) を用いるかである. 実際の場合では大標本の正規分布を用いることが多い.

◇(1)(3) の場合: $U/\sqrt{n} = 7.5/\sqrt{25} = 1.5$ から, 95% 信頼区間は $[297.4 - 1.96 \times 1.5, 297.4 + 1.96 \times 1.5] = [294.5, 300.3]$. (2) の場合: $t_{25-1}(0.025) = 2.06$ から, 95% 信頼区間は $[297.4 - 2.06 \times 1.5, 297.4 + 2.06 \times 1.5] = [294.3, 300.5]$. 両者の違いをどのように見るかは, 現実に利用する場合での判断になる.

192. 【女優の身長】 前に述べた女優 26 人の身長 (cm) のデータ, 161, 168, 164, 157, 175, 169, 168, 171.5, 175, 160, 160, 170, 167, 160, 160, 160, 173, 171, 170, 157, 170, 163, 155, 170, 159, 169 の 95% 信頼区間を求めよ.

単峰性の分布だろうが, 正規分布といえる根拠はない. しかし, は 26

と小さくないので から 分布で近似する.

$$\bar{X} = \text{, } U^2 = \text{, } \frac{U}{\sqrt{n}} = \text{$$

から, 95% 信頼区間は,

$$\left[\bar{X} - 1.96 \frac{U}{\sqrt{n}}, \bar{X} + 1.96 \frac{U}{\sqrt{n}} \right] = \left[\text{, } \text{$$

193. ある錠剤 10 粒の重さは次のようであった.

2.37, 2.27, 2.25, 2.29, 2.33, 2.28, 2.31, 2.24, 2.29, 2.37 (g)

- (a) 10 個のデータの標本平均を求めよ. 単位もつけること.
 (b) 10 個のデータの標本分散を求めよ. 単位もつけること.
 (c) $X_1, X_2, \dots, X_{10}$ が正規分布 $N(m, \sigma^2)$ にしたがうとして,

$$Y = \frac{1}{\sigma^2} \sum_{k=1}^{10} (X_k - \bar{X})^2, \bar{X} = \frac{X_1 + X_2 + \dots + X_{10}}{10}$$

の満たす確率分布を求めよ.

- (d) 母分散の 99% の信頼区間を求めよ.

☞ (d) 自由度 9 の χ^2 分布にしたがう Z に対して, $P(Z > 23.5) = 0.005, P(Z < 1.73) = 0.005$ を用いよ.

194. ある国で内閣支持率を調査したところ 17.5% であった. 調査対象は 2000 人であった. 支持率の 90% 信頼区間を求めよ.

195. ある生産ラインで 1 万個の製品を作った. ランダムに選んだ 40 個の製品の平均重量は 156g であった. この生産ラインの機械的特性から, 生産される製品の重量の標準偏差は 8g である.

- (a) 生産した 1 万個の製品の平均重量の 95% 信頼区間を求めよ.
 (a) 95% 信頼区間の幅を 1g 以下にするためには何個の標本をとる必要があるだろうか.

☞ 信頼区間の幅とは信頼区間の上限と下限の差, あるいは $2 \times$ 標準誤差 である.

196. ある意見に対する賛成率は約 60% と予想されている. この意見に対する賛成率を, 信頼区間の幅が 4% 以下になるように推定したいとする. 信頼度 95% で推定するには, 何人以上抽出して調べればよいか.

197. 東葉高速線の西船橋駅での乗降客は 103,948 人/日で, そのうち他の路線に乗り換えるのは 27,162 人/日だそうである (出典: 平成 17 年大都市交通センサス), 駅で他の路線に乗り換えているのは 1 日あたり 1000 人中何人であるか, 信頼度 95% で推定せよ.

☞ このような実データを用いた例を探してみよう.

198. ある工場は「製品の強さが正規分布にしたがって分布するものとして, 製品の強さの平均値の 95% 信頼区間を求めたとき, その下限が 100kg 以上であること」とい

う条件の製品を受注した。そこで5個の試作品を作って強さを測定した結果, 100, 105, 105, 110, 135(kg) という結果を得た。ところが, 工場長はデータを 100, 105, 105, 105, 105(kg) のようにわざと低い数値に書き直して注文主に提出した。工場長はなぜこのように「良いデータをわざわざ悪く書き換える」ような不正を行なったのか, その動機を考えよ。上側 2.5% 点 $t_{0.025}(5-1) = 2.776$ を用いよ。

☞ 直感的には, 推定に依らずとも, 正規分布の仮定から, ヒストグラムが左右対称なので, 135kg という大きな値が現れている以上, 極端に小さな値も現れる可能性があるということ。また, 正規分布からは 135kg は滅多に起こらないので, 何かの異常あるいは不具合が疑われるため。

§ 5 検定

◆1 検定の基礎

検定の手順は以下ようになる

何をするかを定める: 母平均, 母比率など
 仮説を立てる: 帰無仮説 H_0 , 対立仮説 H_1
 標本調査を行う
 帰無仮説を棄却するか, 採択するかを決める
 結論をまとめる

199. 【背理法】 検定は(確率的な)背理法に基づいている。はじめに, 背理法について復習しておこう。

- (a) 背理法とは何か, 説明せよ。
- (b) $\sqrt{2}$ は無理数であることを証明せよ。

☞ むろん, 背理法以外にいろいろな証明法はある。

200. 二項分布 $B(n, p)$ は $n \rightarrow \infty$ で正規分布 $N(np, np(1-p))$ で近似できることを示せ。



以下の問いに答えよ.

- (a) 目の出方に偏りが無い 1 個のさいころを繰り返し n 回投げるとき $\square$ が出る回数を X とする. 期待値 $E(X)$, 分散 $V(X)$ をそれぞれ求めよ.

☞ 二項分布はしっかり身に付けよう.

- (b) さいころ A を繰り返し 720 回投げたところ $\square$ が 100 回出た. このさいころ A の目の出方は偏っていると判断してよいか, 有意水準 5% で検定せよ. ただし, 確率変数 Z が標準正規分布にしたがうとき, $P(|Z| \geq 1.96) = 0.05$ とする.

☞ 二項分布を正規分布で近似→検定, この一連の手順に慣れよう.

◇【2020 新潟県高】 (a) X は二項分布にしたがうので, $X \sim B\left(n, \frac{1}{6}\right)$. 期待値は

$$E(X) = n \times \frac{1}{6} = \frac{n}{6}, \text{ 分散は } V(X) = n \times \frac{1}{6} \times \left(1 - \frac{1}{6}\right) = \frac{5n}{36}$$

(b) 手順 0 (正規分布で近似): さいころ A を投げた 720 回のうち, $\square$ が出る回数 X は, 二項分布 $B\left(720, \frac{1}{6}\right)$ にしたがう. n が大きいので, 正規分布 $N(120, 100)$ で近似する. さらに, $Z = \frac{X - 120}{10}$ と標準化すると, Z は標準正規分布 $N(0, 1)$ にしたがう.

手順 1 (母数の検定): $\square$ が出る確率を p (母数) とし, 720 回中 100 回しかでないで, $p \neq \frac{1}{6}$ と思っているが確信がもてない. そこで, 仮に $p = \frac{1}{6}$ が正しいとして矛盾を導く (背理法).

手順 2 (仮説の設定): 帰無仮説 $H_0: p = \frac{1}{6}$, 対立仮説 $H_1: p \neq \frac{1}{6}$

・期待しているのは帰無仮説を否定することで, 対立仮説の出番はそうでない場合

・ p が小さい場合も大きい場合も考えるので, 両側検定を使う (下図も参照)

手順 3 (帰無仮説から導かれること): 帰無仮説 H_0 が正しいとすれば, さいころ A を投げた回数 X は, 標準化変数 $Z = \frac{X - 120}{10}$ は標準正規分布 $N(0, 1)$ にしたがう (帰無分布ということもある). これから言えることは

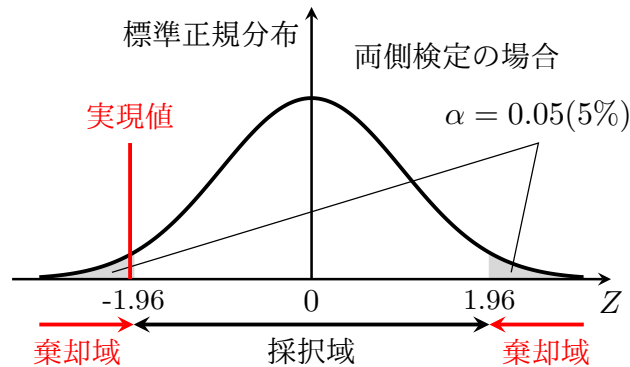
・ $P(|Z| \geq 1.96) = 0.05 \Rightarrow$ データの 5% は棄却域に入る

帰無仮説が正しいならこの区間には滅多に入らないので,

もし落ちることがあれば帰無仮説が正しくないを考える (棄却)

0.05(5%) のことを有意水準という.

・ $P(|Z| \leq 1.96) = 0.95 \Rightarrow$ データの 95% は採択域に入る



手順 4 (実現値を調べる) : 実現値は $X = 100$ で, $Z = \frac{100 - 120}{10} = -2$ である.

よって, 実現値は棄却域に落ちるので, 帰無仮説は棄却される (ギリギリだが).

手順 5 (検定の結論) : 有意水準 5% で帰無仮説は棄却される. つまり, さいころ A の目の出方は偏っているといえる.

(元の変数での棄却域は $X \geq 100.1, X \leq 139.5$)

201. コインを 100 回投げたとき, 表が 68 回出た. コインは公正といえるだろうか? 表の出る確率を p として, 有意水準 5% で検定せよ.

正規分布による近似: $N[\square, \square]$

仮説の設定: 帰無仮説 H_0 : $\square$, 対立仮説 H_1 : $\square$

標準化: $Z = \frac{p - \square}{\square}$

有意水準 5% で両側検定: 棄却域 $(-\infty, -1.96] \cup [1.96, \infty)$

実現値: $\frac{\bar{p} - \square}{\square} = \square$

検定の結論:

☞ 上の例題と同様に進め, 検定の方法に慣れよう.

202. 全国の小学校新入生男子の身長 (単位は cm) は正規分布に従い, 平均は 115.5, 標準偏差は 5.0 であることが知られている. ある地方で 16 人の小学生について平均身長を調べてみたところ, 118.4 であった. このデータから, この地方の平均身長は全国平均と異なるかどうかについて, 仮説検定することにする.

(a) 帰無仮説 H_0 および対立仮説 H_1 を求めよ.

(b) z 値を求めよ.

(c) 有意水準 5% で検定せよ.

203. ある占い師は透視能力があると主張している. そこで, 記号が書かれているカードと白紙のカードを 1 枚ずつ用意し, そのうちの 1 枚を裏向きにして透視する実験を 64 回行ったところ 40 回的中した. 透視能力があるかどうかは, 確率 0.5 以上だと主張を認めることにした場合, この占い師には透視能力があるといえるどうか, 有

有意水準 5% で片側仮説検定を用いて考察せよ.

☞ 片側の仮説検定の棄却領域: $P(x \leq X < \infty) = 0.05$ となるのは $x = 1.64$.

204. ある食料品の製造において, 1 袋 100g 中糖分が 2.0g になるように調整している. この工場の工程から, 糖分量の標準偏差は 0.6g であることが経験的に知られている. あるロットからランダムに 100 個の標本を選んで調査したところ, 糖分が平均 2.15g 含まれていた. この製造工程は正常に機能しているといえるか? 有意水準 5% で仮説検定を用いて考察せよ.

205. ある試験の全国平均は 66 点であった. A 塾から 10 名が受験した. 結果は

78, 72, 65, 86, 58, 64, 76, 88, 74, 59

であり, その平均点 72 点が 66 点を大きく上回ると A 塾は主張している. A 塾の主張は正しいか, 有意水準 5% で仮説検定せよ.

☞ データ数が少ない小標本の問題で, 不偏分散, t 分布を用いる.

以下は少々難しい問題



これまでの演習で, 区間推定, 検定が一通り使えるようになったと思うので. ここで, 平均の区間推定, 検定に関する理論的な背景について考えよう. 問題は, 標本の大きさによって, 標準化した確率変数が異なる分布にしたがうことによる.

正規分布を用いる

$$Z = \frac{\bar{X} - m}{\sigma/\sqrt{n}}$$

正規母集団

大標本

中心極限定理
 $n > 20 \sim 30$

t 分布を用いる

$$T = \frac{\bar{X} - m}{U/\sqrt{n}}$$

小標本

206. 【標本平均の分布: 母集団が正規分布の場合】積率母関数を利用し, 以下の手順にしたがって証明せよ.

- 正規母集団 $N(m, \sigma^2)$ からサイズ n (データ数) のデータ $X_k (k = 1, 2, \dots, n)$ を無作為に抽出したとき, 各変数の積率母関数は $M_X(t) = e^{mt + (t\sigma)^2/2}$ である.
- X_k が独立のとき, $S_n = \sum_{k=1}^n X_k$ の積率母関数は $(M_X(t))^n = e^{n(mt + (t\sigma)^2/2)}$ である.
- 積率母関数の性質から, S_n は, 平均 nm , 分散 $n\sigma^2$ の正規分布 $N(nm, n\sigma^2)$ にしたがう.
- 標本平均 $\bar{X} = \frac{S_n}{n} = \frac{1}{n} \sum_{k=1}^n X_k$ は正規分布 $N\left(m, \frac{\sigma^2}{n}\right)$ にしたがう.

(e) $Z = \frac{\bar{X} - m}{\sigma/\sqrt{n}}$ と標準化すると, $E(Z) = \frac{E(\bar{X} - m)}{\sigma^2/n} = 0$, $V(Z) = \frac{E((\bar{X} - m)^2)}{\sigma^2/n} - E(Z)^2 = 1$ となるので, Z は標準正規分布 $N(0, 1)$ にしたがう.

207. 【標本平均の分布 : (正規分布に限らない) 大標本の場合】 母集団が正規でない場合でも, 【中心極限定理】を利用すると, 標準化した変数は大標本ならば近似的に標準正規分布 $N(0, 1)$ にしたがう. このことを以下の手順にしたがって調べよう.

(a) X と Y が区間 $(0, 1)$ で一様なとき, $T_2 = \frac{X + Y}{2}$ の確率密度関数を求めよ.

$$f_2(t) = \int_{-\infty}^{\infty} f_X(2t - y)f_Y(y)dy = \begin{cases} 4t & (0 \leq t < \frac{1}{2}) \\ 4(1 - t) & (\frac{1}{2} \leq t \leq 1) \end{cases}$$

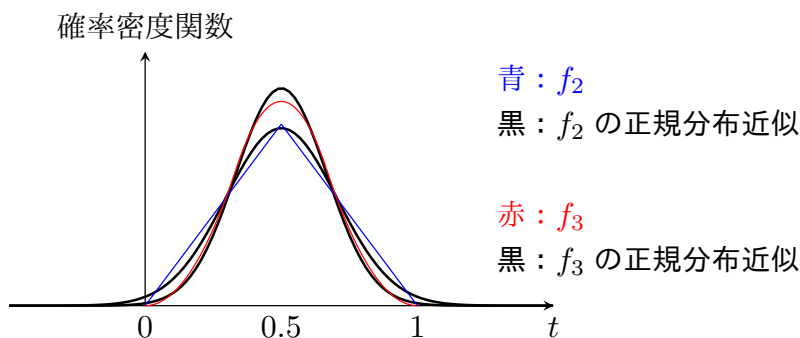
(b) $T_3 = \frac{X + Y + Z}{3}$ の確率密度関数を求めよ. これは少々難しいので, 結果のみを記しておく.

$$f_3(t) = \begin{cases} \frac{27}{2}t^2 & (0 \leq t < \frac{1}{3}) \\ -\frac{9}{2}(1 - 6t + 6t^2) & (\frac{1}{3} \leq t < \frac{2}{3}) \\ \frac{27}{2}(1 - t)^2 & (\frac{2}{3} \leq t \leq 1) \end{cases}$$

(c) 黒線で示したグラフは, T_2, T_3 を正規分布で近似たものである. その正規分布の平均と分散はそれぞれ,

$$E(T_2) = \boxed{}, V(T_2) = \boxed{}, E(T_3) = \boxed{}, V(T_3) = \boxed{}$$

である. $E(T_3) = \frac{1}{2}$, $V(T_3) = \frac{3}{12} \frac{1}{3^2} = \frac{1}{36}$ である. 図からわかるように, 3 個のデータでも正規分布でかなりよく近似できている.



中心極限定理とは

X_k が独立で同一の確率分布 (平均 m , 分散 σ^2) にしたがうとき, n が大きいならば (概ね 20, 30 以上), $S_n = \sum_{k=1}^n X_k$ は近似的に正規分布 $N(nm, n\sigma^2)$ にしたがう.

積率母関数を利用し, 以下の手順にしたがって証明しよう.

- (a) 平均 m , 分散 σ^2 の母集団からサイズ n (データ数) のデータ $X_k (k = 1, 2, \dots, n)$ を無作為に抽出する. 各変数の積率母関数を $M_X(t) = E(e^{tX})$ とする. $t = 0$ の周りでテーラー展開すると, $M_X(t) = M_X(0) + M'_X(0)\frac{t}{1!} + M''_X(0)\frac{t^2}{2!} + M'''_X(0)\frac{t^3}{3!} + \dots = 1 + m\frac{t}{1!} + (\sigma^2 + m^2)\frac{t^2}{2!} + M'''_X(0)\frac{t^3}{3!} + \dots$ となことを示せ.
- (b) X_k が独立のとき, $M_X(t)^n = \left(1 + mt + (\sigma^2 + m^2)\frac{t^2}{2}\right)^n = 1 + nmt + (n\sigma^2 + nm^2)\frac{t^2}{2} + \frac{n(n-1)}{2!}(mt)^2 = 1 + nmt + n(\sigma^2 + m^2 + m^2n - m^2)\frac{t^2}{2!} = 1 + nmt + (n\sigma^2 + m^2n^2)\frac{t^2}{2!}$
- (c) 正規分布の積率母関数は $e^{n(mt + (t\sigma)^2/2)}$ である. $M_X(t)^n$ の展開式を比較することで, $M_X(t)^n$ は平均 mn , 分散 $n\sigma^2$ の正規分布 $N(nm, n\sigma^2)$ になることを示せ.
- (d) こうして, n が大きいとき, 中心極限定理により, 標本平均は正規分布 $N\left(m, \frac{\sigma^2}{n}\right)$ にしたがうことが証明できた. あとは, 正規母集団の場合 (上記の問題) と同じで, 標準化した変数 Z は標準正規分布 $N(0, 1)$ にしたがう.

208. 【小標本, 母集団は平均 m , 分散 σ^2 の同一分布】 標本平均はこれまで通り $\bar{X}$ を用いる. 標本分散はこれまでの標本分散は $s^2 = \sum_{k=1}^n (X_k - \bar{X})^2/n$ にかわり, $U^2 = \sum_{k=1}^n (X_k - \bar{X})^2/(n-1)$ を用いる.

- (a) $E(\bar{X}) = m$, $E(s^2) = \frac{n-1}{n}$ となることを証明せよ.

$E(\bar{X}) = \frac{1}{n} \sum_{k=1}^n E(X_k) = (n \times m)/n = m$ となり, 標本平均の期待値 (標本平均の平均) は母集団の平均に一致する.

$$\begin{aligned} E(s^2) &= \frac{1}{n} E\left(\sum_{k=1}^n (X_k - \bar{X})^2\right) = \frac{1}{n} E\left(\sum_{k=1}^n (X_k - m - (\bar{X} - m))^2\right) = \\ &= \frac{1}{n} E\left(\sum_{k=1}^n ((X_k - m)^2 - (X_k - m)(\bar{X} - m) + (\bar{X} - m)^2)\right) = \frac{1}{n} E\left(\sum_{k=1}^n ((X_k - m)^2 - 2(X_k - m)(\bar{X} - m) + (\bar{X} - m)^2)\right) \\ &= \frac{1}{n} E\left(\sum_{k=1}^n ((X_k - m)^2) - E((\bar{X} - m)^2)\right) = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2 \end{aligned}$$

$U^2 = \frac{\sum_{k=1}^n (X_k - \bar{X})^2}{n-1}$ が不偏分散, つまり, $E(U^2) = \sigma^2$ 満たすこと証明せよ.

- (b) $\frac{\bar{X} - m}{U/\sqrt{n}} = \frac{\bar{X} - m}{\sigma/\sqrt{n}} \times \frac{1}{\sqrt{U^2/\sigma^2}}$ と変形すると, 前者の $\frac{\bar{X} - m}{\sigma/\sqrt{n}}$ は既に述べ

たように正規分布にしたがう．後者の分母 $\sqrt{U^2/\sigma^2}$ は前者と独立であり， $\sqrt{\chi_{n-1}^2/(n-1)}$ と同一の分布である．ここで， χ_{n-1}^2 は自由度 $n-1$ のカイ二乗分布を表す．したがって， $\frac{V}{\sqrt{U/(n-1)}}$ と同じ分布である．ただし， V と U は互いに独立で，それぞれ標準正規分布，自由度 $n-1$ のカイ二乗分布にしたがう．

以下の順序で証明していこう．

① $X \sim N(0, 1)$ として， $Y = X^2$ の確率密度関数を導く：問題 (b)

② $Y = X_1^2 + X_2^2 \dots X_n^2$ がカイ二乗分布：問題 (c)

③ $Z = \frac{X}{\sqrt{XY/n}}$ 自由度 n の t 分布：問題 (d)

(b) $X \sim N(0, 1)$ として， $Y = X^2$ の分布を調べる． $P(Y) = \int_a^b f_Y(y)dy = P(-\sqrt{b} \leq X \leq -\sqrt{a}) + P(\sqrt{a} \leq X \leq \sqrt{b}) = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{b}}^{-\sqrt{a}} f_X(x)dx + \frac{1}{\sqrt{2\pi}} \int_{\sqrt{a}}^{\sqrt{b}} f_X(x)dx$ ．ここで， $t = x^2$ とおく． $\int_a^b f_Y(y)dy = \frac{1}{\sqrt{2\pi}} \int_a^b \exp(-t/2)t^{-1/2}dt =$ となるので，両辺を b で微分し， $b = y$ とおけば， $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y/2} y^{-1/2} = \frac{1}{2^{1/2}\Gamma(1/2)} e^{-y/2} y^{-1/2}$ ．以上から， $X^2 \sim \chi^2(1)$

(c) 確率変数 X と Y がそれぞれ自由度 m と n のカイ二乗分布に従うとき， X と Y が独立であるならば，和 $X+Y$ は自由度 $m+n$ のカイ二乗分布にしたがう（証明略）． $Y = X_1^2 + X_2^2 \dots X_n^2 = \chi^2(1+1+\dots+1) = \chi^2(n)$

(d) $X \sim N(0, 1)$ と $Y \sim \chi^2(n)$ から， $Z = \frac{X}{\sqrt{XY/n}} \sim$ が示せる．

(c) χ_{n-1}^2 分布の証明

$\frac{U^2}{\sigma^2} = \frac{1}{\sigma^2} \frac{\sum_{k=1}^n (X_k - \bar{X})^2}{n-1}$ は χ_{n-1}^2 は自由度 $n-1$ のカイ二乗分布 χ_{n-1}^2 にしたがうことを証明せよ．

(d) t 分布の確率密度関数の導出

$\frac{\bar{X} - m}{U/\sqrt{n}}$ は自由度 $n-1$ の t 分布にしたがうことを証明せよ．

(e) $U^2 = \frac{\sum_{k=1}^n (X_k - \bar{X})^2}{n-1}$ である．標準化は，

$T = \frac{\bar{X} - m}{U/\sqrt{n}}$ ．このとき， T は以下の 自由度 $p = n-1$ の t 分布 (スチューデント分布) にしたがう．

$$f(t) = \frac{\Gamma(p+1)/2}{\sqrt{p\pi}\Gamma(p/2)} \left(1 + \frac{t^2}{p}\right)^{-\frac{p+1}{2}}$$

(b)

209. 【中心極限定理が成り立たない場合が多く見られるようになった：正規分布でなくとも検定はできる】 前に余震の時間間隔がべき関数で近似できることを述べた。これ以外に、保険，経済，インターネットなど様々な分野で必要になっている。べき関数の確率密度関数 $f(x) = ax^{-0.5}$ ($x \geq 1$) について調べる。

$$a = \boxed{}, m = \boxed{}, \sigma = \boxed{}$$

$$Q_1 = \boxed{}, Q_2(\text{中央値}) = \boxed{}, Q_3 = \boxed{}$$

$$\text{標準化 (mini max Z-score とも言う)} : \frac{X_C - Q_2}{(Q_3 - Q_1)/1.3489}$$

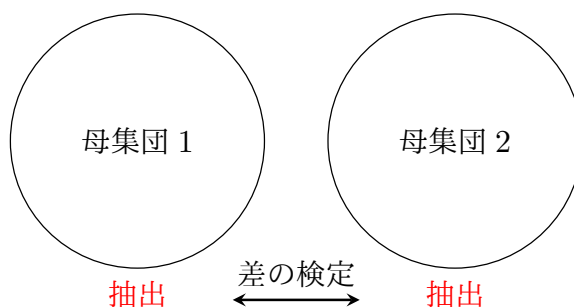
以上から平均値や分散は代表値にはなり得ないので，ここでは中央値を対象とする。標本サイズを $n = 20$ としよう。中央値を $Q2(X_1, X_2, \dots, X_n)$ とおく。

210. t 分布について調べよ。

◆2 母平均の差の推定と検定

母平均の検定の場合と同様に，状況によって以下の分類ができる。

- (1) 正規母集団で母分散が既知あるいは大標本の場合
- (2) 正規母集団で母分散が未知で等しい場合 (t 検定)
- (3) 正規母集団で母分散が未知で異なる場合 (ウェルチ検定)



(1) の場合：

$X_{11}, X_{12}, \dots, X_{1n_1}$ は互いに独立で $N(m_1, \sigma_1^2)$ にしたがう， $X_{21}, X_{22}, \dots, X_{2n_2}$ は互いに独立で $N(m_2, \sigma_2^2)$ にしたがうとき，標本平均 $\bar{X}_1$ は $N(m_1, \sigma_1^2/n_1)$ ，標本平均 $\bar{X}_2$ は $N(m_2, \sigma_2^2/n_2)$ の正規分布にしたがう。よって， $\bar{X}_1 - \bar{X}_2$ は $N(m_1 - m_2, \sigma_1^2/n_1 + \sigma_2^2/n_2)$ の正規分布にしたがう。これを標準化すると Z 値は

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (m_1 - m_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

で, $N(0, 1)$ にしたがう. 母平均の差の検定は, 通常 $m_1 = m_2$ のもとでの検定を行うが, 母平均の差が c の検定を行う場合は $m_1 - m_2 = c$ とする. Z 値は

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

ただし, 大標本で σ_1, σ_2 が未知の場合はそれぞれは U_1, U_2 で置き換える

帰無仮説 $H_0 : m_1 = m_2$, 対立仮説 $H_1 : m_1 \neq m_2$

有意水準 5% で両側検定: 棄却域 $(-\infty, -1.96] \cup [1.96, \infty)$

(2) の場合: 等分散なので $\sigma_1 = \sigma_2 = \sigma$ とおく.

$\bar{X}_1 - \bar{X}_2$ は $N(m_1 - m_2, \sigma^2(1/n_1 + 1/n_2))$ の正規分布にしたがう. ここで, $\frac{n_1 S_1^2}{\sigma^2}$ は $\chi^2(n_1 - 1)$ に, $\frac{n_2 S_2^2}{\sigma^2}$ は $\chi^2(n_2 - 1)$ にしたがうので, $\frac{n_1 S_1^2 + n_2 S_2^2}{\sigma^2}$ は $\chi^2(n_1 + n_2 - 2)$ にしたがう. よって, σ^2 の推定値は $\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$ となり,

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (m_1 - m_2)}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

は自由度 $n_1 + n_2 - 2$ の t 分布にしたがう.

帰無仮説 $H_0 : m_1 = m_2$, 対立仮説 $H_1 : m_1 \neq m_2$

有意水準 5% で両側検定:

棄却域 $(-\infty, -t_{n_1+n_2-2}(0.05/2)] \cup [t_{n_1+n_2-2}(1 - 0.05/2), \infty)$



2つの正規母集団から得られる2組のデータを比べる. 2台の機械 A, B の生産能力を比べる. A から $n_A = 10$ 個, B から $n_B = 9$ 個のデータをサンプリングした.

A : 78 80 79 83 82 85 78 74 76 84 (個/時間)

B : 81 84 82 88 86 83 78 84 89 (個/時間)

機械 A の母集団が $N(m_A, \sigma_A^2)$, 機械 B の母集団が $N(m_B, \sigma_B^2)$ にしたがうとする. ただし, σ_A, σ_B は未知である.

(a) 母平均の差の 95% 信頼区間を求めよ.

(b) また有意水準 5% で母平均に差 (m_A と m_B の値は等しいかどうか) があるかどうか検定せよ.

◀以下に示す手順に慣れよう.

(2) 正規母集団で母分散が未知で等しい ($\sigma_A = \sigma_B$) と仮定

$$\text{標本平均: } \bar{X}_A = \frac{799}{n_A} = 79.9, \bar{X}_B = \frac{755}{n_B} = 83.9, \bar{X}_A - \bar{X}_B = -\frac{359}{90} = -3.99$$

$$\text{標本分散: } S_A^2 = \frac{1149/10}{n_A}, S_B^2 = \frac{854/9}{n_B}$$

$$\text{分散の推定値: } \frac{n_A S_A^2 + n_B S_B^2}{n_A + n_B - 2} = \frac{1149/10 + 854/9}{17} = \frac{18881}{1530} = 12.34$$

$$t_{10+9-2}(0.05) = 2.11$$

信頼度 95% の区間推定

$$\text{信頼区間の下限: } (\bar{X}_A - \bar{X}_B) - t_{17}(0.05) \sqrt{\frac{18881}{1530} \left(\frac{1}{10} + \frac{1}{9} \right)} = -7.394$$

$$\text{信頼区間の上限: } (\bar{X}_A - \bar{X}_B) + t_{17}(0.05) \sqrt{\frac{18881}{1530} \left(\frac{1}{10} + \frac{1}{9} \right)} = -0.584$$

$$\text{母平均の差 } m_A - m_B \text{ の } 95\% \text{ 信頼区間は } [-7.394, -0.584]$$

有意水準 5% の t 検定

$$\text{帰無仮説 } H_0: m_A = m_B, \text{ 対立仮説 } H_1: m_A \neq m_B$$

$$\text{両側検定の棄却域: } (-\infty, -2.11] \cup [2.11, \infty)$$

$$\text{統計量の実現値: } T = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{18881}{1530} \left(\frac{1}{10} + \frac{1}{9} \right)}} = -2.471$$

棄却域に落ちるので有意水準 5% で帰無仮説は棄却される

つまり、機械 A、B の生産能力は同じでない。

(3) 正規母集団で母分散が未知で異なる ($\sigma_A \neq \sigma_B$) と仮定

$$\text{標本平均: } \bar{X}_A = \frac{799}{n_A} = 79.9, \bar{X}_B = \frac{755}{n_B} = 83.9, \bar{X}_A - \bar{X}_B = -\frac{359}{90} = -3.99$$

$$\text{不偏分散: } U_A^2 = \frac{1149/10}{n_A - 1}, U_B^2 = \frac{854/9}{n_B - 1}$$

$$\text{自由度 } d = \frac{\left(\frac{U_A}{10} + \frac{U_B}{9} \right)^2}{\frac{U_A^2}{n_A^2} \frac{1}{n_A - 1} + \frac{U_B^2}{n_B^2} \frac{1}{n_B - 1}} = 16.85$$

自由度は整数なので線形補間で、以下のように近似する。

$$t_d(0.05) = (17 - 16.85) \times t_{16}(0.05) + (16.85 - 16) \times t_{17}(0.05) = 2.111$$

信頼度 95% の区間推定

$$\text{信頼区間の下限: } (\bar{X}_A - \bar{X}_B) - t_d(0.05) \sqrt{\frac{U_A^2}{10} + \frac{U_B^2}{9}} = -7.3897$$

$$\text{信頼区間の上限: } (\bar{X}_A - \bar{X}_B) + t_d(0.05) \sqrt{\frac{U_A^2}{10} + \frac{U_B^2}{9}} = -0.588$$

母平均の差 $m_A - m_B$ の 95% 信頼区間は $[-7.390, -0.588]$

有意水準 5% のウェルチの t 検定

帰無仮説 $H_0 : m_A = m_B$, 対立仮説 $H_1 : m_A \neq m_B$

両側検定の棄却域: $(-\infty, -2.111] \cup [2.111, \infty)$

$$\text{統計量の実現値: } T = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{\frac{U_A^2}{10} + \frac{U_B^2}{9}}} = -2.476$$

棄却域に落ちるので有意水準 5% で帰無仮説は棄却される

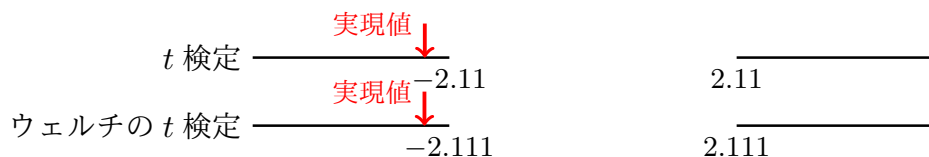
つまり, 機械 A, B の生産能力は同じでない.

以上の結果をまとめてみよう.

(参考) 正規分布で近似してみよう. つまり 1.96 を用いると, 母平均の差の 95% 信頼区間は $[-7.152, -0.825]$



中央の黒点は標本平均の差 $\bar{X}_A - \bar{X}_B$ の位置である. 3 つの信頼区間に大きな差が認められるでしょうか?



210. ある大学の A 学部と B 学部に英語の試験をし, 両学部から 50 人サンプルの学生を抽出した. A 学部は平均点が 360 点で標準偏差が 10 点, B 学部は平均点が 365 点で標準偏差が 95 点であった. 両学部の英語力は同じといってよいか? 両学部全体の標準偏差はわからないが, データは多いのでサンプルの学生のそれを使う. また, 危険率 (有意水準のこと) は 5% とせよ. ただし, $\frac{10^2}{50} + \frac{95^2}{50} = \frac{10025}{50}$ は 200 と近似し, その平方根は 14 と簡単化して答えよ.

☞ (1) の正規母集団で母分散が既知あるいは大標本の場合である.

大標本であるとして正規分布で計算する. つまり, $\bar{X}_A \sim N(m_A, \sigma_A^2/n_A)$, $\bar{X}_B \sim N(m_B, \sigma_B^2/n_B)$, $\bar{X}_A - \bar{X}_B \sim N(m_A - m_B, \sigma_A^2/n_A + \sigma_B^2/n_B)$ を用いることができる.

帰無仮説 $H_0 : m_A = m_B$, 対立仮説 $H_1 : m_A \neq m_B$

両側検定の棄却域： $(-\infty, -1.96] \cup [1.96, \infty)$

帰無仮説のもとで， $N(0, 1)$ にしたがう標準化変数は，

$$Z = \frac{\boxed{}}{\sqrt{\boxed{}}}$$

である．

標本平均はそれぞれ $\bar{X}_A = 360$, $\bar{X}_B = 365$, $\bar{X}_A - \bar{X}_B$ の標準偏差は $\sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B} = \sqrt{10^2/50 + 95^2/50} = \sqrt{10025/50} \simeq 14$ である．したがって，検定量 Z の実現値は

$$Z = \frac{\boxed{}}{\sqrt{\boxed{}}} = \frac{\boxed{}}{\boxed{}} = -0.357\dots$$

結論：棄却域に落ちないので，同じだという帰無仮説は否定できない（違いがあるとは言えない．）

211. 123

◆3 適合度の検定

$k \geq 2$ とする． $A_1, A_2, \dots, A_k$ が生起する確率を $p_1, p_2, \dots, p_k$ とする． $\sum_{j=1}^k p_j = 1$ を満たす． n 回この試行を独立に行って， $A_1, A_2, \dots, A_k$ が生じた回数を $X_1, X_2, \dots, X_n$ とすると，多項分布によって， $P(X_1 = n_1, X_2 = n_2, \dots, X_k = n_k) = \frac{n!}{n_1! n_2! \dots n_k!} p_1^{n_1} p_2^{n_2} \dots p_k^{n_k}$ である．ただし $n = n_1 + n_2 + \dots + n_k$ ．これにスターリングの公式を用いると，

$$P(X_1 = n_1, \dots, X_k = n_k) = \frac{1}{\sqrt{(2\pi)^{k-1} n^{k-1} p_1 \dots p_k}} \left(\frac{np_1}{n_1}\right)^{n_1 + \frac{1}{2}} \dots \left(\frac{np_k}{n_k}\right)^{n_k + \frac{1}{2}}$$

ここで， $x_j \sqrt{np_j} + np_j = n_j$ なる変換をし， $C = 1/\sqrt{(2\pi)^{k-1} n^{k-1} p_1 \dots p_k}$ とおくと，

$$P(X_1 = x_1 \sqrt{np_1} + np_1, \dots, X_k = x_k \sqrt{np_k} + np_k) = C \exp\left(-\frac{1}{2}(x_1^2 + \dots + x_k^2)\right)$$

となる．ここで，

$$\sum_{j=1}^k x_j \sqrt{np_j} = 0$$

が成り立つ．さらに， $Y_j = \frac{X_j - np_j}{\sqrt{np_j}}$ とすれば，

$$P(Y_1 = x_1, \dots, Y_k = x_k) = C \exp\left(-\frac{1}{2}(x_1^2 + \dots + x_{k-1}^2) - \frac{1}{2}\left(\sqrt{\frac{p_1}{p_k}}x_1 + \dots + \sqrt{\frac{p_{k-1}}{p_k}}x_{k-1}\right)^2\right)$$

となる．また，

$$z_1^2 + \dots + z_{k-1}^2 = x_1^2 + \dots + x_{k-1}^2 + \left(\sqrt{\frac{p_1}{p_k}}x_1 + \dots + \sqrt{\frac{p_{k-1}}{p_k}}x_{k-1}\right)^2$$

となるような新たな変数が導入でき，

$$P(Z_1 = z_1, \dots, Z_k = z_k) = D \exp \left(-\frac{1}{2}(z_1^2 + \dots + z_{k-1}^2) \right)$$

と表せる. こうして, 初めの統計量は n が十分に大きいと, 自由度 $k-1$ の χ^2 分布にしたがうことは分かる.

$$\text{統計検定量: 食い違い度} = \sum_{j=1}^k \frac{(\text{実現値}_j - \text{期待値}_j)^2}{\text{期待値}_j}$$

- 212. 【血液型】** 100 人の血液型を調べて, A 型が 52 人, B 型が 13 人, O 型が 27 人, AB 型が 8 人という調査結果を得た. 日本人の血液型は A 型が 40%, B 型が 20%, O 型が 30%, AB 型が 10% といわれているが, この調査結果はこの事実に適合していると言えるか, 危険率 5% で検定せよ. ただし, 自由度 3 の $\chi^2(3)$ 分布にしたが確率変数が 9.35 以上をとる確率は 0.025.

◆4 独立性の検定

独立性とは何か, 簡単な例を用いて説明していこう. ある製品を製造する工程が二つあり, 不良品が発生する. このとき, 工程に差があるだろうか. つまり不良品が発生する割合が工程によって違いがあるだろうか. 差がなければ, 工程 1 と工程 2 は関係がなく, 独立といえる.

	行程 1	行程 2	計
良品	a	b	a+b
不良品	c	d	c+d
計	a+c	b+d	n

$n = a + b + c + d$. 独立ならば, $P(\text{行程 1, 良品}) = P(\text{行程 1})P(\text{良品})$. つまり, $\frac{a}{n} = \frac{a+c}{n} \frac{a+b}{n}$ から, $ad - bc = 0$ が成り立つ. 直感的に言えば, $ad - bc$ が 0 に近いかどうかで独立性を検定する.

212. 123