「確率・統計」講義録
明治大学　2015 年度春学期　月 5 限・木 4 限

明治大学理工学部数学科　宮部賢志

2015 年 7 月 11 日

目次

前書き

　このノートは明治大学において 2015 年度春学期月 5 限・木 4 限に行う「確率・統計」の講義録です．予習復習などに利用してください．

休講情報

月 5 限ー 5 月 4 日，5 月 11 日，7 月 13 日は休講です．
木 4 限ー 5 月 7 日は休講です．

1　第 1 回　確率を学ぶ意味

1.1　オリエンテーション

この講義では「確率・統計」を学ぶ.

内容としては高校で学んだ確率と似ているところが多いが，全く異なるものとして理解してもらいたい．全体は大きく 3 つに分けられる.

(1) 離散確率
(2) 連続確率
(3) 統計

離散確率はサイコロの目やカードの柄などの確率について学ぶ．高校で学んだ計算に似ているが，確率の定義が異なることに注意する．連続的な確率変数とは，確率変数が実数など連続的に変化する値を取りうる確率である．本質的には離散確率と変わらないのだが，微分積分などの道具が使えるようになるため，この方が便利なことがある．例えば，1 時間あたり A 人の客が来る場合，平均的な待ち時間はどのくらいか？などの問題を考えることができる．統計とは確率の逆である．確率の場合，確率的な現象がどのように値をとるのかを考えるが，統計の場合，データが先に与えられてその確率分布を予想する，という考え方をする．そのため確率を理解していないと，統計を理解することは困難である．具体的な問題としては，誤差と有意な差との判定がある.

毎回課題および感想のレポートを提出してもらう．次の回までに採点をして返却する．課題は点数が悪かった場合，解き直して再提出しても良い．期末テストも行う．単位は課題 30% と期末テスト 70% を総合して判定する.

単位は努力賞ではない．目標のレベルに達しているかどうかで判定する．全員達していれば全員に単位を出すし，全員達していなければ全員に単位を出さない.

1.2　確率を学ぶ意味

なぜ確率を学ぶのか．統計を学ぶ意味は何か．高校で似たようなことを学んだからとか，必修だからとか，単位が楽そうだからとか，友達がいるからとか，そのような理由ではなく，学問としてどんな意味があるのかを今日は考えよう.

「これを学ぶとこういうことができる」「これを学ぶとこういう役に立つ」そのために勉強するという考え方もある．しかし，「これを理解していないと世の中の仕組みが理解できなくて困る」というものもある．確率や統計に関して言えば，後者の理由をいくらでも挙げることができる.

例 1.1. A 君はある予備校の模試を受けたら明治大学の合格確率は 10% で，志望校を変えた方が良いという判定が出た．そのほか 9 校でも同じように 10% と出た．とても落ち込んだが，よく考えると 10% は合格する確率があるのだから，10 校受ければ 1 校はきっと受かるだろうと考えた．親に頼み込んで 10 校の受験料を出してもらって受験したが 1 校も受からなかった．A 君は合格確率 10% と言った予備校に騙されたと思い，予備校を訴えることを考えている.

確率の概念について学んでいれば，このような不適切な怒りは起きてこないであろう．しかし本人は自分の苦しみが，勉強していないから起こっているとは思っていない．予備校が悪いと思っている．社会の仕組みを理解するということは，無駄な怒りや無駄な努力をなくすために必要なことだと思う.

例 1.2. 「原発の電源が落ちる確率は 0%」と言っていたのに，実際にそのようなことが起きたのだから，確率という学問は当てにならない.

だから確率は教えるべきでも研究すべきでもないというのだろうか？確率に関してその程度の理解の人が，国民の代表というのは問題ではないだろうか.

今回は自分の生活に関係がある，もしくは将来関係が出てきそうな事柄で，確率や統計の概念に関わる事柄をでき

るだけたくさん挙げてみよう.

以下，確率や統計についてのよくある疑問や誤解を挙げてみる.

例 1.3. どんな手術の成功確率も 50% である．成功するか，失敗するかのどちらかであるから.

例 1.4. 天気予報で降水確率が何 % などと言っているが，どういう意味だろうか？雨がふるかどうかは決まっていないのだろうか？決まっているのだろうか？

例 1.5. 「量子力学によるとこの世の中は確率的に決まるらしい．確率的に決まるのであれば，予測不可能である．将来が予測不可能であれば，今将来のために努力することは無意味である．だいたい未来がハッキリ分かってしまっていればつまらない．次何が来るか分からないワクワク感が大事だ．ゲームだってそうじゃないか．だから今楽しいことをすれば良い.」この考え方のどこが間違っているだろうか？

例 1.6. 賭けに勝つ確率が 50% ならば永遠に破産せずにゲームを続けられる.

例 1.7. コインの表と裏を大量に投げると，表と裏の比は 1:1 に近づいていく．なので，表が多く出ているときは裏が出やすく，表が少ない時には表が出やすい．これを大数の法則という.

例 1.8. これまで受けたテストの偏差値の平均は 55 なので，私は偏差値 55 の人間ということができる.

例 1.9. 視聴率を測った所，1 位と 2 位が同率になった．このようなことがあっては困るという要望を受けて，視聴率はより正確に小数点以下 5 桁まで発表することになった.

例 1.10. 曲のランダム再生で同じアーティストの曲が 3 曲連続で流れたので，これはランダムではないと思い Apple に抗議の手紙を送った.

例 1.11. 一部だけ検査してもたまたま正常なものだけにあたることもあるのだから，検査をするときには必ず全部を検査しなければならない.

例 1.12. 手元にあるデータから未来を予測するためには，まずデータに完全に合致するモデルを考え，そのモデルによる予測を考えるのが良い.

例 1.13. 機械には学習や成長ということはない.

2　第 2 回　必要となる数学

2.1　前回の感想から

「確率統計のは地味な単元というイメージがあった」応用や目標は華やかですが，そのために必要な数学を学ぶ作業は地味なものです．実際，今日からその地味な作業を行います．

「身の回りに確率に関係する事柄が多いことに驚いた」はい，本当にその通りです．

「確率なのか割合なのかよくわからなかった」「確率と一言でまとめてしまっているが，世の中には様々なタイプの確率が存在していると気づくことができた」確かにその通りで，確率とは何かについては第 3 回の授業で行います．

「さいころの出る目が 1/6 ではないと聞いて驚いた」これは多少誤解を招いているようですので，もう少し詳しく説明します．サイコロは本当に 1/6 出るのか？という疑問のもと，機械を作って実験をした人がいます．ウェルドンという統計学者は 1894 年に 12 個のサイコロを 2 万 5000 回投げて，5 や 6 が他の目よりも有意に出やすいことを発見しています．これは普通のサイコロは角がなかったり，目の窪みのためにわずかに重心がずれているからです．

それはそれとして，ではサイコロは確率的に出るのか？という問題もあります．確率的に出るということは，出る目が予想できないということですが，現実には，手を離した瞬間の初速度，位置，机の材質，サイコロの硬さ，などの情報がすべて集まればサイコロの目はかなり正確に予想できるはず．その意味でサイコロの目は確率的に出ているわけではない．サイコロの目が確率的に出ていると考えるのは 1 つの考え方でしかないです．この辺りの話は統計のところでもう少し詳しく話す予定です．

「0% は起こりうるかというのがまだよく分からない」例えば，ダーツで「この部分にあたる確率」というのは考えることができるが，「この点にあたる確率」というのは 0% でしかない．しかし現実にはどこかの点にあたるわけで，その意味で 0% のことが起こっている．これは離散的確率と連続的確率の大きな違いの 1 つ．原発の話はそれとはまた異なる話で，それについては次回．

「確率は言語で考えれば考えるほど疑問が増え解決しなくなってくるので，数学できちんと考え悩みを解決したい」その通りで，数学で語ることが重要です．

「あいまいな理解で中高と学習してきたのだと思い知らされました．」というより高校まではあいまいな教え方をしていたのです．

「将来研究するときにも何回か検証することがあると思うから，そのときにこれから学ぶことが役立てばいいと思う」その通りで，実験したり調査をしたりする時には，統計の知識が必須です．

「「統計」が名前からしてある情報を整理するだけだと思っていて，確率とは何の接点もないと思っていたのですが，そこから確率が出せることには驚きました．」「情報を整理する」というタイプの統計は，記述統計と言われます．それに対し確率と逆の関係にあると言った統計は，推測統計と言われます．記述統計は推測統計の基礎ですので，この講義では両方扱います．

「確率と統計は区別できるのか」これは非常に重要な点です．確率と統計を学ぶ時には，はっきりと区別して理解することが必要です．例えば，確率でも統計でも分散という概念が出てきますが，確率の分散と統計の分散は異なります．区別して理解しないと混乱します．それを現実に応用する場面においては，確率と統計の区別を曖昧にして使うことが多いです．だからこそ，混乱しそうになった時に，これはどういうことなのか，常に考えればわかるという状態にしておくことが必要です．

「確率は興味深い話のネタの宝庫だと自分は勝手にでるが信じています．なのでそのような確率にまつわるお話をしていただけるととても嬉しいです．」いくつかは用意するつもりでいます．面白さを理解するためには数学を理解する必要があるので頑張ってください．

「確率って大事なんだなあと思いました．確率ができなくて落ちた大学もあるので今身につけようと思います．」はい．大学受験などという小さな目標のためではなく，自分の成長のために頑張ってください．

「確率に反してそれが起こったりした時に，文句を言ったり，色々言う人は変だなと思いました」確率というのは難しいところで，逆に文句を言いたくてもいいにくいという欠点もあります．

「先生は確率の先生ということで，友達や家族とトランプ，麻雀をするときにもいちいち確率とか考えながらやって

いるんですか？それをやると勝率はやはり上がるのでしょうか？」確率の授業を持っている人がすべて確率の専門家ではありませんが，私は確率やランダムのことを研究しています．トランプはめったにやりませんが，麻雀をするときにはやはり切る牌は，大雑把な確率を考えます．しかしこれは麻雀の本にならたいてい書いてあることで，確率の専門家であるということは関係ないと思います．

「中学の先生が確率を使い宝くじで 70 万ほど当てていました当時の私は嘘だろうと思っていましたが，今日では本当かもしれないと思うようになりました.」いくら使って 70 万円を当てたのかが問題です．ギャンブルで負けない唯一の方法はギャンブルをやらないことです．

「期末テストは何点満点ですか？成績評価の通り 70 点満点ですか？」100 点満点ですが，後で 7 掛けします．期末テストの答案は返却できませんが，成績発表後に見に来ることはできます．

「統計は高校の時に課題研究で突然変異して飛べないハエの研究をしたときに，ハエをシャーレに入れて，シャーレの下に色紙を 2 枚しき，10 匹のハエがどちらに移動するかという実験で使いました．例えば色紙は赤と青にして赤に 4 匹，青に 6 匹いたとき，ハエは青に移動しやすいと言い切ることができるのか，という疑問の時に統計が役に立ちました．この実験の時は，64:36 以上の差が出た時に移動しやすいと言えるものでした.」まさにこういうのが統計です．

2.2 集合と論理

集合とはものの集まり．

a が集合 A に含まれている，集合 A の要素である，ことを $a \in A$ と書く．$A = \{1, 2, 3\}$ などとすべてを書き出す場合もあれば，$B = \{x \in \mathbb{R} \mid 0 < x\}$ のように満たすべき性質を書き表すこともある．

例 2.1. $A = \{1, 2, 3, 4, 5, 6\}$, $B = \{$ トランプの絵柄 $\}$, $\mathbb{N} = \{1, 2, 3, \cdots\}$, $\mathbb{R} = \{x \mid x$ は実数 $\}$

よく使う省略記号として，$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}$ や $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$ などがあり，それぞれ閉区間，開区間と呼ばれる．$[a, b)$ や $(a, b]$ は半開区間という．

集合 A のすべての要素が集合 B に含まれる時，A は B の部分集合であるといい，$A \subset B$ と書く．$x \in A$ であるとき，$\{x\} \subset A$ であり，逆も成り立つ．A を偶数の集合，$\mathbb{Z}$ を整数の集合とすると，$A \subset \mathbb{Z}$

A と B のすべての要素からなる集合を和集合 $A \cup B$ といい，A と B の共通要素からなる集合を積集合 $A \cap B$ という．

要素をまったく含まない集合を空集合 $\emptyset$ といい，$A \cap B = \emptyset$ であるとき，A と B は排反であるという．

全体集合 X が与えられた時，A に含まれない集合を A の補集合といい，A^c や $\overline{A}$ で表す．A に含まれるが B に含まれない要素を集めた集合を差集合といい，$A \setminus B$ と書く．

2.3 ランダウの記法

高校まででは，コインを 10 回投げてそのうちの 3 回表が出る確率，のような問題を考えていた．しかしこれからはもっとたくさん投げることを考える．例えば，1 万回投げるとか，100 万回投げるとか．考える確率も，表の出る回数が 1000 回から 2000 回の間となる確率，などのように幅をもたせて考えることがある．こうなってくると，「だいたいこのくらい」のような近似がよく使われる．n が十分大きい時や，x が十分小さい時に，何かと何かがだいたい同じ，という表現ができるようになろう．

例 2.2. $|x|$ が十分小さい時，$\log(1 + x) = x - \frac{x^2}{2} + O(x^3)$

この表現はどういう意味か．例えば，$x = 0.01$ のとき，$\log(1 + x) = 0.00995033085\cdots$ で，$x - \frac{x^2}{2} = 0.00995$, $x^3 = 0.000001$ であり，誤差が x^3 と同じくらいのオーダーであることがわかる．もう少し厳密に言えば，

$$\frac{\log(1 + x) - (x - \frac{x^2}{2})}{x^3}$$

という値が $x \to 0$ とした時に有限で収まることをいう．特に，$\lim_{x \to 0}$ が存在すれば，十分である．

上記の $\log(1 + x)$ の近似式はよく出てくるので覚えておこう．

2.4 級数と e の定義

$|r| < 1$ のとき,

$$\frac{1}{1-r} = 1 + r + r^2 + r^3 + \cdots$$

という式を高校で習ったであろう．このことから,

$$\frac{1}{1-r} = 1 + r + O(r^2)$$

などと書いたりする．ここで，r を $-r$ で置き換えると,

$$\frac{1}{1+r} = 1 - r + r^2 - r^3 + \cdots$$

となり，この両辺を積分すると,

$$\log(1+r) = r - \frac{r^2}{2} + \frac{r^3}{3} - \frac{r^4}{4} + \cdots$$

と先ほどの式が出てくる．これを項別積分という．いつこのようなことができるのかについては，微分積分の講義で学んでもらうことにして，今はこのような変形が可能であるということだけ知っておいて欲しい．

$$e^x = \exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

という式もよく出てくる．高校では,

$$e = \lim_{t \to 0} (1+t)^{1/t}$$

として習っているはずである．厳密な証明ではないが以下のように思うと納得できるかもしれない．

$$e^x = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^{nx} = \lim_{n \to \infty} \sum_{k=0}^{nx} {}_{nk}C_k \frac{1}{n^k} = \lim_{n \to \infty} \sum_{k=0}^{nx} \frac{nx \cdots (nx-k+1)}{nx \cdots nx} \frac{x^k}{k!} = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

ここで二項定理

$$(a+b)^n = \sum_{k=0}^{n} {}_nC_k a^k b^{n-k}$$

を利用していることに注意しよう．

2.5 積分計算

積分の計算として,

- $\int x^\alpha dx = \frac{1}{\alpha+1} x^{\alpha+1},\ \alpha \neq -1$
- $\int \frac{1}{x} dx = \log|x|$
- $\int e^{\alpha x} dx = \frac{1}{\alpha} e^{\alpha x}$
- $\int f(x)g(x)dx = F(x)g(x) - \int F(x)g'(x)dx$

などは思い出しておきたい．例えば,

$$\int \log x dx = x \log x - \int x (\log x)' dx = x \log x - x$$

などである．また広義積分

$$\int_a^\infty f(x)dx = \lim_{\alpha \to \infty} \int_a^\alpha f(x)dx$$

もよく使われる．例えば,

$$\int_0^\infty e^{-x} dx = \lim_{\alpha \to \infty} \int_0^\alpha e^{-x} dx = \lim_{\alpha \to \infty} [-e^{-x}]_0^\alpha = \lim_{\alpha \to \infty} (1 - e^{-\alpha}) = 1.$$

2.6 演習問題

問題 2.3. $I = [0, 1]$ を全体集合として，$A = [0, 1/2]$, $B = [0, 2/3]$ としたとき，$A \cup B$, $A \cap B$, A^c, $B \setminus A$ をそれぞれ求めよ.

問題 2.4. $\lambda > 0$ として，次の式を示せ.

$$\sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = 1, \qquad \sum_{x=0}^{\infty} \frac{xe^{-\lambda}\lambda^x}{x!} = \lambda.$$

問題 2.5. $0 < p < 1$ として，次の式を示せ.

$$\sum_{x=1}^{\infty} x(1-p)^{x-1}p = \frac{1}{p}$$

問題 2.6. $\lambda > 0$ として，次の式を示せ.

$$\int_0^{\infty} \lambda e^{-\lambda x} dx = 1, \qquad \int_0^{\infty} x\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

解答.

$$\int_0^{\infty} \lambda e^{-\lambda x} dx = \lim_{\alpha\to\infty} \int_0^{\alpha} \lambda e^{-\lambda x} dx = \lim_{\alpha\to\infty} [-e^{-\lambda x}]_0^{\alpha} = \lim_{\alpha\to\infty} 1 - e^{-\lambda\alpha} = 1.$$

$$\int_0^{\infty} x\lambda e^{-\lambda x} dx = [x \cdot (-e^{-\lambda x})]_0^{\infty} - \int_0^{\infty} (-e^{-\lambda x}) dx = [-\frac{e^{-\lambda x}}{\lambda}]_0^{\infty} = \frac{1}{\lambda}.$$

□

3　第 3 回　確率の概念と計算

3.1　前回のテスト感想から

{, } と [,] の区別をしっかり
年組番号名前は上に書きましょう
赤ペンでお願いします
復習しましょう
問題を写しましょう

3.2　確率の哲学

確率論は数学の中では最近始まった学問で，その起源は 1654 年のパスカルとフェルマーの間で交わされた手紙によると言われる．

なぜこんなに最近まで確率論は研究されなかったのか．大きく分けて 3 つあると思われる．1 つはアリストテレスの影響で，アリストテレスによるとすべてのものには原因があるのだから，偶然ということはあり得ないと言った．古代においてはこのような考え方は広く普及していた．2 つ目はキリスト教の影響で，神がすべてを決めるのだから，偶然ということはあり得ないと言った．3 つ目は，確率に関連する事柄は賭け事に関することで，下等に見られていた．賭け事が研究対象となり得なかったのは，神学の影響もあるので，これも広い意味で言えばキリスト教の影響と言えるかもしれない．

しかし確率という概念が存在しないのは非常に不便で，特に問題になったのは裁判においてであったらしい．数値で表せないのはともかく，「間違いなくそうだ」「もっともらしい」「おそらくそうだ」「そうかもしれない」「ありえない」というようなグレーの概念が使えないのは，裁判において非常に大変だったであろう．ちなみにこのような概念自体は，西洋以外，例えばインドなどでは，紀元前から頻繁に使われていたので，西洋におけるキリスト教の影響の大きさを感ぜずにおれない．

パスカルとフェルマーの間で交わされた手紙において問題になったのは，途中で終わったゲームの賞金はどう分配されるべきか，という問題．（他にもあったが，特にこの問題が重要である．）この手紙の中では「確率」という言葉は出てこないが，「ゲームをいつまでも続けたと仮定して均等な場合を数え上げる」という考え方が発見されたというのが転機であった．このやりとりをおそらくは聞いて，アントワーヌアルノーはポートロワイヤル論理学の中で，初めて確率（probability）という言葉を使った．「それは，10 人でゲームをして 1 人が勝つ確率は 1/10 である．なぜならばお互いに起こりうる可能性は等しいから．」

ここから確率の計算そして哲学について様々な議論を経て，1814 年ラプラスは『確率の哲学的試論』を表し，古典的確率を確立した．古典的確率では「同様に確からしい場合の数で，求める場合の数を割ったものが確率である」と定義する．この定義の有用性は計り知れない．しかし，一方で「同様に確からしい」とは一体どういうことなのか．ここにすでに確率の概念が出てきている．人によって異なる場合は？そもそも数えられない場合は？など様々な問題が起こってくる．もう少し詳しく知りたい人は，ベルトランの逆説で調べると面白いかもしれない．

決定的であったのは 1890 年代，ブラウン運動の研究の中で，確率概念が混乱をきたしていることを皆が感じ始めた．そこで，1900 年ヒルベルトが当時主流であった公理的な数学を使って，確率概念を公理化できないかと問題提起した．これに答えたのがコルモゴロフで，1933 年『確率論の基礎概念』においてである．コルモゴロフの確率論は公理的確率論と呼ばれ，確率の意味は問わず，確率が満たすべき性質だけを問題にする．そのため確率についてどんな立場を取っている人にとっても使いやすく，広く普及し，今でもただ確率論と言えばコルモゴロフの公理的確率論のことを指す．

確率とは何か，という哲学的問いにも，長い歴史があり，多くの問題がある．ここでは初歩として，頻度説，主観説，傾向説の 3 つの説について説明しよう．

1 つ目の頻度説とは，確率とは頻度であるという考え方．頻度とは繰り返し実験を行った時に，何回くらいそれが起こるか？ということ．6 面のサイコロにおいて 1 が出る確率は，サイコロを振る前には分からない．600 回振って

1 が 112 回出たとすれば，1 が出る確率は 112/600 回に近いらしいと思う．厳密な確率はその相対頻度の極限として定義する．極限なので実際に実験することはできず，それゆえ確率は求められない．しかし，確率が分かっていたとすれば，別の確率を求めることはできる．そういう考え方が頻度説である．確率の哲学の中でも最も素朴で，受け入れやすい反面，実際に確率を求めることができないので，数学的に使いにくいという欠点がある．

2 つ目は主観説で，確率とは信念の度合いであるという考え方．頻度説では「宇宙人のいる確率」というのは意味がない．実験を行うことができないから．しかし私たちはそのような言葉を使うこともある．1 度しか起こらないことに対しての確率とは，その人がそれについてどのくらい強く信じているか，というものが確率である．だから確率が人によって異なるということがありうる．私とあなたで宇宙人のいる確率は異なる（かもしれない）．しかし信念の度合いという以上は整合性が取れている必要がある．例えば，A という事象を信じている度合いが p であるとすれば，A でないという事象を信じている度合いは $1-p$ でなければならない．その意味で信念の度合いが満たすべき性質が，確率の満たすべき性質であるという考え方である．世にはベイズ主義とも呼ばれ，例えばメールのスパムフィルターなどに使われている．

3 つ目は傾向説で，確率とはその現象を引き起こす度合いであるという考え方．例えばサイコロで 1 の目が出る確率が 1/6 であるとは，サイコロ自身にそれぞれの目の出やすさという固有の数値を持っていて，サイコロが 1 の目を確率 1/6 で出すのだ，という考え方．これは事実としては間違っている．サイコロは完全に決定的に出る目が決まる．しかし，このような考え方をすると，非常に物事が単純化され，計算しやすくなる．なので間違っている，嘘だとは知っているが，このような考え方をしましょう，ということ．この考え方を「うまく」使うことが世の中を渡っていく上で必要なのだが，この確率が本物だと思うと様々な誤解や混乱の元になる．

では，次の事柄は頻度説，主観説，傾向説のうち，どれだと考えるのが最も自然だろうか．複数の立場がありうるということもあるかもしれないし，そもそも確率と考えるべきではないということもあるかもしれない．

- 降水確率
- 宝くじが当たる確率
- 打率
- 地震が起こる確率
- 原発事故の確率
- 合格確率
- 視聴率
- 留年率
- 私が留年する確率
- 就職率
- 生涯独身率
- 私が結婚できない確率
- ガンで死ぬ確率

3.3 確率の計算方法について

この講義で学ぶのは確率の哲学ではなく，数学としての確率論であり，それはコルモゴロフの公理的確率論である．よって，確率はどんな性質を持つのかを知り，その性質から確率や期待値などの確率に関連する値を求めることができるようになるのが目標である．

では確率とは何か，どんな性質を持つものか．確率はある全体集合の部分集合に対して定められる．全体集合を $X=\{1,2,3,4,5,6\}$ とすると，$A=\{2,4,6\}$ は部分集合であり，例えば，$P(A)=1/2$ と定める．このような部分集合を事象 (event) と呼ぶ．ここで「確率はどんな値でも良い」ことに注意しよう．必ずしもすべてが 1/6 でなくても良い．ただし，「どんな組み合わせでも良い」わけではない．取りうる事象 1 つ 1 つを根元事象と呼び，その集合を標本空間と呼ぶ．全体集合も 1 つの事象であり，全事象と呼ぶ．

$P(X)=1$, $P(\emptyset)=0$.

$P(A)+P(\overline{A})=1$.

和事象，$P(A\cup B)\geq P(A), P(B)$. 特に $A\cap B=\emptyset$ のときは，$P(A\cup B)=P(A)+P(B)$.

$A\cap B$ を積事象，$A\cap B=\emptyset$ のときは排反事象，$\overline{A}$ を余事象と呼ぶ.

例 3.1. $P(A)=1/3$, $P(B)=1/2$, $A\cap B=\emptyset$ のとき，$P(A\cup B)$, $P(\overline{A})$, $P(\overline{A}\cap B)$ を求めよう.

3.4 演習問題

問題 3.2. $P(A)=1/5$, $P(B)=1/4$, $P(A\cup B)=1/3$ のとき，$P(A\cap B)$, $P(A\cap\overline{B})$, $P(\overline{A}\cap B)$ を求めよ.

問題 3.3. 2 つのサイコロを投げたとき，異なる目が出る事象を A，少なくとも 1 つは 1 の目である事象を B とする.

(1) A が起こる確率を求めよ.

(2) B が起こる確率を求めよ.

(3) $A\cap B$ の確率を求めよ.

証明. (1) $\frac{5}{6}$

(2) $\frac{11}{36}$

(3) 10 通りなので，$\frac{5}{18}$

□

問題 3.4. 頻度説，主観説，傾向説の具体例を 1 つづつ挙げよ．面白いものであれば加点する.

4　第 4 回　場合の数と独立試行

4.1　前回の感想から

頻度説

- 麻雀の天和の確率
- 帰り道に赤信号で止まる確率
- ババ抜きで最初の手札にジョーカーがある確率
- ボーリングでストライクが出る確率
- 故障確率
- 金環日食が起こる確率
- ごま塩から 1 粒とってゴマである確率
- 1 日 3 食食べる確率

主観率

- 宿題を忘れた時に，先生に当てられる確率
- 占いの当たる確率
- 明日自分が死ぬ確率
- 神様が助けてくれる確率
- 地獄に行く確率
- 好きな人と付き合える確率

傾向説

- 明日台風が来る確率
- 不良品の確率
- 光の反射率
- 再婚した人の離婚率
- 洗剤の除菌率
- 15 時になるとお腹がすく確率
- ある人がじゃんけんでどの手を出すか
- 雨の日に車がスピンする確率
- 自販機に投入したお金が認識されずに戻ってくる確率（マシンに依存）
- クラスの男女が付き合う確率

「確率には約 200 年という長い歴史があることに驚いた.」これは驚き方がまちがっている．数学は数千年の歴史がある学問で，ほとんどの分野は昔にさかのぼることができる．しかし確率はわずか数百年前に発見された．なぜこんなに遅いのか，という議論がなされている．

「積事象だが，掛け算ではない！」気をつけましょう．

4.2　独立試行

確率の性質は

(1) 面積のような性質を持つ.
(2) 独立は乗法を表す.
(3) 条件付き確率.

の 3 つにまとめられる．今日はそのうちの 2 番目の独立について．独立にも独立試行と独立事象があるが，今日は独立試行について．独立事象についてはまた今度．

次のような問題を考えよう．「さいころを 2 つ投げて，2 つとも 1 の目が出る確率は？」答えはもちろん $\frac{1}{36}$ である．ではどのように計算してのことか．2 種類考えられる．

(1) $\frac{1}{6\times 6}$
(2) $\frac{1}{6}\times\frac{1}{6}$

それぞれどのように考えてのことか．(1) では，6×6 の表を考えて，それぞれが同じ確率だから，という考え方．いわゆる古典確率の考え方．これは分かりやすい．(2) では，さいころ 2 つに順番をつけて，1 番目のサイコロが 1 の目である確率が $\frac{1}{6}$．そのとき，2 番目のサイコロが 1 の目である確率が $\frac{1}{6}$ である．よって，その積が求める確率である．

この (2) の考え方にはずいぶんと多くの疑問がある．まず，さいころに順番をつけて良いのか．区別する場合と区別しない場合は非常にややこしい．同時に振っているのに，順番に振っているとして計算して良いのか．順番をつける順番が違ったら答えが違うということはないのか．何より，なぜ掛けるのか．足し算でも引き算でも割り算でもなさそうだが，掛け算で確率が求まるのはなぜか．

ここでは「2 つのサイコロは，お互いの目に依存しない」と仮定している．一方が 1 なら，もう一方は 6 が出やすいとか，1 が出やすいとか，そういうことはないと仮定している．どうしてかと言われてもそれはそういう約束だから．問題文に書いていないけど，それは暗黙の約束．本当にサイコロにそういう性質があるかは別問題．それは数学の問題ではない．数学ではそう仮定する．

このようなお互いに影響しない試行（実験）のことを独立試行という．独立試行の場合には確率は乗法によって求まる．「なぜ独立ならば乗法によって求まるのか」と聞いてはいけない．「独立の時には乗法で求まる」という性質を持つものを確率と呼んでいる．

教えられたことは理論であって正しいとは限らないし信じる必要もない．みなさんは理解するのが仕事であって，正しいと信じる必要はない．「納得できなければ信じない」と言って，理解しないのは視野を狭める．「教えられたことはすべて正しい」とやみくもに信じるのは，危うい．「自分は信じないけれども，そういう理論があることを理解する」というのが正しい態度．

では，具体例を見ていこう．誕生日が同じ人のいる確率は $1-\frac{{}_{365}P_n}{365^n}$．

順列 ${}_nP_k=\dfrac{n!}{(n-k)!}$．

- 区別できる n 個のものから k 個を選んで並べる場合の数
- 1 から 9 までの数字からできる同じ数字が出てこない 3 桁の数の場合の数は ${}_9P_3$
- トランプを 3 枚選んで並べる時の場合の数は ${}_{52}P_3$

組み合わせ ${}_nC_k=\dfrac{n!}{k!(n-k)!}$．

- n 個のものから k 個を選ぶ組み合わせの数
- $(ax+b)^n$ を展開した時の k 次の係数は ${}_nC_k a^k b^{n-k}$
- トランプを 5 枚ひいたときの組み合わせの数は ${}_{52}C_5$

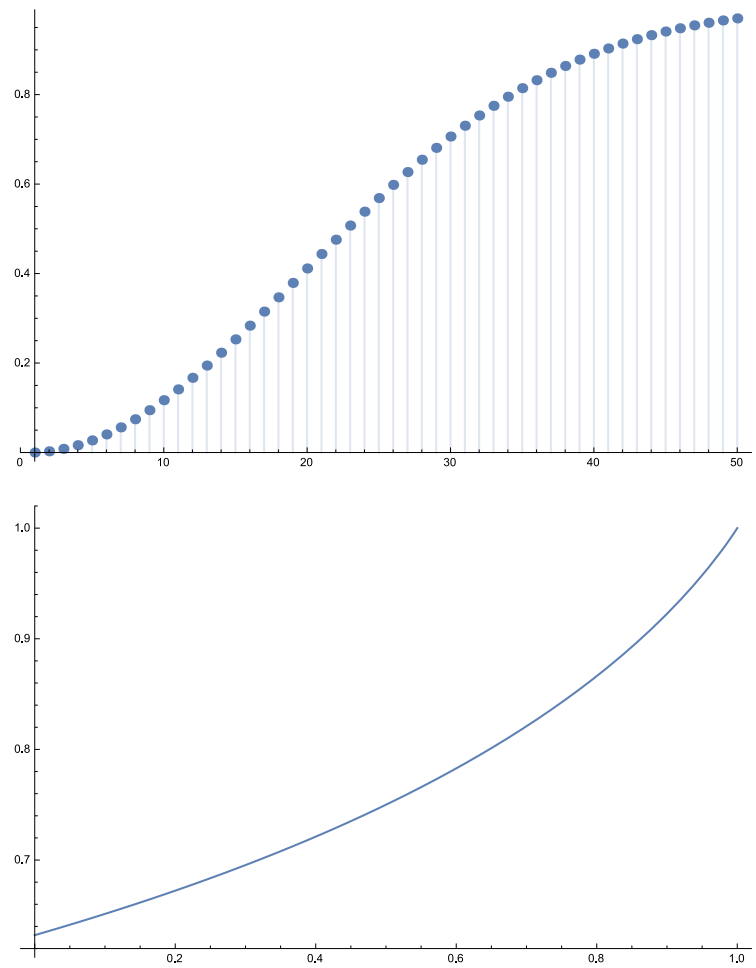
${}_nC_k={}_{n-1}C_k+{}_{n-1}C_{k-1}$

独立試行．試行が独立ならば確率は乗法で求まる．n 回の独立試行において，k 回成功する確率は，${}_nC_k p^k(1-p)^{n-k}$

確率 $\frac{1}{10}$ の事柄を 10 回行って，1 度は成功する確率は，$1-0.1^{10}\approx 0.65$．一般に $1-(1-x)^{1/x}$ のグラフは次のようになる．

ここで，$\lim_{x\to 0}(1-x)^{1/x}=\frac{1}{e}\approx 0.37$ である．これを 37% の法則といい，様々な場所で出てくる．例として「秘書問題」（もしくは「結婚問題」）を挙げよう．

(1) 秘書を 1 人雇いたい．
(2) n 人が応募してきている．n は既知とする．
(3) 応募者には重複なく順位が付けられる．



(4) 無作為に面接を行う.

(5) 毎回の面接の後，採用するかどうかをその場で決定する.

(6) その応募者を採用するかどうかはそれまでの相対的順位のみによって決定する.

(7) 不採用にしたら後から採用することはできない.

(8) 最も良い応募者を選択するにはどのようにすれば良いか.

戦略として考えられるのは，最初の r 人はスキップして，その後の面接者の中で，それまでの応募者の中で最もよければ採用する，という方法をとることになる．この時，最善の応募者を選択できる確率を最も高くするには，r をいくつにすればよいだろうか？またその時の確率はいくらだろうか.

最善の応募者が i 番目にいるとする．ただし，$r+1 \leq i \leq n$ とする．その確率は $\frac{1}{n}$ である．この応募者が選ばれるためには，$i-1$ 番目までの人の中で最高の順位の人が，r 番目までにいる必要がある．その確率は $\frac{r}{i-1}$ である．よって求める確率は,

$$p(r) = \sum_{i=r+1}^{n} \frac{1}{n} \cdot \frac{r}{i-1}$$

である．十分 n が大きい時，区分求積法より，$x = \frac{r}{n}$ として,

$$p(r) = \sum_{i=r+1}^{n} \frac{1}{n} \frac{x}{(i-1)/n} \to x \int_x^1 \frac{dt}{t} = -x \log x = f(x)$$

これが最大となるような x は,

$$f'(x) = -\log x - 1 = 0$$

を解いて，$x = \frac{1}{e}$．このとき最善の応募者を選択できる確率は,

$$f(e^{-1}) = \frac{1}{e}.$$

4.3 演習問題

問題 4.1. 銃で標的に命中する確率が $\frac{3}{5}$ であるとする.

(1) 4 回打った時に，3 回以上命中する確率を求めよ．

(2) 少なくとも 1 回標的に命中する確率を 0.99 よりも大きくするためには，何回打たなければならないか．

証明.　(1) ${}_4C_3\left(\frac{3}{5}\right)^3\left(\frac{2}{5}\right)^1+{}_4C_4\left(\frac{3}{5}\right)^4=\frac{189}{625}$

(2) $\left(\frac{2}{5}\right)^n\leq\frac{1}{100}$ を解けば良い．5 ではだめで，6 なら良いので，答えは 6 回．

□

問題 4.2. $P(A)=p$ とする．n 回の重複試行において，A が k 回起こる確率を $p(k)$ とする．

(1) $\dfrac{p(k)}{p(k-1)}$ を求めよ．

(2) $p(k)$ が最大となるような k は n,p を使ってどのように表されるか．

証明. $p(k)={}_nC_kp^k(1-p)^{n-k}$ であるから，計算して，$\dfrac{n-k+1}{k}\cdot\dfrac{p}{1-p}$. この値 >1 を解くと $k<(n+1)p$ となるので，答えは $[(n+1)p]$. □

問題 4.3. 答えが ${}_nP_k$, ${}_nC_k$ となるような数にはどんなものがあるか．面白いものであれば加点する．

5　第 5 回　条件付き確率

${}_nP_k$-アイドルがステージに立つ時の並び方，色鉛筆を並べる．${}_nC_k$-野菜ジュースに入れる野菜の選び方，n 個のうち k 個の窓を開ける．

「問題を解く時間が短い」10 点満点は 5 人くらい，8 点以上は 15 人くらい．

確率の持つ性質の中で条件付き確率は，数学的には難しくないのに，直感に反することが度々あるため，混乱しやすい内容である．

5.1　条件付き確率とベイズの定理

あるツボに赤玉 3 個，青玉 2 個入っている．ここから 1 個の玉を取り出す (戻さない) という操作を 2 回行う．1 回目の玉が赤である確率を A，2 回目の玉が赤である確率を B とする．

$P(A) = \frac{3}{5}$ である．では $P(B)$ はどう計算したら良いか．1 回目が終わって 2 回目の操作を行う時のツボの中の状態は 1 回目の操作に依存する．もし 1 回目が赤ならば，赤玉 2 個青玉 2 個なので確率 $\frac{1}{2}$ である．もし 1 回目が青ならば，赤玉 3 個青玉 1 個なので確率 $\frac{3}{4}$ である．そこで，これを条件付き確率といい，$P(B|A) = \frac{1}{2}$, $P(B|\overline{A}) = \frac{3}{4}$ と表す．

1 回目が赤で 2 回目も赤という事象は $A \cap B$ と表される．この確率 $P(A \cap B)$ は 1 回目が赤の確率 $P(A)$ に $P(B|A)$ を掛ければ良い．この条件のもとでは独立のように振舞う．そこで，$P(A \cap B) = P(A)P(B|A)$ という性質が成り立つ．これは図に書くと理解しやすい．

$P(B) = P(A \cap B) + P(\overline{A} \cap B)$ であったから，$P(B) = \frac{3}{5} \times \frac{1}{2} + \frac{2}{5} \times \frac{3}{4} = \frac{6+6}{20} = \frac{3}{5}$ である．すなわち $P(A) = P(B)$ であり，これはくじ引きにおいて当たる確率は操作の順番に依らないことを表している．

「2 回目が赤玉であった時の 1 回目が赤玉である確率」$P(A|B)$ を求めよう．先ほどのように考えることはできないので，$P(A|B) = \frac{P(A \cap B)}{P(B)}$ という性質を使おう．$P(A|B) = \frac{3/10}{3/5} = \frac{1}{2}$ である．$P(A) = \frac{3}{5} > \frac{1}{2} = P(A|B)$ なので，「2 回目が赤」という条件があったほうが，「1 回目が赤」の確率は低くなる．なぜならば 1 回目が青のほうが，2 回目が赤である可能性が高いからだ．

このような考え方は様々な場面で使われるが，最も身近な例はメールのスパムフィルターだろう．予め学習セットと呼ばれるメールを与えて，そこにはスパムかスパムでないかが判定されているとしよう．スパムに含まれる単語とスパムでないメールに含まれる単語を調べる．次に新しいメールが来た時にどの単語が含まれているかを見て，スパムである可能性を計算する．90% 以上ならばスパムと判定する．このような判定基準を閾値と言ったりする．間違えてスパム判定した場合には修正し，学習するメールの量が増えれば，精度は高くなっていく．人によって異なる部分もあれば，ほどんどの人がスパムと判定するメールもある．この辺りをどのようにするかなど，凝り出すとキリがない．

以下では条件付き確率にまつわる有名なパラドックスを 2 つ紹介する．

5.2　診断

ある病原菌の検査試薬は，病原菌がいるのに誤って陰性と判断する確率は 0.1%，病原菌がいないのに誤って陽性とする確率が 1% である．全体の 0.01% にこの病原菌が感染している集団から 1 つの個体を取り出す．この検査が陽性であったときに，実際に病原菌に感染している確率を求めよう．

取り出した個体が感染しているという事象を A，検査結果が陽性であるという事象を B とする．求めるのは

$P(A|B)$ である．条件より $P(A) = 10^{-4}$, $P(\overline{B}|A) = 10^{-3}$, $P(B|\overline{A}) = 10^{-2}$．よって，

$$\begin{aligned}
P(A \cap B) =& P(A)P(B|A) = 10^{-4} \times (1 - 10^{-3}),\\
P(\overline{A} \cap B) =& P(\overline{A})P(B|\overline{A}) = (1 - 10^{-4}) \times 10^{-2},\\
P(B) =& P(A \cap B) + P(\overline{A} \cap B) = 10^{-2} + 10^{-4} - 10^{-6} - 10^{-7},\\
P(A|B) =& \frac{10^{-4} - 10^{-7}}{10^{-2} + 10^{-4} - 10^{-6} - 10^{-7}} \approx 10^{-2}.
\end{aligned}$$

よって，約 1%.

5.3 モンティホール問題

3 つの扉のうち 1 つにだけ賞品が入っていて，回答者はそれを当てたら賞品がもらえる．ただし扉は次のように 2 段階で選ぶことができる．

(1) まず回答者は 3 つの扉からどれか 1 つを選ぶ.
(2) 次に答えを知っている司会者が，選んでいない扉の中から賞品の入っていない扉を 1 つ開ける．回答者が当たりの扉を選んでいる場合は，残りの扉からランダムに 1 つ開けるとする．この後，回答者は扉を選び直しても良い.

扉を換えるのと換えないのでは，どちらが当たる確率が高いか.

深く考えないと確率は 1/2 ずつで，確率が同じなら最初に選んだほうを選び続けるほうが良いと多くの人は考える．変えて外れると悔いが残るので，それを嫌うのであろう.

回答者が最初に選んだ扉を A，司会者が開けた扉を B，もう一つの扉を C とし，それぞれの扉に賞品がある事象も表すとする．賞品は 3 つの扉に等確率で入っているしているので，$P(A) = \frac{1}{3}$．B には賞品は入っていないので，$P(B) = 0$. C に賞品が入っているのは，A に賞品が入っていなかった場合なので，$P(C) = \frac{2}{3}$.

確率が上がるのはなぜだろうか？それは「司会者が選ばない」という情報が増えるからである．扉が 100 個の場合を考えると状況がより明確になるだろう.

問題 5.1. 子供が男か女かは確率 $\frac{1}{2}$ ずつであるとしよう.

(1) 2 人の子供のうち少なくとも 1 人は男の子であることが分かっているとしよう．この家庭に女の子の子供がいる確率はいくらか.
(2) 2 人子供のいる家を訪ねたら 1 人の男の子が顔を出した．もう一人の子供が女の子である確率はいくらか.

問題 5.2. あるネジ工場にはネジをつくる機械 3 台，A, B, C があり，それぞれ全体の $50\%, 30\%, 20\%$ を生産している．A, B, C の各機械でつくるネジのうち $2\%, 3\%, 4\%$ が不良品である．今，製品全体の中から 1 個のネジを取り出すと，それは不良品であった．それが B で生産されたものである確率を求めよ.

証明. 取り出したネジが A, B, C の各機械でつくられたという事象をそれぞれ A, B, C，取り出したネジが不良品であるという事象を D とすると，

$$P(A) = 0.5,\ P(B) = 0.3,\ P(C) = 0.2,\ P(D|A) = 0.02,\ P(D|B) = 0.03,\ P(D|C) = 0.04$$

である．よって，

$$P(D) = 0.5 \times 0.02 + 0.3 \times 0.03 + 0.2 \times 0.04 = 0.027,$$

$$P(B|D) = \frac{P(B)P(D|B)}{P(D)} = \frac{0.3 \times 0.03}{0.027} = \frac{9}{27} = \frac{1}{3}$$

□

6　第 6 回　確率変数と期待値

6.1　確率変数の概念について

これまではサイコロやトランプ，ツボなどできるだけ具体的な対象の確率を考えてきた．しかしこれからは確率変数という抽象的な概念が出てくる．そしてこれからずっとこの確率変数という言葉を使いつづける．昨年 1 年間教えてみて分かったことは，この確率変数の概念を理解し損ねた人が案外多いこと．ここでつまづくとこの後全部分からなくなる．できる限りゆっくり話をするので，きちんと理解して欲しい．

確率変数とは確率的に変化する値である．定義はこれだけ．この意味をしっかり理解しよう．

例えば，サイコロを振る．サイコロの目は 1 から 6 までのどれかが出る．サイコロの目を X とすると，$X = 1$ となる確率は $\frac{1}{6}$．これを $P(X = 1) = \frac{1}{6}$ とも書く．この X を確率変数と言い，英語では random variable という．

X と書いてあるが，これは何か 1 つの値ではない．$X = 1$ の時もあるし，$X = 2$ の時もある．変化する．なので変数である．これはちょうど関数の媒介変数に似ている．$f(x) = x^2$ と書いた時には，x は実数を動く．0 のときもあるし，1 の時もある，変化する．問題は f がどのように変化するかであって，x の値は重要ではない．$X = 1$ と書くのは，関数で言えば $f(x) = 1$ と書くようなもので，そうなるような x はいくつか存在するのと同様に，そうなるような世界がいくつか存在して，そうなる世界の確率を考えることになる．

サイコロの場合，$i = 1, 2, 3, 4, 5, 6$ に対し $P(X = i) = \frac{1}{6}$，という式ですべての情報を表す．数学として興味があるのはサイコロではなく，確率なので，サイコロという情報を忘れよう．なので，「$i = 1, 2, 3, 4, 5, 6$ に対し $P(X = i) = \frac{1}{6}$ となる確率変数 X を考える」という言い方をする．最初のうちはこの確率変数という考え方が慣れないかもしれない．その場合は「ーとなるようなサイコロを考える」と置き換えると良いかもしれない．今後は取る値が実数だったり，確率が平等でなかったりするので，「ーとなるような変なサイコロを考える」と思う．

重要なのは確率の値だけなので，それを表にしたものを確率分布表という．例えば，2 個のサイコロの目の和を Y とすると，確率分布表は $P(Y = 2) = \frac{1}{36}, P(Y = 3) = ...$ となる．この時，Y を 2 個のサイコロの目の和と考えるのではなく，新しい確率変数として，このような確率分布を取る新しい変な 1 個のサイコロの目と考えよう．

今後は「何かの確率」を求めることが重要なのではなく，「確率変数の振る舞い」つまり「確率がどのように分布しているか」に注目する．

6.2　期待値

次のようなゲームを考えよう．2 つのサイコロを投げてそのサイコロの目の差を X とする．賞金 $Y = X \times 100$ 円を受け取る．この確率分布表は次のようになる．

X	0	1	2	3	4	5
$P(X)$	3/18	5/18	4/18	3/18	2/18	1/18

さてこのゲームの参加費としてはいくらが適切だろうか？ちなみに日本の法律では賭博は禁じられている．賭博の成立要件には微妙な部分があるので，不安に思ったら詳しい人にきちんと相談してからの方が良い．0 円では商売あがったりだし，500 円では誰も挑戦しないだろう．その間のはずだ．今，利益とか人件費などを無視して，長い間繰り返し行った時に平均的に客と商売人が同等となるような金額を設定しよう．これを期待値といい，$E(X)$ で表す．これは，それぞれの値に確率を掛けたものの和で計算できる．

$$E(X) = \sum_{i=0}^{5} i \times 100 \times P(X = i) = 100 \times \frac{5}{18} + 200 \times \frac{4}{18} + 300 \times \frac{3}{18} + 400 \times \frac{2}{18} + 500 \times \frac{1}{18} = \frac{35}{18} \cdot 100 \approx 194$$

2 つのサイコロを投げて，偶数だったら賞金 100 円，3 の倍数だったら賞金 200 円がもらえるゲームを考えよう．今，サイコロの目を X とする．これは確率変数である．賞金の値段は確率変数でこれを Y としよう．これも確率変

数だが，X によって決まる．つまり，

$$Y = \begin{cases} 0 & \text{if } X = 1, 5 \\ 100 & \text{if } X = 2, 4 \\ 200 & \text{if } X = 3 \\ 300 & \text{if } X = 6 \end{cases}$$

これより Y の確率分布表を書くと，

Y	0	100	200	300
$P(Y)$	1/3	1/3	1/6	1/6

となる．よって，Y の期待値 $E(Y)$ は，

$$E(Y) = 100 \times \frac{1}{3} + 200 \times \frac{1}{6} + 300 \times \frac{1}{6} = \frac{100 + 100 + 150}{3} = \frac{350}{3}.$$

この期待値は次のようにも求められる．Y_1 を X が偶数の時には 100，奇数の時には 0 を取る確率変数とし，Y_2 は X が 3 の倍数の時には 200，そうでない時には 0 を取る確率変数とする．明らかに $Y = Y_1 + Y_2$ である．$E(Y_1) = 100 \times \frac{1}{2} = 50$ であり，$E(Y_2) = 200 \times \frac{1}{3} = \frac{200}{3}$ なので，

$$E(Y_1) + E(Y_2) = 50 + \frac{200}{3} = \frac{350}{3} = E(Y_1 + Y_2)$$

となっている．これは偶然ではない．和の期待値は期待値の和になる．

サイコロを 2 つ振った時の目の和を Z とする．$E(Z)$ を求めたい．サイコロ 2 つの目をそれぞれ Z_1, Z_2 とすると，$Z = Z_1 + Z_2$ である．その期待値は $E(Z_1) = E(Z_2) = \frac{7}{2}$ である．よって，$E(Z) = E(Z_1) + E(Z_2) = 7$ である．

積の期待値は期待値の積には必ずしもならない．積については来週．

問題 6.1. ある宝くじは 1 枚 300 円を 2 億枚販売している．その当選金額は以下のとおりである．

等級	当選金額	当選本数
1 等	5 億円	20 枚
2 等	500 万円	2000 枚
3 等	300 円	2 千万枚

当選金額の期待値を求めよ．

証明.

$$5 \cdot 10^8 \times \frac{20}{2 \cdot 10^8} + 5 \cdot 10^6 \times \frac{2000}{2 \cdot 10^8} + 300 \times \frac{2 \cdot 10^7}{2 \cdot 10^8} = 130$$

□

問題 6.2. 100 種類のメダルが等確率で出るゲームを考える．今，50 種類のメダルを持っているとしよう．

(1) 次にゲームを行った時に新しい種類のメダルが出る確率を求めよ．
(2) k 回目に初めて新しい種類のメダルが出る確率を求めよ．
(3) 新しい種類のメダルが出るまでに行うゲームの回数を X としたとき，$E(X)$ を求めよ．

証明. (1) $\frac{1}{2}$
(2) 2^{-k}
(3)

$$E(X) = \sum_{k=1}^{\infty} k 2^{-k} = 1 \cdot 2^{-1} + 2 \cdot 2^{-2} + \cdots$$

この値を S とすると，

$$2^{-1} S = 1 \cdot 2^{-2} + 2 \cdot 2^{-3} + \cdots$$

より，

$$2^{-1} S = 1 \cdot 2^{-1} + 1 \cdot 2^{-2} + \cdots = 1$$

よって，$S = 2$.

□

問題 6.3. コインを表が出るまで投げ続け，表が出た時に賞金がもらえるゲームを考える．もらえる賞金は，1 回目ならば 1 円，2 回目ならば 2 円，3 回目ならば 4 円，k 回目ならば 2^{k-1} 円である．このゲームの参加費として期待値を設定したい．いくらにすべきだろうか？

7　第 7 回　期待値，分散，標準偏差

7.1　確率と統計

期待値は平均と深い関係がある．その関係を理解するためには確率と統計の違いについてハッキリ理解する必要がある．

(1) サイコロを振ると，確率 $\frac{1}{6}$ で 1 から 6 の目が出る．この時，どのくらいの値が出ると期待されるかというのが期待値で計算すると $\frac{7}{2}$ となる．
(2) サイコロを 10 回振ったら，$1, 5, 4, 3, 3, 1, 1, 2, 3, 2$ であった．その平均はすべてを足して回数で割って $\frac{25}{10} = 2.5$ である．

期待値はサイコロを振る前の話，平均はサイコロを振った後の話．平均は期待値に近いはず．コンピュータでシミュレーションしてみると，平均は 10 回で 2.7, 100 回で 3.15，1000 回で 3.504，10000 回で 3.5154，10 万回で 3.49481，100 万回で 3.500391 となった．これがどのくらいの速さで近づくのかについては，もう少し後で勉強する．

さて今確率分布が分かっている時には期待値や平均値という話をした．しかし，確率分布が分かっていなくても，データさえあれば平均値については語ることができる．確率的に振舞っていない現象であってでも，データさえあれば平均値について語ることができる．例えばテストの点数の平均点は，すべての生徒の点数を合計し生徒数で割ったものである．しかし，確率的に振舞っているわけではない．事実としては確率的に振舞っているわけではないが，確率的に振舞っていると見なして計算することもある．ここが確率論の分かりにくいところである．

さて，テストでよく出てくるのが偏差値である．今日は偏差値が何を意味しているのかを理解しよう．

今同じクラスで数学と英語のテストがあったとする．どちらのテストも平均点は 60 点だった．A さんは数学は 90 点，英語は 80 点であった．どちらが「すごい」だろうか．どちらが「取りにくい点数」だろうか．

点数だけでは自分がクラスの中でどのくらいの位置にいるか分からない．順位も 1 つの指標ではあるが，あまり良い指標ではない．偏差値はそれよりは「ましな」指標である．

数学の方が点数が高いのだから，数学の方が「すごい」ではダメなのだろうか？そんな単純にはいかない．例えば数学の方は 100 点も 0 点も多く，100 点から 0 点まで万遍なくいるとしよう．それに対して英語はほとんどの人が 60 点で，70 点以上の人はほとんどいないとしよう．この場合，点数は数学の方が高くても，英語の 80 点の方が取りにくい．その点数にどのくらい価値があるかは，平均点だけではなく，点数の散らばり具合によって変わる．

この点数の散らばり具合を表すのが分散 (variance) である．これからデータに対する分散の式を書こう．

$$V = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2$$

ここで，x_i はそれぞれの点数で，m は平均である．この分散 V は平均からの距離の 2 乗の和であり，データの散らばり具合を表す．分散が大きいということは，データの散らばり具合が大きいということである．この式を展開すると，

$$V = \frac{1}{n}\sum_{i=1}^{n}(x_i^2 - 2mx_i + m^2) = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2m\frac{1}{n}\sum_{i=1}^{n}x_i + \frac{1}{n}m^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - 2m\frac{1}{n}m + \frac{1}{n}m^2 = \frac{1}{n}\sum_{i=1}^{n}x_i^2 - m^2$$

となって，分散は (2 乗の平均)-(平均の 2 乗) となっていることも分かる．

分散は 2 乗されているので，これの平方根の値も重要な値で，標準偏差 (standard deviation) と言われている．σ や s で表される．

$$\sigma = \sqrt{V}$$

各個人の偏差値 (standard score) は，

$$T_i = 50 + \frac{10(x_i - m)}{\sigma}$$

として定義される．平均点と同じ点数であったときには，偏差値が 50 となる．偏差値が 100 を越えたり，0 を下回ったりすることも理論上はあり得る．

テストの点数など多くの場合，データは正規分布をする．正規分布の場合，偏差値が 60 以上は 15% ほど，70 以上は 2.27%，80 以上は 0.13%，90 以上は 0.003%，100 以上は 0.00002%，くらいとなる．

7.2 分散，標準偏差

今まではデータに対する分散の話をしてきた．今度は確率変数に対する分散の話をしよう．確率変数の分散は，確率分布が $P(X = x_k) = p_k$ であるとき，

$$V(X) = \sum_{k=1}^{s} (x_k - E(X))^2 p_k$$

で定義される．

サイコロを振った時の分散は，

$$V(X) = (1 - 7/2)^2 \times \frac{1}{6} + (2 - 7/2)^2 \times \frac{1}{6} + \cdots + (6 - 7/2)^2 \times \frac{1}{6} = \frac{35}{12}$$

である．

確率変数の場合にも，

$$V(X) = E(X^2) - (E(X))^2$$

が成り立つ．

$$E(X^2) = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + \cdots + 6^2 \times \frac{1}{6} = \frac{91}{6}$$

よって，

$$V(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{182 - 147}{12} = \frac{35}{12}$$

7.3 確率変数の演算

サイコロを 2 回振った場合，それぞれの目を X, Y として，

$$E(XY)$$

を求めてみよう．

$$E(XY) = \sum_{i=1}^{6} \sum_{j=1}^{6} i \times j \times \frac{1}{36} = \sum_{i=1}^{6} \frac{i}{6} \sum_{j=1}^{6} \frac{j}{6} = E(X)E(Y) = \frac{49}{4}$$

実は一般に X, Y が独立の時，

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

が成立する．また，X, Y が独立の時，

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

が成立する．例えば，サイコロを 2 回振った時の和を Z とすると，

$$V(Z) = 2 \times \frac{35}{12} = \frac{35}{6}$$

と求まる．

複雑なものの期待値や分散を求める時には，独立なものをうまく利用するのが良い．

また線形変換に関しては，

$$E(aX + b) = aE(X) + b, \quad V(aX + b) = a^2 V(X)$$

などが知られている．

問題 7.1. 確率変数 X の確率分布が，$P(X = -1) = \frac{1}{3}$，$P(X = 0) = \frac{1}{6}$，$P(X = 2) = \frac{1}{2}$ で与えられているとする．X の期待値，分散，標準偏差を求めよ．

証明. $E(X) = -\frac{1}{3} + 2 \cdot \frac{1}{2} = \frac{2}{3}$. $E(X^2) = \frac{1}{3} + 4 \cdot \frac{1}{2} = \frac{7}{3}$. $V(X) = \frac{7}{3} - \left(\frac{2}{3}\right)^2 = \frac{17}{9}$. $\sigma(X) = \frac{\sqrt{17}}{3}$. □

問題 7.2. (1) あるテストで平均点は 55.0 点，A さんの点数は 70 点で偏差値が 55.0 であった．B さんの偏差値が 65.0 であったとき，B さんの点数をいくらか.

(2) テストである人の偏差値が 100 を越えるためには最低何人が受験しなければならないか.

証明. (1) 標準偏差を σ とすると,

$$50 + \frac{10 \times (70 - 55)}{\sigma} = 55.0$$

より，$\sigma = 30$. B さんの点数を x とすれば,

$$50 + \frac{10 \times (x - 55)}{30} = 65$$

より，$x = 100$.

(2) n 人が受験して, $n-1$ 人が 0 点, 1 人が 100 点であったとしよう．平均点は $100/n$, 分散は $100^2/n - (100/n)^2 = 100^2 \frac{n-1}{n^2}$. よって，100 点をとった人の偏差値は,

$$50 + \frac{10 \times (100 - \frac{100}{n})}{100 \times \frac{\sqrt{n-1}}{n}} = 50 + 10\sqrt{n-1}$$

これが 100 を越えるためには $n \geq 26$ でなければならない.

□

問題 7.3. あるテストで A 君は 60 点で，平均点が 53.0000 点，偏差値が 58.7500 であった．その後ある 1 人の別室受験者の点数が反映されていないことが分かり，再計算したところ，平均点が 53.9545 点，A 君の偏差値は 55.9940 となった．この別室受験者の点数を求めよ．(このタイプの計算は誤差が蓄積されやすい．有効数字を十分とって計算せよ．)

証明. 別室受験者を除いた受験者数を n 人とし，n 人の点数の和を A，n 人の点数の 2 乗の和を B，別室受験者の点数を x とする.

平均点を式で表すと,

$$\frac{A}{n} = 53.0000, \quad \frac{A + x}{n + 1} = 53.9545$$

別室受験者を入れずに計算した場合の標準偏差を σ_1 とすると,

$$50 + \frac{10 \times (60 - 53.0000)}{\sigma_1} = 58.7500$$

より，$\sigma_1 = 8.00$. よって，点数の 2 乗の平均は

$$\frac{B}{n} = 53.00^2 + 8.00^2 = 2873$$

別室受験者を入れて計算した場合の標準偏差を σ_2 とすると,

$$50 + \frac{10 \times (60 - 53.9545)}{\sigma_2} = 55.9940$$

より，$\sigma_2 = 10.0859$. よって，点数の 2 乗の平均は

$$\frac{B + x^2}{n + 1} = 53.9545^2 + 10.0859^2 = 3012.81$$

この 4 つの式から A, B, n, x を求める．求めたいのは x なので A, B, n を順に消去しよう．まず，A, B を消去して,

$$\frac{53.00n + x}{n + 1} = 53.9545, \quad \frac{2873n + x^2}{n + 1} = 3012.81$$

第 1 式から $n = \frac{x - 53.9545}{0.9545}$．第 2 式から $n = \frac{x^2 - 3012.81}{139.81}$ に代入して,

$$0.9545x^2 - 139.81x + 4667.6515 = 0$$

$D = 1725.742673$ より，$\sqrt{D} = 41.542059$ なので，

$$x = 94.998459,\ \ 51.476135$$

平均点が上昇していることから，$x = 95$ と推測できる．また，$n = 43$ も分かる． □

8　第 8 回　幾何分布，ポアソン分布

今日は次のような問題を考える．

1 回 100 円でメダルが出るゲームを考える．メダルは 100 種類あり，すべて等確率で出る．このメダルを 50 種類，80 種類，100 種類集めるまでに必要な金額の期待値を求めよう．

この問題は次のように考える．今，$k-1$ 種類のメダルを持っているとして，k 種類目のメダルが出るまでのゲームの回数を X_k とすると，求める期待値は，$n = 50, 80, 100$ として，

$$E(100\sum_{k=1}^{n} X_k) = 100\sum_{k=1}^{n} E(X_k)$$

となる．よって，$E(X_k)$ を求めよう．

$k-1$ 種類のメダルを持っているときに新しいメダルが出る確率は

$$1 - \frac{k-1}{100}$$

である．繰り返し行ったときにこの確率のことが初めて起こるまでの回数の期待値を求めたい．

このタイプの確率分布を幾何分布という．確率 p で起こる事柄が何回目で起こるか？という問題．合格確率 10% で大学を順番に受けていったら，何校目で初めて合格するか．ナンパの成功確率 3% だったとして，何人目で成功するか．などなど．

1 回目で起こる確率は

$$P(X = 1) = p$$

であり，2 回目で起こる確率は,

$$P(X = 2) = (1-p)p$$

同様にして k 回目で初めて起こる確率は,

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

である．よって，期待値は,

$$E(X) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$$

を計算すれば良い．

このタイプの計算は高校でもやったように公比をかけてずらして引くという方法で求まる．しかし，ここでは大学で学ぶ方法で求めてみよう．$|x| < 1$ に対しては,

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$$

が成立する．これを両辺 x で微分する．右側は項別微分と呼ばれる．どういうときにこれができるのかは微分積分学で学んで欲しい．

$$\frac{1}{(1-x)^2} = \sum_{k=1}^{\infty} kx^{k-1}$$

これと先ほどの式を見比べて $x = 1-p$ を代入し，p をかけてやると,

$$E(X) = \frac{1}{p}$$

と求まる．

さらに先ほどの式に x をかけてもう一度微分してやると,

$$\frac{(1-x)^2 + 2x(1-x)}{(1-x)^4} = \sum_{k=1}^{\infty} k^2 x^{k-1}$$

より，

$$\frac{1+x}{(1-x)^3} = \sum_{k=1}^{\infty} k^2 x^{k-1}$$

ここで $x = 1-p$ を代入し，p をかけてやると，

$$\frac{2-p}{p^2} = \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p = E(X^2)$$

よって，

$$V(X) = E(X^2) - (E(X))^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

結論を見ると，確率 10% の事柄が最初に起こるまでにかかる期待値は 10 回となって，直感にもよく適合するだろう．しかし，一般に「10 回くらいだから最初の方はまだ出ないだろう」とか「10 回に近づいてきたからそろそろ出やすくなったかな」などと考えることがある．しかし，毎回確率 p で起こるのだからそういうことはありえない．これを幾何分布の「無記憶性」という．

最初から s 回以内で起こる確率

$$P(X \leq s)$$

と，t 回起こらなかった後で s 回以内で起こる確率

$$P(X \leq s+t \mid X > t)$$

を比べてみよう．

$$P(X > t) = \sum_{k=t+1}^{\infty} (1-p)^{k-1} p = (1-p)^t p \frac{1}{1-(1-p)} = (1-p)^t$$

よって，

$$P(X = s+t \mid X > t) = \frac{P(X = s+t)}{P(X > t)} = \frac{(1-p)^{s+t-1} p}{(1-p)^t} = (1-p)^{s-1} p = P(X = s)$$

すなわち，

$$P(X \leq s) = P(X \leq s+t \mid X > t).$$

最初の問題に戻ろう．今，少し一般化して $N = 100$, $c \in [0,1]$ とすると，

$$E(X_k) = \frac{1}{1-(k-1)/N} = \frac{N}{N-(k-1)}$$

であるから，

$$\sum_{k=1}^{cN} E(X_k) = \frac{N}{N} + \frac{N}{N-1} + \cdots + \frac{N}{N-(cN-1)} = N(\sum_{k=1}^{N} \frac{1}{k} - \sum_{k=1}^{(1-c)N} \frac{1}{k})$$

ここで，N が十分大きいとして，以下のように近似式が知られている．

$$\sum_{k=1}^{N} \frac{1}{k} = \ln N + \gamma$$

ここで γ はオイラーの定数と呼ばれる数で約 0.57 である．これより，$c \neq 1$ のときは，

$$\sum_{k=1}^{cN} E(X_k) = N(\ln N - \ln(1-c)N) = -N \ln(1-c)$$

で，$c = 1$ のときは，

$$\sum_{K=1}^{N} E(X_k) = N(\ln N + \gamma)$$

となる．例えば，$N = 100$ で $c = 0.5, 0.8, 0.9, 1$ の時にはそれぞれ，

$$69.3, \quad 160, \quad 230, \quad 517$$

となる．

問題 8.1. ポアソン分布は，$\lambda > 0$ として，

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

で表される確率分布であり，「単位時間当たり λ 回発生する事象が単位時間に k 回起こる確率」を表す．例として，

- ある交差点を 1 時間当たりに通過する台数
- 1 日に受け取る電子メールの数
- 1 日に店に来る客の人数
- 1 時間当たりの Wikipedia の更新数

などがある．

(1) ポアソン分布が確率分布であることを示せ．
(2) $E(X), V(X)$ を求めよ．
(3) X, Y が独立でそれぞれ λ, μ をパラメータとするポアソン分布に従うとき，$X + Y$ は $\lambda + \mu$ をパラメータとするポアソン分布に従うことを示せ．

証明. (1)

$$\sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = 1$$

(2)

$$E(X) = \sum_{k=0}^{\infty} kP(X = k) = \sum_{k=0}^{\infty} k\frac{\lambda^k e^{-\lambda}}{k!} = \lambda$$

$$E(X^2) = \sum_{k=0}^{\infty} k^2 P(X = k) = \sum_{k=1}^{\infty} k\frac{\lambda^k e^{-\lambda}}{(k-1)!} = \sum_{k=2}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-2)!} + \sum_{k=1}^{\infty} \frac{\lambda^k e^{-\lambda}}{(k-1)!} = \lambda^2 + \lambda$$

よって，

$$V(X) = E(X^2) - (E(X))^2 = \lambda$$

(3)

$$\begin{aligned} P(X + Y = k) &= \sum_{n=0}^{k} P(X = n)P(Y = k - n) \\ &= \sum_{n=0}^{k} \frac{\lambda^n e^{-\lambda}}{n!} \frac{\mu^{k-n} e^{-\mu}}{(k-n)!} \\ &= \frac{1}{k!} e^{-(\lambda+\mu)} \sum_{n=0}^{k} {}_kC_n \lambda^n \mu^{k-n} \\ &= \frac{(\lambda + \mu)^k}{k!} e^{-(\lambda+\mu)} \end{aligned}$$

□

問題 8.2. 二項分布 $B(n, p)$ において，$np = \lambda$ を保ったまま $n \to \infty$ とすると，ポアソン分布に収束することを示せ．

証明. $B(n, p)$ においては，

$$P(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

である．これとポアソン分布の確率 $\frac{\lambda^k e^{-\lambda}}{k!}$ と比較して，次のように変形する．

$$P(X = k) = \frac{(np)^k}{k!} \frac{n \cdot (n-1) \cdots (n-k+1)}{n \cdot n \cdots n} (1-p)^{\lambda/p - k}$$

ここで，k を固定しているので，$n \to \infty$ のとき $\frac{n-k+1}{n} \to 1$．また，$n \to \infty$ のとき $p \to 0$ で，$(1-p)^{\lambda/p-k} \to e^{-\lambda}$．
これより，$P(X = k) \to \frac{\lambda^k e^{-\lambda}}{k!}$．
□

9　第 9 回　チェビシェフの不等式，大数の弱法則

9.1　チェビシェフの不等式

チェビシェフの不等式は「平均から離れた値はあまり多くない」ということを表している．つまり，多くは平均に近いところにあるよ！ということ．このこと自身も重要な事実だが，今日の後半ではこの事実を使って大数の法則というとても重要な事実を証明する．

確率と統計は区別しなければならない．チェビシェフの不等式は，確率においても，統計においても成り立つが，おそらくはわかりやすいと思うので，まず統計の方から説明する．

100 人のクラスでテストを行う．平均点 m は 60 点だったとしよう．また，標準偏差 σ は 8 だったとする．平均点から 3σ 以上離れている人は全体のごく一部で，10 人ほどしかいないはずである．なぜか．分散 V は，

$$V = \frac{1}{100}\sum_{k=1}^{100}(x_k - m)^2$$

と表されるので，$|x_k - m| \geq 3\sigma$ という人が 12 人いたら，

$$V \geq \frac{1}{100} \times 12 \times 9V$$

となって矛盾が起こるからである．よって，せいぜい 11 人までということが分かる．

一般に平均から $k\sigma$ 以上離れている人は全体の $1/k^2$ 以下である．これがチェビシェフの不等式と言われるものである．

現実の社会では 3σ というのがよく使われる．すべての分布に対して，全体の $1/9$ になることがチェビシェフの不等式から分かるが，ほとんどの場合は，特に後に習う正規分布の場合，1% 以下に収まる．例えば工場などで誤差が出ることがあっても，ほとんどの製品の誤差は 3σ 以内と考えられるし，それ以上の誤差が起こっていたら不良品であったり，何かおかしなことが起こっていると考えるのが自然だからである．3σ でなければならない特別な理由はないが，一般によく使われる数字である．

さて，全く同じことが確率の場合にも成り立つ．つまり，

$$P(|X - E(X)| \geq k\sigma(X)) \leq \frac{1}{k^2}$$

である．本来これがチェビシェフの不等式と言われる．

9.2　大数の弱法則

さてこのことを使って大数の弱法則を示そう．大数の弱法則とは「たくさん回数を重ねると高い確率で平均は期待値に近い」というものである．例えば，表の出る確率が $\frac{1}{3}$ であるようなコインを考えよう．これを 3000 回振って表の出る回数を数える．もちろん 3000 回とも表が出るということは起こりうる．その確率は 3^{-3000} ととても小さい．実は非常に高い確率で表の出る回数は $3000 \times \frac{1}{3} = 1000$ 回に近いのである．「そりゃそうだろう」と思ってもらえると有難い．では具体的にどれくらい近いのか．

表の出る確率が p であるコインを n 回投げる．X_k を k 回目が表の時に 1，裏の時には 0 となる確率変数とする．知りたいのは $X = \sum_{k=1}^{n} X_k$ の振る舞いについてである．$E(X_k) = p$, $V(X_k) = p(1-p)$ であるから，$E(X) = np$, $\sigma(X) = \sqrt{V(X)} = \sqrt{np(1-p)}$. 相対頻度の誤差を ϵ とすると，

$$P(|X/n - p| > \epsilon) = P(|X - np| > \epsilon n) = P(|X - np| > \frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}\sqrt{np(1-p)}) \leq \frac{p(1-p)}{\epsilon^2 n}$$

すなわち，

$$P(|X/n - p| \leq \epsilon) > 1 - \frac{p(1-p)}{\epsilon^2 n}$$

右辺は，n が ϵ に対して十分大きい時には，1 にかなり近い

問題 9.1. 100 点満点で 1 点きざみの試験を行ったところ，受験者が 54 名，平均値が 62.3 点，標準偏差が 8.7 点であった．得点が 36 点から 89 点の間にある受験者は何人より多いか？

証明. $\lambda = 3$ でチェビシェフの不等式を適用して，$54 \times \frac{8}{9} = 48$ □

問題 9.2 (第 10 章 A-3)**.** 以下では次の事実（大数の弱法則・ベルヌーイの定理）を使う．$B(n,p)$ に従う確率変数 X について，$\alpha > 0$ に対し，

$$P\left(p - \alpha < \frac{X}{n} < p + \alpha\right) \geq 1 - \frac{p(1-p)}{n\alpha^2}$$

硬貨を繰り返し投げるとする．

(1) 1,000 回投げて表の出る回数が 500 回より 40 回以内の偏りである確率をベルヌーイの定理を用いて評価せよ．
(2) 表の出る回数の割合が 0.5 より 5% 以内にある確率が 90% 以上であるようにするためには，少なくとも何回以上投げればよいか．ベルヌーイの定理を用いて評価式を求めよ．

証明. (1) ベルヌーイの定理を $n = 1000$, $p = 1/2$, $\alpha = 40/1000 = 1/25$ として適用すると，

$$P(\frac{1}{2} - \frac{1}{25} < X/1000 < \frac{1}{2} + \frac{1}{25} \geq 1 - \frac{1/4}{100(1/25)^2} \approx 0.844$$

(2) ベルヌーイの定理を $p = 1/2$, $\alpha = (1/2)(5/100) = 1/40$ として適用すると，

$$P(\frac{1}{2} \times 0.95 < X/n < 0.5 \times 1.05) \geq 1 - \frac{1/4}{n(1/40)^2} = \frac{4n - 1600}{4n}$$

題意を満たすためには，

$$\frac{4n - 1600}{4n} \geq 0.9$$

を解けばよい．これより $n \geq 4000$ なので 4000 回．

□

問題 9.3 (第 10 章 B-2)**.** 確率変数の列 $X_1,\ X_2,\ \cdots$ は互いに独立で，$E(X_k) = m$, $V(X_k) = \sigma^2$ $(k = 1, 2, \cdots)$ とする．

(1) $\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$ とするとき，任意の $\epsilon > 0$ に対して，

$$\lim_{n\to\infty} P(|\overline{X}_n - m| > \epsilon) = 0$$

が成り立つことを示せ．
(2) 任意の $\epsilon > 0$ と任意の $\alpha > 0$ に対して，

$$\lim_{n\to\infty} P(|\frac{X_1 + X_2 + \cdots + X_n - nm}{n^{1/2+\alpha}}| > \epsilon) = 0$$

が成り立つことを示せ．

証明. (2) を示せば，$\alpha = 1/2$ と置くことで (1) が導かれる．

$Y = X_1 + \cdots + X_n$ とおくと，$E(Y) = nm$, $V(Y) = n\sigma^2$ であるから，チェビシェフの不等式より，

$$P(|Y - nm| > k\sqrt{n}\sigma) \leq 1/k^2$$

示したい式と見比べて $k = \epsilon n^{\alpha}/\sigma$ と置けば，

$$P(|\frac{Y - nm}{n^{1/2+\alpha}}| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2 n^{2\alpha}}$$

ここで $n \to \infty$ とすると，右辺は 0 に収束する． □

10　第 10 回　相関係数，回帰直線

10.1　相関係数

大量に観察するとそこにはある一定の法則のようなものが見つかる．特に 2 つのデータに「関係がある」すなわち「相関がある」という考え方について話そう．歴史上は次のような話が有名である．

(1) ゴルトンによる身長と上腕の長さの相関
(2) ゴルトンによるスイートピーの種子の直系の測定．親を x 軸に子を y 軸に取ると，直線の傾きはだいたい $1/3$ になる．

相関 (correlation) があるかどうかは，2 つの変数 x, y に関係があるかどうかを見るので，2 つのデータを対等に見ていることに注意する．まずそのデータを図に書いてみよう．このような図を散布図 (scattergram) と呼ぶ．

(1) 人口と小売商店数の散布図
(2) 8 月の不快指数とエアコン保有率の散布図
(3) 1 世帯当たりの米の消費支出と 1 世帯当たりのパンの消費支出の散布図
(4) 出生率と死亡率の散布図

2 つの変数の間に直線関係に近い傾向が見られる時,「相関関係がある」という．もう少し身近なところでは,

(1) 身長が高い人は体重も重い.
(2) 数学の点数が高い人は国語の点数も高い.
(3) 気温が高い日はアイスクリームの売り上げも高い.
(4) 交通量が多い交差点は交通事故も多い.
(5) 収入と結婚率
(6) 人口密度とコンビニ密度
(7) 父親の身長と子供の身長
(8) 出席率とテストの点数

などがある.

その関係の強さを表す指標が，相関係数 (correlation coefficient) であり，ピアソンにより導入された.

$$r = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum (x_i - \overline{x})^2} \cdot \sqrt{\sum (y_i - \overline{y})^2}}$$

ここで,

$$C_{xy} = \frac{1}{n} \sum (x_i - \overline{x})(y_i - \overline{y}) = \frac{1}{n} \sum x_i y_i - \overline{xy}$$

を共分散 (covariance) と呼ぶので,

$$r = \frac{C_{xy}}{\sigma_x \sigma_y}$$

とも表せる.

相関係数は,

$$-1 \leq r \leq 1$$

の式を満たす．まず，$\overline{x} = \overline{y} = 0$, $\sigma_x = \sigma_y = 1$ のときは,

$$\frac{1}{n} \sum (x_i \pm y_i)^2 = \frac{1}{n} \sum x_i^2 \pm 2 \frac{1}{n} \sum x_i y_i + \frac{1}{n} \sum y_i^2 = 2(1 \pm r_{xy}) \geq 0$$

さらに，$x' = ax + b$, $y' = cy + d$, $ac > 0$, と線型変換した共分散を C'_{xy}，相関係数を r'_{xy} とすると,

$$C'_{xy} = \frac{1}{n} \sum ((ax_i + b) - (a\overline{x} + b))((cy_i + d) - (c\overline{y} + d)) = acC_{xy}$$

より，

$$r'_{xy} = \frac{acC_{xy}}{|a|\sigma_x|c|\sigma_y} = r_{xy}$$

このことから，相関係数 r は $-1 \leq r \leq 1$ となることが示された．また，$|r| = 1$ の場合は，適当な線型変換で $x_i = y_i$ となる場合であるから，一直線上に乗っている場合であることが分かる．

相関関係と因果関係は異なる．

(1) アイスクリームの売り上げが伸びると水死者数も増える．
(2) 朝食を食べている生徒は成績が良い．
(3) 景気が良くなれば株価は上がる．

実際に相関係数を手計算で求めることは滅多にない．Excel にデータを入力して，関数を適切に指定すれば出てくる．

10.2 回帰直線

今度は一方の変数がもう一方を説明していいると思うことにしよう．独立変数 (independent variable) と従属変数 (dependent variable) と呼ぶ．説明変数，被説明変数ということもある．

n 個のデータの組 (x_i, y_i) があったとして，x が y を説明するとして，$y = b + ax$ とする．

$$D(a,b) = \frac{1}{n}\sum_{i=1}^{n}(y_i - (b + ax_i))^2$$

これが最小となる a, b を求める．この方法を最小二乗法という．

$$\frac{\partial D}{\partial a} = -2\sum_{i=1}^{n}(y_i - b - ax_i)x_i = 0,$$
$$\frac{\partial D}{\partial b} = -2\sum_{i=1}^{n}(y_i - b - ax_i) = 0$$

を解こう．a, b に関して整理すると，

$$(\sigma_x^2 + m_x^2)a + m_x b = \sigma_{xy} + m_x m_y,$$
$$m_x a + b = m_y$$

となる．これより，

$$a = \frac{\sigma_{xy}}{\sigma_x^2}, \quad b = m_y - \frac{\sigma_{xy}}{\sigma_x^2}m_x$$

となる．前者は $a = r\frac{\sigma_y}{\sigma_x}$ とも書ける．後者の式は (m_x, m_y) が回帰直線上にあることを意味している．また，この時 $D(a,b)$ の最小値は，

$$R^2 = \sum(y_i - (b + ax))^2 = (1 - r_{xy})^2\sigma_y^2$$

とも書ける．よって，相関係数 r_{xy} は直線関係のあてはまりの良さの尺度であるとも言える．

実例を Excel で見てみよう．

問題 10.1. 次の表は，同じ種類の 5 本の木の太さ x(cm) と高さ y(m) を測定した結果である．x と y の相関係数 r を求めよ．また，x を説明変数，y を被説明変数としたときの回帰直線を求めよ．

木の番号	1	2	3	4	5
x	22	27	29	19	33
y	13	15	18	14	20

証明.

$$\begin{aligned}
\overline{x} =& (22+27+29+19+33)/5 = 26 \\
\overline{y} =& (13+15+18+14+20)/5 = 16 \\
v_x =& (22^2+27^2+29^2+19^2+33^2)/5 = 24.8 \\
v_y =& 6.8 \\
\frac{1}{n}\sum x_i y_i =& 427.8 \\
\sigma_{xy} =& 11.8 \\
r =& 0.908659\cdots \\
y =& 0.476x + 3.624
\end{aligned}$$

□

問題 10.2. 「相関関係はあるが因果関係はない」例を挙げよ．面白いものに加点する．

11　第 11 回　連続的確率分布，指数分布

これまで確率変数は自然数などの飛び飛びの値を取ると仮定してきた．このような確率分布を離散的確率分布という．これに対し，実数のような連続的に値が変化する場合には，連続的確率変数や連続的確率分布という．

例えば，ある工場で 100g のパンを作っているとしよう．100g のパンと一言で言うが，厳密には 100g ではない．だいたい 100g くらいということであろう．実際に測ってみれば，105g や 98g ということもあるだろう．平均が 100g に十分近くて，分散が小さいほど，このパンの重さに関する精度が良い，と言えるだろう．そこでこのパンの重さを確率変数 X として，確率分布を考える．測った場合には整数値になるだろうが，重さは連続的に変化すると考え，X は連続的に変化する確率変数と考えるのが良いだろう．

離散的確率変数の場合には，$P(X = k)$ という値を考えた．連続的確率変数の場合には，$P(X = k)$ という値を考えても，ほとんど 0 となり，あまり意味はない．それよりも，$P(a < X < b)$ という値を考える方が自然である．離散的な場合，確率分布と言った場合には $P(X = k)$ の組もしくは表を指した．連続歴確率変数の場合には，累積分布関数 $F_X(x) = P(X \leq x)$ を指すことが多い．累積分布関数が微分可能であれば，$F'(x) = f(x)$ として，

$$P(a < X < b) = \int_a^b f(x)dx$$

と表すことができる．この f を X の密度関数という．f が密度関数であれば，$\int_{-\infty}^{\infty} f(x)dx = 1$ となるので，$P(a < X < b)$ とは，f のグラフの a から b までの面積にあたる．

例として 0 から 1 までを均等に取る確率変数 X を考えよう．このような分布を一様分布という．明らかに，$0 < a < b < 1$ ならば，

$$P(a < X < b) = b - a$$

である．また，分布関数は

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

となる．よって，密度関数は

$$f_X(x) = \begin{cases} 1 & (0 \leq x \leq 1) \\ 0 & (o.w.) \end{cases}$$

となる．

今，$a > 0, b$ を定数として，$Y = aX + b$ という確率変数を考えてみよう．Y の分布関数は，

$$F_Y(x) = P(Y \leq x) = P(aX + b \leq x) = P(X \leq \frac{x-b}{a}) = F_X(\frac{x-b}{a})$$

であるから，密度関数は

$$f_Y(x) = \frac{1}{a} f_X(\frac{x-b}{a})$$

である．これを用いて，離散的確率変数の場合と同様に期待値と分散が次のように定義される．

$$\begin{aligned} E(X) =& \int_{-\infty}^{\infty} x f(x)dx \\ V(X) =& E((X - E(X))^2) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)dx \end{aligned}$$

離散の場合と同様に，

$$\begin{aligned} E(aX + b) =& aE(X) + b \\ E(X + Y) =& E(X) + E(Y) \\ V(aX + b) =& a^2 V(X) \\ V(X) =& E(X^2) - (E(X))^2 \end{aligned}$$

などが成り立つ.

例えば，0 から 1 までの一様分布 $U(0,1)$ の場合，密度関数は $f(x)=1,\ 0\leq x\leq 1$ なので,

$$E(X)=\int_0^1 x\ dx=\frac{1}{2}$$

また,

$$V(X)=\int_0^1 x^2\ dx=\frac{1}{3}$$

である.

問題 11.1. $U(a,b)$ の期待値と分散を求めよ.

証明. 密度関数は $f(x)=\frac{1}{b-a},\ a\leq x\leq b$ なので,

$$E(X)=\int_a^b x\frac{1}{b-a}dx=[\frac{x^2}{2(b-a)}]_a^b=\frac{a+b}{2}$$

である．また,

$$E(X^2)=\int_a^b x^2\frac{1}{b-a}dx=\frac{b^3-a^3}{3(b-a)}=\frac{b^2+ab+b^2}{3}$$

よって,

$$V(X)=\frac{a^2+ab+b^2}{3}-\frac{a^2+2ab+b^2}{4}=\frac{1}{12}(a^2+b^2-2ab)=\frac{(b-a)^2}{12}$$

□

問題 11.2. 正のパラメータ λ に対して，$f(x)=\lambda e^{-\lambda x},\ x>0$ を密度関数とする確率分布を指数分布という．次に頻繁には起こらない現象に対して，起こるまでの時間の分布を表す．期待値と分散を求めよ.

証明.

$$\int_0^\infty ae^{-ax}dx=[-e^{-ax}]_0^\infty=1$$

より，確かに確率分布になっている．期待値は,

$$E(X)=\int_0^\infty xae^{-ax}dx=[-xe^{-ax}]_0^\infty-\int_0^\infty(-e^{-ax})dx=[-e^{-ax}/a]_0^-\infty=\frac{1}{a}.$$

分散は,

$$E(X^2)=\int_0^\infty x^2ae^{-ax}dx=[-x^2e^{-ax}]_0^\infty-\int_0^\infty(-2xe^{-ax})dx=\frac{2}{a^2}$$

より,

$$V(X)=E(X^2)-(E(X))^2=\frac{1}{a^2}$$

□

指数分布．工場製品の寿命，次に事故が起こるまでの時間，次に電話がかかってくるまでの時間.

幾何分布同様に無記憶性の性質を持つ.

$$P(X>t)=\int_t^\infty ae^{-ax}dx=[-e^{-ax}]_t^\infty=e^{-at}$$

よって,

$$P(X>s+t\mid X>t)=\frac{P(X>s+t)}{P(X>t)}=\frac{e^{-a(s+t)}}{e^{-at}}=e^{-as}=P(X>s).$$

ポアソン分布は単位時間当たり平均 λ 回起こる事象が，単位時間当たり k 回起こる確率を表している．指数分布は平均待ち時間 $\frac{1}{a}$ の事象の待ち時間の分布を表している．なのでちょうど逆の関係になっている．このことをもう少し詳しく見てみよう.

ポアソン分布 $P(X=k)=\frac{\lambda^k e^{-\lambda}}{k!}$ において，時間 t の間には平均 $t\lambda$ 回起きるので，時間 t の間に起こる回数を Y とすると，

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

初めて起こるまでの待ち時間 Z が t よりも大きいという事象は，時間 t の間にこの事象が起こらない確率で，

$$P(Z>t)=P(Y=0)=e^{-t\lambda}$$

よって，$P(Z\leq t)=1-e^{-t\lambda}$ であり，その密度関数は微分して $f(x)=\lambda e^{-t\lambda}$ である．これは指数分布である．

まず，互いに独立な確率変数 X,Y に対して，$X+Y$ の確率密度関数を求めよう．$X,Y,X+Y$ の密度関数を f,g,h，累積分布関数を F,G,H とする．

$$\begin{aligned}H(t)=P(X+Y\leq t)&=\int_{x+y\leq t} f(x)g(y)dxdy\\&=\int_{-\infty}^{\infty}(\int_{-\infty}^{t-y} f(x)dx)g(y)dy\\&=\int_{-\infty}^{\infty} F(t-y)g(y)dy\end{aligned}$$

よって，

$$\begin{aligned}h(t)=&\frac{d}{dt}\int_{-\infty}^{\infty} F(t-y)g(y)dy\\=&\int_{-\infty}^{\infty} f(t-y)g(y)dy\end{aligned}$$

これを f と g のたたみ込みという．

平均待ち時間 $1/\lambda$ の事象について，1 回目に起こるまでの時間を X_1, $k\geq 2$ に対しては $k-1$ 回起こってから k 回目に起こるまでの時間を X_k とする．すると，X_i はすべて独立で，すべてパラメータ λ の指数分布に従う．

$$Y_k=X_1+X_2+\cdots+X_k$$

とおくと，Y_k の確率密度関数は，

$$f_k(x)=\frac{\lambda^k x^{k-1}}{(k-1)!}\exp(-\lambda x)$$

となることを k の帰納法で示そう．$k=1$ のときは通常の指数分布となることが確認できる．k のときにこの形で表されるとすると，

$$\begin{aligned}f_{k+1}(x)=&\int_{-\infty}^{\infty} f_1(x-y)f_k(y)dy\\=&\int_0^x \lambda e^{-\lambda(x-y)}\frac{\lambda^k y^{k-1}}{(k-1)!}\exp(-\lambda y)dy\\=&\frac{\lambda^{k+1}e^{-\lambda x}}{(k-1)!}\int_0^x y^{k-1}dy\\=&\frac{\lambda^{k+1}x^k}{k!}\exp(-\lambda x)\end{aligned}$$

となる．これより，

$$\begin{aligned}P(Y_{k+1}\leq 1)=&\int_0^1 \frac{\lambda^{k+1}x^k}{k!}\exp(-\lambda x)dx\\=&[-\frac{\lambda^k x^k}{k!}\exp(-\lambda x)]_0^1+\int_0^1\frac{\lambda^k x^{k-1}}{(k-1)!}\exp(-\lambda x)dx\\=&-\frac{\lambda^k e^{-\lambda}}{k!}+P(Y_k\leq 1)\end{aligned}$$

さて単位時間あたりに起こる事象の回数を Z とすると，$Z = k$ ということは，$Y_k \leq 1$ かつ $Y_{k+1} > 1$ ということである．よって，

$$P(Z = k) = P(Y_k \leq 1) - P(Y_{k+1} \leq 1) = \frac{\lambda^k e^{-\lambda}}{k!}$$

これはポアソン分布である．

12　第 12 回　正規分布

12.1　正規分布とは何か

連続的確率分布の中でも最も重要な分布が正規分布である．その英語名 Normal distribution からもそのことが感じられるだろう．

正規分布する例としては,

(1) ジュースやお菓子の袋の重さ（すなわち誤差が正規分布するということ）
(2) 身長や体重，成績など

とにかく非常に多くのものが正規分布することが知られている．

正規分布 $N(\mu, \sigma^2)$ は Gaussian distribution とか Normal distribution とも呼ばれる．密度関数は

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

である．特に $N(0,1)$ を標準正規分布という．その形から釣鐘型などとも言われる．標準正規分布 $N(0,1)$ において,

$$\phi(u) = P(0 \leq X \leq u)$$

の値は非常に重要なので，標準正規分布表がたいていの統計の教科書には載っている．

二項分布 $B(n,p)$ は期待値が np，分散が $np(1-p)$ であった．X が二項分布するとき，$Y = \frac{X-np}{\sqrt{np(1-p)}}$ は，ほとんど正規分布することが知られている．より精度の良い式として，以下の半数補正した定理が知られている．

定理 12.1. n が大きいとき，二項分布 $B(n,p)$ における整数 a, b の間の確率 $P(a \leq X \leq b)$ は標準正規分布 $N(0,1)$ における $(a-0.5-np)/\sqrt{np(1-p)}$ と $(b+0.5-np)/\sqrt{np(1-p)}$ の間の確率に近い．

例 12.1,12.2

問題 12.2.　(1) 標準正規分布に従う確率変数 X の平均 0 からの距離が 2.00 以下である確率を，標準正規分布表を使って求めよ．
(2) 標準正規分布に従う確率変数 X について，$P(|X| > d) = 0.05$ となるような d を，標準正規分布表を使って求めよ．
(3) 確率変数 X は正規分布 $N(m, \sigma^2)$ に従うとする．確率変数 X の平均 m からの距離が 3σ 以下である確率を，標準正規分布表を使って求めよ．

証明.　(1) $2 \times \phi(2.00) = 2 \times 0.4772 = 0.9544$
(2) $\phi(d) = 0.475$ となる d は 1.96
(3) $2 \times \phi(3.00) = 2 \times 0.49865 = 0.9973$

□

問題 12.3. コインを 100 回投げた時，表が出る相対頻度が 0.48 から 0.52 の間にある確率を，以下の事実を使って評価せよ．

正規分布近似．n が大きい時，$B(n,p)$ における $P(a \leq X \leq b)$ は，$N(0,1)$ における $P((a-0.5-np)/\sqrt{np(1-p)} \leq X \leq (b+0.5-np)/\sqrt{np(1-p)})$ に近い．

証明. $n = 100$, $p = 1/2$, $a = 50-2$, $b = 50+2$, $\sqrt{np(1-p)} = 5$ より，$N(0,1)$ における $P(-0.5 \leq X \leq 0.5)$ に近い．つまり，$2 \times \phi(0.5) = 2 \times 0.1915 = 0.383$　□

12.2 期待値と分散

これが密度関数になり得ることは次のように確かめられる.

$$I = \int_{-\infty}^{\infty} \exp(-x^2/2)dx$$

とおくと,

$$I^2 = \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dxdy = \int_0^{2\pi}\int_0^{r} e^{-r^2/2}rdrd\theta = 2\pi$$

よって, $I = \sqrt{2\pi}$.

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-y^2/2\right)\sigma dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx = 1$$

ここで $y = \frac{x-\mu}{\sigma}$ と置換した.

期待値は

$$E(X) = \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi}} \exp(-x^2/2)dx = [-\exp(-x^2/2)]_{-\infty}^{\infty} = 0$$

分散は

$$V(X) = \int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi}} \exp(-x^2/2)dx = [\frac{x}{\sqrt{2\pi}}(-\exp(-x^2/2))]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}(-\exp(-x^2/2))dx = 1$$

と求まる. 一般の場合は $Y = \frac{X-\mu}{\sigma}$ から求まる.

12.3 正規分布の再帰性

X, Y はそれぞれ $N(\mu_x, \sigma_x^2), N(\mu_y, \sigma_y^2)$ に従う独立な確率変数としよう. このとき $X+Y$ は $N(\mu_x+\mu_y, \sigma_x^2+\sigma_y^2)$ に従うことを示そう. 今, $\mu_x = 0$ として一般性を失わない. $\mu_y = u, \sigma_x^2 = s^2, \sigma_y^2 = t^2$ と書く.

$$\begin{aligned} h(x) &= \int_{-\infty}^{\infty} f(x-y)g(y)dy \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi s^2}} \exp\left(-\frac{(x-y)^2}{2s^2}\right) \frac{1}{\sqrt{2\pi t^2}} \exp\left(-\frac{(y-u)^2}{2t^2}\right) dy \end{aligned}$$

exp の中身だけ取り出して計算すると,

$$-\frac{(x-y)^2}{2s^2} - \frac{(y-u)^2}{2t^2} = -\frac{1}{2s^2t^2}\left((s^2+t^2)\left(y - \frac{xt^2+us^2}{s^2+t^2}\right)^2 + x^2t^2 + s^2u^2 - \frac{(xt^2+us^2)^2}{s^2+t^2}\right)$$

平方完成の部分は y で積分されて x を含まないなんらかの定数になる. 後半部分をさらに計算すると,

$$-\frac{(x-y)^2}{2s^2} - \frac{(y-u)^2}{2t^2} = -\frac{(x-u)^2}{2(s^2+t^2)}$$

となる. すなわち,

$$h(x) = C\exp\left(-\frac{(x-u)^2}{2(s^2+t^2)}\right)$$

これがなんらかの分布の密度関数になることは分かっているから, $N(u, s^2+t^2)$ であり, $C = \sqrt{2\pi(s^2+t^2)}$. これが示したいことであった.

12.4　二項分布の正規分布による近似の証明

二項分布が正規分布で近似されることを示そう．二項分布 $B(n,p)$ において，$q=1-p$ として $Y=\frac{X-np}{\sqrt{npq}}$ を考える．$X=k$ のとき $Y=t$ となるとすれば，当然 $P(Y=t)=P(X=k)$ となる．Y の確率分布は $\sqrt{npq}$ ごとに分かれていることを考えると，極限の確率密度関数としての値は，

$$f_n(t)=\sqrt{npq}\frac{n!}{k!(n-k)!}p^k q^{n-k}$$

$n\to\infty$ としたときのこの値が，$\frac{1}{\sqrt{2\pi}}\exp(-t^2/2)$ であることを示す．

まずスターリングの公式 $n!\approx\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$ より，

$$f_n(t)\approx\sqrt{npq}\frac{\sqrt{2\pi n}}{\sqrt{2\pi k}\sqrt{2\pi(n-k)}}\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}$$

となる．ここで $t=\frac{k-np}{\sqrt{npq}}$ より，$\frac{k}{n}\to p$, $\frac{n-k}{n}\to q$ だから，前半部分は $\frac{1}{\sqrt{2\pi}}$ に収束する．そこで，後半部分を

$$g_n(t)=\left(\frac{np}{k}\right)^k\left(\frac{nq}{n-k}\right)^{n-k}$$

とおくと，

$$\frac{k}{np}=1+t\sqrt{\frac{q}{np}}$$

より，

$$\log\frac{k}{np}\approx t\sqrt{\frac{q}{np}}-\frac{t^2q}{2np}$$

同様にして，

$$\frac{n-k}{nq}=1-t\sqrt{\frac{p}{nq}}$$

より，

$$\log\frac{n-k}{nq}\approx -t\sqrt{\frac{p}{nq}}-\frac{t^2p}{2nq}$$

よって，

$$\begin{aligned}\log g_n(t)&\approx-(np+t\sqrt{npq})(t\sqrt{\frac{q}{np}}-\frac{t^2q}{2np})-(nq-t\sqrt{npq})(-t\sqrt{\frac{p}{nq}}-\frac{t^2p}{2nq})\\&=-t\sqrt{npq}+t\sqrt{npq}-t^2q-t^2p+\frac{t^2q}{2}+\frac{t^2q}{2}+\frac{t^3q^{3/2}}{2}-\frac{t^3p^{3/2}}{2}\\&\approx-\frac{t^2}{2}\end{aligned}$$

よって，$f_n(t)\to\frac{1}{\sqrt{2\pi}}\exp(-\frac{t^2}{2})$.

13　第 13 回　検定

13.1　統計的仮説検定

コインを n 回投げて，何回表が出るかを調べよう．

もしコインの表が出る確率が $\frac{1}{2}$ だと分かっているならば，確率論により表が出る回数の確率を求めることができる．大数の法則によりだいたい $\frac{n}{2}$ に近く，その誤差の収束の速さは重複対数の法則により $\frac{1}{\sqrt{n}}$ くらいであることも分かる．このようなコインを振る前の話が確率である．

では，逆に表の出る確率は分からないとして，コインを n 回投げたところ，表の出た回数が k 回であった．この時に表の出る確率は確率はどう推測できるだろうか．このようなコインを降った後の話が統計である．

今，コインが公平かどうかを知りたいとする．n, k がいくつの時に，「コインの表が出る確率が $\frac{1}{2}$ である」と言うことができるだろうか．逆に n, k がいくつの時には，「コインの表が出る確率は $\frac{1}{2}$ ではない」と言うことができるだろうか．少し考えると，n, k がいくつであったとしても，どちらとも言うことはできないことが分かるだろう．このような歯切れの悪さが「統計は数学ではない」と言われる所以だと思われる．

しかし，何も言えないという訳ではない．「コインの表が出る確率は $\frac{1}{2}$ に近そうだ」とか，「コインの表が出る確率が $\frac{1}{2}$ であるというのは不自然だ」ということなら，言うことができる．そしてその不自然の度合いを数字で表すことができるのである．

「コインの表が出る確率は $\frac{1}{2}$ であるというのは不自然だ」ということを結論するために，今，仮に，「コインの表が出る確率は $\frac{1}{2}$ である」と仮定して，その帰無仮説が不自然であることを示す．n が十分大きければ，表の出る回数 k はほとんど正規分布に従うことが分かっている．そこで「表の出た回数だけに注目し，平均から離れている 5% が出たら不自然である」と結論することにしよう．なぜ「表の出た回数」だけに注目するのか，なぜ平均から離れている場所だけ棄却するのか，このあたりは「先輩方がそうしたから」であり，恣意的である．後にそうではないものも考える．このように問題を設定して，1000 回降って，表が 953 回出たとする．これが平均からとても離れているので，不自然であり，もとの仮説は棄却される．

問題 13.1 (第 13 章 A-2)**.** ある自動車メーカーはある車種のガソリン 1ℓ 当たりの走行距離を $18.5\mathrm{km}/\ell$ にしている．ある時期に製造された 8 台の車の実験走行の平均走行距離は，$17.95\mathrm{km}/\ell$ であった．この期間の車の走行距離は規格からずれているかどうか，危険率 5% で検定せよ．ただし，走行距離は正規分布に従うことが知られていて，標準偏差は $1.5\mathrm{km}/\ell$ であるとする．

証明. 8 台の車の燃費 X_i はそれぞれ $N(18.5, 1.5^2)$ に従う．よって，その平均燃費 X は $N(18.5, 1.5^2/8)$ に従う．$X = 17.95$ を $Z = \frac{X-18.5}{1.5/\sqrt{8}}$ で変換すると，$Z = -1.03709$ である．危険率 5% の両側検定では，$-1.960 < Z < 1.960$ の範囲に入っているので，規格からはずれているとは認められない．　□