

標本分布について

- 黒木玄
- 2022-04-11~2022-04-29

このノートでは[Julia言語 \(https://julialang.org/\)](https://julialang.org/)を使用している:

- [Julia言語のインストールの仕方の一例 \(https://nbviewer.org/github/genkuroki/msfd28/blob/master/install.ipynb\)](https://nbviewer.org/github/genkuroki/msfd28/blob/master/install.ipynb)

自明な誤りを見つけたら、自分で訂正して読んで欲しい。大文字と小文字の混同や書き直しが不完全な場合や符号のミスは非常によくある。

このノートに書いてある式を文字通りにそのまま読んで正しいと思ってしまうとひどい目に会う可能性が高い。しかし、数が使われている文献には大抵の場合に文字通りに読むと間違っている式や主張が書いてあるので、内容を理解した上で訂正しながら読んで利用しなければいけない。実践的に数学を使う状況では他人が書いた式をそのまま信じていけない。

このノートの内容よりもさらに詳しいノートを自分で作ると勉強になるだろう。膨大な時間を取られることになるが、このノートの内容に関係することで飯を食っていく可能性がある人にはそのためにかけた時間は無駄にならないと思われる。

目次

- ▼ [1 標本分布とその使い方](#)
 - [1.1 同時確率質量関数と同時確率密度関数](#)
 - ▼ [1.2 確率変数の独立性の定義](#)
 - [1.2.1 独立な確率変数達の同時確率質量関数](#)
 - [1.2.2 独立な確率変数達の同時確率密度関数](#)
 - [1.2.3 独立性に関する大雑把なまとめ](#)
 - [1.2.4 分散を0に近付けたときの正規分布について](#)
 - [1.3 独立同分布\(i.i.d., iid\)の定義](#)
 - [1.4 標本分布の定義](#)
 - [1.5 確率分布の積と \$n\$ 乗の定義](#)
 - [1.6 分布 \$D\$ の標本分布の主な使用用途](#)
 - [1.7 試行回数 \$n\$ のBernoulli試行の分布はBernoulli分布の標本分布](#)
 - [1.8 二項分布による推定の確率的揺らぎの記述](#)
- ▼ [1.9 問題: 大阪都構想に関する住民投票の結果について](#)
 - [1.9.1 Julia言語による計算の例](#)
 - [1.9.2 WolframAlphaによる計算の例:](#)
 - [1.9.3 Clopper-Pearsonの信頼区間とそれを与えるP値](#)
 - [1.9.4 信頼区間よりも情報量が大きなP値関数のプロット](#)
 - [1.9.5 Sternの信頼区間とそれを与えるP値関数](#)
 - [1.9.6 Sternの信頼区間を与えるP値関数の実装例](#)
 - [1.10 対ごとに独立であっても全体が独立であるとは限らない](#)
 - [1.11 確率変数の独立性の現実における解釈に関する重大な注意](#)
- ▼ [2 確率変数達の共分散と相関係数と無相関性](#)
 - [2.1 確率変数達の共分散と相関係数の定義](#)
 - [2.2 確率変数達の無相関の定義](#)
 - [2.3 問題: 確率変数の相関係数の計算例](#)
 - [2.4 問題: Cauchy-Schwarzの不等式](#)
 - [2.5 問題: 等確率有限離散分布の相関係数と \$\cos \theta\$ の関係](#)
 - [2.6 問題: 相関係数の計算](#)
 - [2.7 問題: 共分散が0に近くても相関係数が0から遠い場合がある](#)
 - [2.8 問題: 独立ならば無相関である\(実質1行で解ける\)](#)
 - [2.9 問題: 無相関でも独立とは限らない](#)
 - [2.10 問題: 無相関な確率変数達の和の分散はそれぞれの分散の和になる](#)
 - [2.11 問題: 二項分布と負の二項分布の平均と分散のBernoulli分布と幾何分布の場合への帰着](#)
 - [2.12 問題: 番号が異なる確率変数達が無相関なときの確率変数の和の共分散](#)
 - [2.13 分散共分散行列とその半正定値性](#)
- ▼ [3 標本\(サンプル, データ\)の平均と分散と共分散と相関係数](#)
 - [3.1 標本平均の定義](#)
 - [3.2 問題: 無相関な確率変数達の標本平均の分散](#)
 - [3.3 標本分散と不偏分散の定義](#)
 - [3.4 不偏推定量について: 不偏分散の定義ではどうして \$n\$ ではなく \$n-1\$ で割るか](#)
 - [3.5 データの共分散の定義](#)
 - [3.6 問題: 標本平均達の共分散](#)

- 3.7 問題: 不偏共分散の定義ではどうして n ではなく $n - 1$ で割るか
- 3.8 データの相関係数の定義 (以上の定義のまとめにもなっている)
- 3.9 問題: 最小二乗法
- 3.10 問題: 計算例
- ▼ 4 モーメントとキュムラントと歪度と尖度
 - 4.1 モーメントとその母関数と特性関数とキュムラント母関数の定義
 - 4.2 特性関数による期待値の表示
 - 4.3 問題: 分布のアフィン変換のキュムラント
 - 4.4 問題: 標準正規分布のモーメント母関数と特性関数とキュムラント母関数
 - 4.5 確率変数の標準化と標準化キュムラントと歪度と尖度
 - 4.6 問題: 正規分布の歪度と尖度
 - 4.7 問題: 一様分布のキュムラント母関数と歪度と尖度
 - 4.8 問題: 独立な確率変数達の和のモーメント母関数と特性関数とキュムラント母関数
 - 4.9 問題: ガンマ分布のキュムラント母関数と歪度と尖度
 - 4.10 問題: 二項分布のキュムラント母関数と歪度と尖度
- ▼ 5 独立同分布な確率変数達の不偏分散の分散
 - 5.1 期待値 0, 分散 1 の場合への帰着
 - 5.2 標本平均と不偏分散の共分散の計算
 - 5.3 不偏分散の分散の計算
 - 5.4 歪度と尖度に関する不等式
- 6 正規分布の標本分布の場合

```
In [1]: 1 ENV["LINES"], ENV["COLUMNS"] = 100, 100
2 using BenchmarkTools
3 using Distributions
4 using Printf
5 using QuadGK
6 using Random
7 Random.seed!(4649373)
8 using Roots
9 using SpecialFunctions
10 using StaticArrays
11 using StatsBase
12 using StatsFuns
13 using StatsPlots
14 default(fmt = :png, titlefontsize = 10, size = (400, 250))
15 using SymPy
```

1 標本分布とその使い方

1.1 同時確率質量関数と同時確率密度関数

確率変数達 X_1, \dots, X_n が同時確率質量関数 $P(x_1, \dots, x_n)$ を持つとは, $P(x_1, \dots, x_n) \geq 0$ がかつ, $P(x_1, \dots, x_n)$ の総和が 1 であり,

$$E[f(X_1, \dots, X_n)] = \sum_{x_1} \cdots \sum_{x_n} f(x_1, \dots, x_n) P(x_1, \dots, x_n)$$

が成立することである. このとき, X_i 単独の確率質量関数 $P(x_i)$ は

$$P(x_i) = \sum_{x_1} \cdots \widehat{\sum_{x_i}} \cdots \sum_{x_n} P(x_1, \dots, x_i, \dots, x_n)$$

になる. ここで, $\widehat{\sum_{x_i}}$ は \sum_{x_i} を除くという意味である. なぜならば,

$$\begin{aligned} E[f(X_i)] &= \sum_{x_1} \cdots \sum_{x_i} \cdots \sum_{x_n} f(x_i) P(x_1, \dots, x_i, \dots, x_n) \\ &= \sum_{x_i} f(x_i) \left(\sum_{x_1} \cdots \widehat{\sum_{x_i}} \cdots \sum_{x_n} P(x_1, \dots, x_i, \dots, x_n) \right). \end{aligned}$$

注意: $i \neq j$ のとき $P(x_i)$ と $P(x_j)$ は一般に異なる確率質量関数になることに注意せよ. 区別を明瞭にするためには $P(x_i)$ を $P_i(x_i)$ のように書くべきであるが, 統計学の世界での慣習に従ってこのような記号法上の省略を行った. これが気に入らない人は自分でノートを作るときに, 区別されるべきものに異なる記号を割り振るようにすればよいだろう. 以下についても同様である.

確率変数達 X_1, \dots, X_n が同時確率密度関数 $p(x_1, \dots, x_n)$ を持つとは, $p(x_1, \dots, x_n) \geq 0$ がかつ, $p(x_1, \dots, x_n)$ の積分が 1 になり,

$$E[f(X_1, \dots, X_n)] = \int \cdots \int f(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_1 \cdots dx_n$$

が成立することである。ここで \int は定積分を表す。このとき、 X_i 単独の確率密度関数 $p(x_i)$ は

$$p(x_i) = \int \cdots \int \cdots \int p(x_1, \dots, x_i, \dots, x_n) dx_1 \cdots \widehat{dx_i} \cdots dx_n$$

になる。ここで、 $\widehat{\int}$, $\widehat{dx_i}$ はそれらを除くことを意味する。なぜならば、

$$\begin{aligned} E[f(X_i)] &= \int \cdots \int \cdots \int f(x_i) p(x_1, \dots, x_i, \dots, x_n) dx_1 \cdots dx_i \cdots dx_n \\ &= \int f(x_i) \left(\int \cdots \int \cdots \int p(x_1, \dots, x_i, \dots, x_n) dx_1 \cdots \widehat{dx_i} \cdots dx_n \right) dx_i. \end{aligned}$$

1.2 確率変数の独立性の定義

確率変数達 X_1, \dots, X_n が与えられており、それらの関数の期待値 $E[f(X_1, \dots, X_n)]$ が定義されているとする。このとき、 X_1, \dots, X_n が **独立** (independent) であるとは、

$$E[f_1(X_1) \cdots f_n(X_n)] = E[f_1(X_1)] \cdots E[f_n(X_n)] \quad (1)$$

のように X_i ごと関数達 $f_i(X_i)$ の積の期待値が各々の $f_i(X_i)$ の期待値の積になることだと定める。

注意: 厳密には関数 f_i 達を動かす範囲を確定させる必要があるが、このノートではそのようなことにこだわらずに解説することにする。厳密な理論の展開のためには測度論的確率論の知識が必要になるが、そのような方向はこのノートの目標とは異なる。測度論的確率論に興味がある人は別の文献を参照して欲しい。ただし、測度論的確率論の理解と統計学の理解は別な話題になってしまうことには注意して欲しい。測度論的確率論と統計学では興味の方向が異なる。

1.2.1 独立な確率変数達の同時確率質量関数

確率変数達 X_1, \dots, X_n が同時確率質量関数 $P(x_1, \dots, x_n)$ を持つとき、 X_i 単独の確率質量関数を $P_i(x_i)$ と書くならば、確率変数達 X_1, \dots, X_n が独立であることと、

$$P(x_1, \dots, x_n) = P_1(x_1) \cdots P_n(x_n) \quad (2)$$

が成立することは同値である。すなわち、 (X_1, \dots, X_n) の関数の期待値が次のように表されることが同値である：

$$E[f(X_1, \dots, X_n)] = \sum_{x_1} \cdots \sum_{x_n} f(x_1, \dots, x_n) P_1(x_1) \cdots P_n(x_n)$$

注意: ここでは気分を変えて、 $P_i(x_i)$ をずばらに $P(x_i)$ と書く流儀をやめてみた。

証明: (2)が成立しているならば、

$$\begin{aligned} E[f_1(X_1) \cdots f_n(X_n)] &= \sum_{x_1} \cdots \sum_{x_n} f_1(x_1) \cdots f_n(x_n) P_1(x_1) \cdots P_n(x_n) \\ &= \sum_{x_1} f_1(x_1) P_1(x_1) \cdots \sum_{x_n} f_n(x_n) P_n(x_n) \\ &= E[f_1(X_1)] \cdots E[f_n(X_n)] \end{aligned}$$

と(1)が成立する。

逆に(1)が成立しているならば、 $f_i(x_i)$ として $x_i = a_i$ のときにのみ 1 でそれ以外ときに 0 になる関数を取ると、

$$P(a_1, \dots, a_n) = E[f_1(X_1) \cdots f_n(X_n)] = E[f_1(X_1)] \cdots E[f_n(X_n)] = P_1(a_1) \cdots P_n(a_n)$$

なので、(2)が成立する。

1.2.2 独立な確率変数達の同時確率密度関数

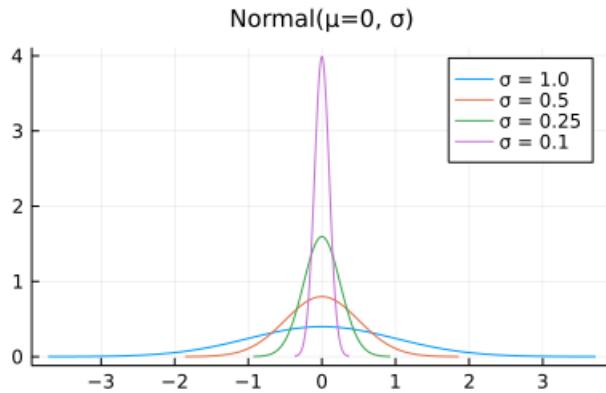
確率変数達 X_1, \dots, X_n が同時確率密度関数 $p(x_1, \dots, x_n)$ を持つとき、 X_i 単独の確率密度関数を $p_i(x_i)$ と書くならば、確率変数達 X_1, \dots, X_n が独立であることと、

$$p(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n) \quad (3)$$

が成立することは同値である。すなわち、 (X_1, \dots, X_n) の関数の期待値が次のように表されることが同値である：

In [3]: 1 var"正規分布は分散が小さくなると一点に集中して行く."

Out[3]:



$E[f(X_0)] = f(\mu)$ を満たす確率変数 X_0 は実質定数 μ に等しく、実質定数に等しい確率変数が従う分布を **Dirac分布** と呼ぶことがある。

1.3 独立同分布(i.i.d., iid)の定義

確率変数達 X_1, \dots, X_n が **独立同分布** (independent and identically distributed, **i.i.d.**, **iid**) であるとは、それらが独立でかつ X_i 達が従う分布が互いにすべて等しいことであると定める。

1.4 標本分布の定義

n 個の確率変数達 X_1, \dots, X_n は独立同分布であり、各 X_i は共通の確率分布 D に従うと仮定する:

$$X_1, \dots, X_n \sim D \quad (\text{independent})$$

このとき、 X_1, \dots, X_n を分布 D の **標本** または **サンプル** (sample) と呼び、 X_1, \dots, X_n の同時確率分布を **分布 D のサイズ n の標本分布** (distribution of samples of size n from distribution D) と呼ぶことにする。

注意: 独立同分布な確率変数達 X_1, \dots, X_n (確率変数は数ではなく関数であった) そのものではなく、それらの実現値 x_1, \dots, x_n (関数としての確率変数の値達) を標本もしくはサンプルと呼ぶことがある。1つひとつの値 x_i をサンプルと呼ぶのではなく、数の列 x_1, \dots, x_n をサンプルと呼ぶことに注意せよ。その辺が紛らわしい場合には数の列 x_1, \dots, x_n を **データ** (data) と呼ぶことがある。 X_1, \dots, X_n もデータと呼ぶことがある。この辺の用語の使い方はかなりイーカゲンになり易いので注意して欲しい。

確率変数 X_i 達が同一の確率密度関数 $p(x_i)$ を持つとき、 X_1, \dots, X_n の同時確率密度関数が

$$p(x_1, \dots, x_n) = p(x_1) \cdots p(x_n)$$

であることと、 X_1, \dots, X_n の同時確率分布が確率密度関数 $p(x)$ が定める連続分布の標本分布であることは同値である。

確率変数 X_i 達が同一の確率質量関数 $P(x_i)$ を持つとき、 X_1, \dots, X_n の同時確率質量関数が

$$P(x_1, \dots, x_n) = P(x_1) \cdots P(x_n)$$

であることと、 X_1, \dots, X_n の同時確率分布が確率質量関数 $P(x)$ が定める離散分布の標本分布であることは同値である。

1.5 確率分布の積と n 乗の定義

X_1, \dots, X_n が独立な確率変数達であり、 X_i が従う分布が D_i であるとき、 (X_1, \dots, X_n) が従う分布を

$$D_1 \times \cdots \times D_n$$

と書き、分布 D_1, \dots, D_n の **積** (product, 直積, direct product) と呼ぶ:

$$(X_1, \dots, X_n) \sim D_1 \times \cdots \times D_n \iff X_i \sim D_i \text{ for } i = 1, \dots, n \text{ and } X_1, \dots, X_n \text{ are independent.}$$

例えば、分布 D_i が確率密度関数 $p_i(x_i)$ を持つとき、積分布 $D_1 \times \cdots \times D_n$ は確率密度関数

$$p(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n)$$

によって定義される多変量連続分布になる。確率密度関数が積の形になるので、分布の積と呼ぶ。

D_i がすべて同一の分布 D であるとき、積 $D_1 \times \cdots \times D_n$ を D^n と書き、分布 D の **べき乗** (冪乗, power) もしくは **累乗** と呼ぶ。前節で説明したように分布 D の n 乗 D^n は分布 D のサイズ n の標本分布と呼ばれる:

$$(X_1, \dots, X_n) \sim D^n \iff (X_1, \dots, X_n) \text{ is a sample of size } n \text{ from the distribution } D.$$

例: 試行回数 n , 成功確率 p のBernoulli試行の分布は成功確率 p のBernoulli分布の n 乗になる。以下の節も参照せよ。

1.6 分布 D の標本分布の主な使用用途

分布 D のサイズ n の標本分布は、現実で得られる長さ n の数値列 x_1, \dots, x_n の形式のデータの生成法則のモデル化としてよく使われている。そのとき、分布 D やその標本分布 D^n は **統計モデル** と呼ばれる。

ただし、我々は、分布 D のサイズ n の標本分布 D^n を各々が分布 D に従う独立同分布な n 個の確率変数達 X_1, \dots, X_n の同時確率分布のことだと定義したので、データ x_1, \dots, x_n の生成法則が独立同分布という強い条件を近似的に満たしていなければ、そのようなモデル化は妥当ではなくなる。

例: データ x_1, \dots, x_n がコンピュータで `rand()` 関数を n 回繰り返して得られた乱数列とすると、そのデータの生成法則は一様分布 `Uniform(0, 1)` のサイズ n の標本分布に近似的に従っているとみなされる。

例(基本!): S市の中学3年生男子の総人数は N 人であるとし、その身長の数全体を a_1, a_2, \dots, a_N と書く。等確率 $1/N$ で a_i 達のどれかの値が選ばれるという条件で定義される有限離散分布を Q と書き、S市の中3男子の身長の分布と呼ぶことにする。(全数調査前なので分布 Q は未知であると仮定する。) $1, 2, \dots, N$ から n 個の数 i_1, \dots, i_n を無作為に選んで、 n 人分の中3男子の身長を測定してデータ $(x_1, \dots, x_n) = (a_{i_1}, \dots, a_{i_n})$ が得られたとする。このとき、データ x_1, \dots, x_n の生成法則はS市の中3男子の身長の分布 Q のサイズ n の標本分布になっていると考えられる。このとき、未知の分布 Q についての推定・推測・推論をデータ x_1, \dots, x_n を用いて行いたい。そのときに、 Q のモデル化として平均 μ と分散 σ^2 をパラメータに持つ正規分布を設定し、データ x_1, \dots, x_n の生成法則のモデル化として、正規分布のサイズ n の標本分布 $\text{Normal}(\mu, \sigma)^n$ を採用することが考えられる。この場合には正規分布モデルを統計モデルとして採用したことになる。パラメータ μ, σ はデータ x_1, \dots, x_n を使って推定されることになる。未知であるS市の中3男子の身長の分布 Q (よく **真の分布** と呼ばれる)が正規分布からかけ離れた形をしていると、正規分布モデルの採用は妥当ではなくなるが、真の分布が十分に正規分布に近いことが十分に確からしい場合には正規分布モデルの採用は適切になる。

例: 当たりが常に同じ確率で出続けると考えられるルーレットを n 回まわして、当たりなら 1 を記録し、外れなら 0 を記録することによって得られる $1, 0$ で構成された長さ n の数列 x_1, \dots, x_n の生成法則は、未知の成功確率 p を持つ試行回数 n のBernoulli試行の確率分布 $\text{Bernoulli}(p)^n$ (これはBernoulli分布の標本分布に等しい)に近似的に従うと考えられる。

1.7 試行回数 n のBernoulli試行の分布はBernoulli分布の標本分布

試行回数 n , 成功確率 p のBernoulli試行の確率質量関数は

$$P(x_1, \dots, x_n) = p^{x_1 + \cdots + x_n} (1 - p)^{n - (x_1 + \cdots + x_n)} = \prod_{i=1}^n (p^{x_i} (1 - p)^{1 - x_i}) \quad (x_i = 1, 0)$$

であった。これは成功確率 p のBernoulli分布の確率質量関数

$$P(x_i) = p^{x_i} (1 - p)^{1 - x_i} \quad (x_i = 1, 0)$$

の積になっているので、試行回数 n , 成功確率 p のBernoulli試行の確率分布は、成功確率 p のBernoulli分布のサイズ n の標本分布になっている。

未知の確率 p で当たりが出るルーレットを n 回まわしたときの、長さ n の当たりと外れからなる列をそのルーレットの出目のサイズ n の **標本 (サンプル, sample)** と呼ぶ。

その意味でのルーレットの出目のサンプルの確率的揺らぎは、Bernoulli分布の標本分布(すなわちBernoulli試行の分布)でモデル化される。

例えば、未知の確率 p で当たりが出るルーレットを $n = 1000$ 回まわしてサンプル(データ)を取得したら、1000 回中当たりが 308 回で外れが 692 回ならば、そのルーレットで当たりが出る確率は3割程度だろうと推定できる。実際にはルーレットを 1000 回まわし直すたびに当たりの回数は確率的に揺らぐので、推定結果も確率的に揺らぐことになる。そのような揺らぎを数学的にモデル化するために標本分布は使用される。

言葉使いに関する重要な注意: 「標本」「サンプル」は一般に複数の数値の集まりになる。上のルーレットの場合には当たりには 1 を対応させ、外れを 0 に対応させると、サンプルは 1 と 0 からなる長さ n の列になる。一つひとつの数値をサンプルと呼ぶのではなく、複数の数値の集まりをサンプルと呼ぶ。この専門用語的な言葉遣いは日常用語的なサンプルという言葉の使い方からはずれているので注意が必要である。この辺の言葉遣いで誤解を防ぎたい場合には「データ」と呼ぶこともある。

1.8 二項分布による推定の確率的揺らぎの記述

前節の設定を引き継ぐ。

Bernoulli分布のサイズ n の標本分布における 1 の個数の分布は二項分布になるのであった。仮に

n 回中 k 回の当たりが出たときに、未知である当たりが出る確率は k/n に近いだろうと推定

することにしたときの、 k/n の確率的な揺らぎは二項分布によって計算できる。

K は二項分布 $\text{Binomial}(n, p)$ に従う確率変数であるとし、確率変数 \hat{p} を $\hat{p} = K/n$ と定める。この確率変数 $\hat{p} = K/n$ は上のルールで定めた未知の確率 p の推定の仕方の数学的記述になっている。このとき、 \hat{p} はパラメータ p の **推定量** (estimator) であるという。確率変数 \hat{p} は「 n 回中 k 回当たりが出た」というデータに対応する確率変数 K の関数になっている。

$\hat{p} = K/n$ の期待値と分散は、二項分布の期待値と分散がそれぞれ np と $np(1-p)$ であることより、以下のように計算される：

$$E[\hat{p}] = E[K/n] = \frac{E[K]}{n} = \frac{np}{n} = p,$$
$$\text{var}(\hat{p}) = \text{var}(K/n) = \frac{\text{var}(K)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

例えば、 $n = 1000$, $p = 0.3$ のとき、 \hat{p} の標準偏差は

$$\text{std}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.3 \cdot 0.7}{1000}} \approx 1.45\%.$$

モデル内の設定では大雑把に推定結果はこの2倍の $\pm 3\%$ 程度揺らぐと考えられる。

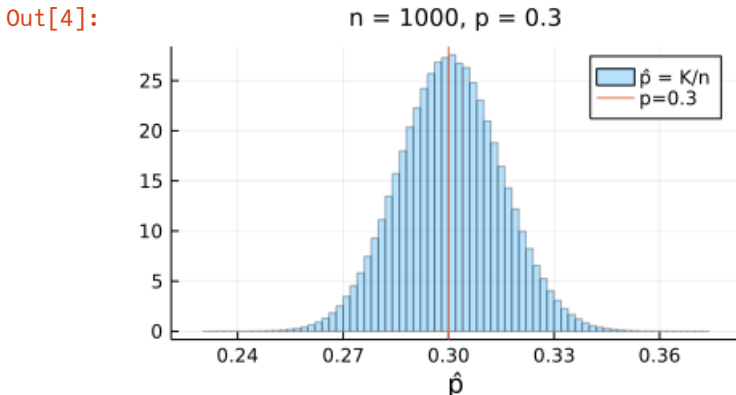
未知の p を使わずに $\text{SE} = \text{std}(\hat{p})$ の値を推定するためには、その式の中の p に p の推定量 \hat{p} を代入して得られる公式

$$\widehat{SE} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

を使えばよいだろう。(SE = $\text{std}(\hat{p})$) は **標準誤差** (standard error) と呼ばれる。 \widehat{SE} は標準誤差の推定量である。 \widehat{SE} のことも標準誤差と呼ぶことがある。この辺の言葉遣いもイカゲンな場合が多いので注意が必要である。何をどう呼ぶかよりも、それが正確には何を意味しているかが重要である。

このような統計分析の結果が現実において信頼できるかどうかは、Bernoulli分布の標本分布によるモデル化の現実における妥当性に依存する。モデルが現実において妥当である証拠が全然無ければ、このような推測結果も信頼できないことになる。モデルの現実における妥当性の証拠の提示は統計モデルのユーザー側が独自に行う必要がある。

```
In [4]: 1 n, p = 1000, 0.3
        2 dist = Binomial(n, p)
        3 L = 10^6
        4 K = rand(dist, L)
        5 p_hat = K/n
        6 histogram(p_hat; norm=true, alpha=0.3, label="p_hat = K/n", bin=100)
        7 vline!([p]; label="p=p", xlabel="p_hat", xtick=0.30-0.12:0.03:0.30+0.12)
        8 title!("n = $n, p = $p")
```



確かにランダムに決まる $\hat{p} = K/n$ の値は、推定先の値である $p = 0.30 = 30\%$ を中心に大雑把に $\pm 3\%$ 程度の範囲に分布している。

実際には小さな確率でもっと大きく外れることもあることに注意せよ。

現実の統計分析では、データを上のグラフの例のように100万回撮り直したりできないので、このようなグラフを描くことはできない。たった1つの「 n 回中 k 回成功」(もしくは「 n 人中 k 人」)の型のデータだけを使って判断を下さなければいけない。真っ暗闇の中を手探りで進むような感じになる。

1.9 問題: 大阪都構想に関する住民投票の結果について

2015年と2020年の大阪都構想に関する住民投票の結果は

- 2015年: 賛成: 694,844 (49.6%) 反対: 705,585 (50.4%)
- 2020年: 賛成: 675,829 (49.4%) 反対: 692,996 (50.6%)

であった([検索 \(https://www.google.com/search?q=%E5%A4%A7%E9%98%AA%E9%83%BD%E6%A7%8B%E6%83%B3+%E4%BD%8F%E6%B0%91%E6%8A%95%E7%A5%A](https://www.google.com/search?q=%E5%A4%A7%E9%98%AA%E9%83%BD%E6%A7%8B%E6%83%B3+%E4%BD%8F%E6%B0%91%E6%8A%95%E7%A5%A)

どちらでも僅差で反対派が勝利した。パーセントの数値を見ると大変な僅差であったようにも見える。この数値に二項分布モデルを適用したらどうなるかを(それが妥当な適用であるかどうかを度外視して)計算するのがこの問題の内容である。

確率変数 K は二項分布 $\text{Binomial}(n, p)$ に従うと仮定する。

(1) $n = 694844 + 705585 = 1400429$, $p = 0.5$ のとき, 確率 $P(K \leq 694844)$ の2倍の値を求めよ。

(2) $n = 675829 + 692996 = 1368825$, $p = 0.5$ のとき, 確率 $P(K \leq 675829)$ の2倍の値を求めよ。

(3) $n = 694844 + 705585 = 1400429$ のとき, 以下を求めよ:

- $P(K \geq 694844) = 2.5\%$ になるようなパラメータ p の値 p_L ,
- $P(K \leq 694844) = 2.5\%$ になるようなパラメータ p の値 p_U .

(4) $n = 675829 + 692996 = 1368825$ のとき, 以下を求めよ:

- $P(K \geq 675829) = 2.5\%$ になるようなパラメータ p の値 p_L ,
- $P(K \leq 675829) = 2.5\%$ になるようなパラメータ p の値 p_U .

確率やパラメータの数値は有効桁4桁まで求めよ。0.000000000000000000001234 のように 0 を沢山含む表示は見難いので,

$$1.234 \times 10^{20} = \underbrace{0.000000000000000000001234}_{20}$$

のように書かずに,

$$1.234\text{E}-20 = 1.234\text{e}-20$$

のように書くこと。

注意: この問題の内容を一般化するだけで **検定** (統計的仮説検定, statistical hypothesis testing)や **信頼区間** (confidence interval)の一般論が得られる。(1), (2)はP値と求める問題になっており, (3), (4)は 95% 信頼区間 $[p_L, p_U]$ を求める問題になっている(ただし両方Clopper-Pearsonの信頼区間の場合)。

解答例:

(1) $P(K \leq 694844) \approx 1.130\text{e}-19$

(2) $P(K \leq 675829) \approx 9.687\text{e}-49$

(3) $[p_L, p_U] \approx [0.4953, 0.4970]$

(4) $[p_L, p_U] \approx [0.4929, 0.4946]$

解答終

1.9.1 Julia言語による計算の例

```
In [5]: 1 # 確率計算を素朴に行うには対数を取った結果を主な対象にしないと失敗する。
        2 # 次は二項分布における確率質量関数の対数である。
        3 logP(n, p, k) = logabsbinomial(n, k)[1] + k*log(p) + (n-k)*log(1-p)
```

```
Out[5]: logP (generic function with 1 method)
```



```
In [6]: 1 # (1)
2 @show 2exp(logsumexp(logP(694844 + 705585, 0.5, k) for k in 0:694844))
3 @show 2cdf(Binomial(694844 + 705585, 0.5), 694844);

2 * exp(logsumexp((logP(694844 + 705585, 0.5, k) for k = 0:694844))) = 1.130422573100686e-19
2 * cdf(Binomial(694844 + 705585, 0.5), 694844) = 1.1304225734350622e-19
```

```
In [7]: 1 # (2)
2 @show 2exp(logsumexp(logP(675829 + 692996, 0.5, k) for k in 0:675829))
3 @show 2cdf(Binomial(675829 + 692996, 0.5), 675829);

2 * exp(logsumexp((logP(675829 + 692996, 0.5, k) for k = 0:675829))) = 9.687442124914083e-49
2 * cdf(Binomial(675829 + 692996, 0.5), 675829) = 9.687442131513807e-49
```

```
In [8]: 1 # (3)
2 f(t) = ccdf(Binomial(694844 + 705585, t), 694844-1) - 0.025
3 g(t) = cdf(Binomial(694844 + 705585, t), 694844) - 0.025
4 @show p_L = find_zero(f, (0, 1))
5 @show p_U = find_zero(g, (0, 1));

p_L = find_zero(f, (0, 1)) = 0.4953366705819138
p_U = find_zero(g, (0, 1)) = 0.4969935526649652
```

```
In [9]: 1 # (3)
2 n = 694844 + 705585
3 k = 694844
4 α = 0.05
5 @show p_L = quantile(Beta(k, n-k+1), α/2)
6 @show p_U = quantile(Beta(k+1, n-k), 1 - α/2);

p_L = quantile(Beta(k, (n - k) + 1), α / 2) = 0.4953366705819138
p_U = quantile(Beta(k + 1, n - k), 1 - α / 2) = 0.49699355266496525
```

```
In [10]: 1 # (3)
2 n = 694844 + 705585
3 k = 694844
4 α = 0.05
5 @show p_L = beta_inc_inv(k, n-k+1, α/2)[1]
6 @show p_U = beta_inc_inv(k+1, n-k, 1 - α/2)[1];

p_L = (beta_inc_inv(k, (n - k) + 1, α / 2))[1] = 0.4953366705819138
p_U = (beta_inc_inv(k + 1, n - k, 1 - α / 2))[1] = 0.49699355266496514
```

```
In [11]: 1 # (4)
2 f(t) = ccdf(Binomial(675829 + 692996, t), 675829-1) - 0.025
3 g(t) = cdf(Binomial(675829 + 692996, t), 675829) - 0.025
4 @show p_L = find_zero(f, (0, 1))
5 @show p_U = find_zero(g, (0, 1));

p_L = find_zero(f, (0, 1)) = 0.49289139358630474
p_U = find_zero(g, (0, 1)) = 0.4945672196912542
```

```
In [12]: 1 # (4)
2 n = 675829 + 692996
3 k = 675829
4 α = 0.05
5 @show p_L = quantile(Beta(k, n-k+1), α/2)
6 @show p_U = quantile(Beta(k+1, n-k), 1 - α/2);

p_L = quantile(Beta(k, (n - k) + 1), α / 2) = 0.49289139358630474
p_U = quantile(Beta(k + 1, n - k), 1 - α / 2) = 0.4945672196912543
```

```
In [13]: 1 # (4)
2 n = 675829 + 692996
3 k = 675829
4 α = 0.05
5 @show p_L = beta_inc_inv(k, n-k+1, α/2)[1]
6 @show p_U = beta_inc_inv(k+1, n-k, 1 - α/2)[1];

p_L = (beta_inc_inv(k, (n - k) + 1, α / 2))[1] = 0.49289139358630474
p_U = (beta_inc_inv(k + 1, n - k, 1 - α / 2))[1] = 0.4945672196912543
```

1.9.2 WolframAlphaによる計算の例:

(1) $2 \text{cdf}(\text{BinomialDistribution}(694844 + 705585, 0.5), 694844)$ (<https://www.wolframalpha.com/input?i=2+cdf%28BinomialDistribution%28694844+%2B+705585%2C+0.5%29%2C+694844%29>)

(2) $2 \text{cdf}(\text{BinomialDistribution}(675829 + 692996, 0.5), 675829)$ (<https://www.wolframalpha.com/input?i=2+cdf%28BinomialDistribution%28675829+%2B+692996%2C+0.5%29%2C+675829%29>)

(3) p_L : $\text{solve cdf}(\text{BinomialDistribution}(694844 + 705585, q), 705585) = 0.025$ (<https://www.wolframalpha.com/input?i=solve+cdf%28BinomialDistribution%28694844+%2B+705585%2C+q%29%2C+705585%29+%3D+0.025>) として、これを1から引いた値を求める: $\text{InverseBetaRegularized}(1/40, 694844, 705586)$ (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%281%2F40%2C+694844%2C+705586%29>). もしくは、 $\text{InverseBetaRegularized}(0.025, 694844, 705585 + 1)$ (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%281%2F40%2C+694844%2C+705586%29>).

p_U : $\text{solve cdf}(\text{BinomialDistribution}(694844 + 705585, p), 694844) = 0.025$ (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%280.025%2C+694844%2C+705585+%2B+1%29>). もしくは、 $\text{InverseBetaRegularized}(1 - 0.025, 694844 + 1, 705585)$ (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%281+-+0.025%2C+694844+%2B+1%2C+705585%29&lang=ja>).

(4) p_L : $\text{solve cdf}(\text{BinomialDistribution}(675829 + 692996, q), 692996) = 0.025$ (<https://www.wolframalpha.com/input?i=solve+cdf%28BinomialDistribution%28675829+%2B+692996%2C+q%29%2C+692996%29+%3D+0.025>) として、これを1から引いた値を求める: $\text{InverseBetaRegularized}(1/40, 675829, 692997)$ (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%281%2F40%2C+675829%2C+692997%29>). もしくは、 $\text{InverseBetaRegularized}(0.025, 675829, 692996 + 1)$ (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%280.025%2C+675829%2C+692996+%2B+1%29>).

p_U : $\text{solve cdf}(\text{BinomialDistribution}(675829 + 692996, p), 675829) = 0.025$ (<https://www.wolframalpha.com/input?i=solve+cdf%28BinomialDistribution%28675829+%2B+692996%2C+p%29%2C+675829%29+%3D+0.025>). もしくは、 $\text{InverseBetaRegularized}(1 - 0.025, 675829 + 1, 692996)$ (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%281+-+0.025%2C+675829+%2B+1%2C+692996%29>).

ここで、(3)、(4)の p_L の計算では $\text{Binomial}(n, p)$ で k 以上の確率は $\text{Binomial}(n, 1 - p)$ で $n - k$ 以下になる確率に等しいことを使った。

1.9.3 Clopper-Pearsonの信頼区間とそれを与えるP値

(3)と(4)で計算した値から得られる区間 $[p_L, p_U]$ は **母比率に関する信頼度95%のClopper-Pearsonの信頼区間** として統計学ユーザーのあいだでよく知られている。

(1)と(2)での2倍する前の確率は **片側検定のP値** になっている。2倍後の値は **両側検定のP値** (通常はこちらを使う)の一種になっており、Clopper-Pearsonの信頼区間を与える。

P値は採用した統計モデルとデータの整合性の指標である(P値が小さければ整合性が低い)。(1)と(2)で求めたP値の値は極めて小さいということは、成功確率 $p = 0.5$ の二項分布モデルと2015年と2020年の大阪都構想に関する住民投票の結果の整合性が極めて低いということを意味している。2015年と2020年の大阪都構想に関する住民投票の結果については、成功確率 $p = 0.5$ の二項分布モデルは捨てる必要がある。

(3)と(4)で求めた区間 $[p_L, p_U]$ はデータから計算されるP値が 5% 以上になるパラメータ p の値全体の集合になっている。すなわち、P値に関する 5% の閾値(**有意水準** と呼ばれる)で整合性が低すぎるという理由で捨て去られずにすむ p の値全体が信頼度 $1 - 5\% = 95\%$ の信頼区間になっている。このパターンは一般の場合にもそのまま通用する。

Clopper-Pearsonの信頼区間の効率的計算には、二項分布の累積分布関数はベータ分布の累積分布関数で書けることが使われる:

$$1 - \text{cdf}(\text{Binomial}(n, p), k - 1) = \text{cdf}(\text{Beta}(k, n - k + 1), p), \\ \text{cdf}(\text{Binomial}(n, p), k) = 1 - \text{cdf}(\text{Beta}(k + 1, n - k), p).$$

これらの公式から、 n, k が与えられていて $K \sim \text{Binomial}(n, p)$ のとき、 $P(K \geq k) = \alpha/2$ の解 p_L と $P(K \leq k) = \alpha/2$ の解 p_U はそれぞれ次のように書ける:

$$p_L = \text{quantile}(\text{Beta}(k, n - k + 1), \alpha/2), \quad p_U = \text{quantile}(\text{Beta}(k + 1, n - k), 1 - \alpha/2).$$

ベータ分布の累積分布関数が **正則化された不完全ベータ関数** (regularized incomplete Beta function)になっている:

$$P(T \leq \theta) = I_\theta(\alpha, \beta) = \frac{\int_0^\theta t^{\alpha-1} (1-t)^{\beta-1} dt}{B(\alpha, \beta)} \quad \text{if } T \sim \text{Beta}(\alpha, \beta)$$

このことから、quantile 関数の代わりに $\theta \mapsto p = I_\theta(\alpha, \beta)$ の逆関数を直接使ってもよい:

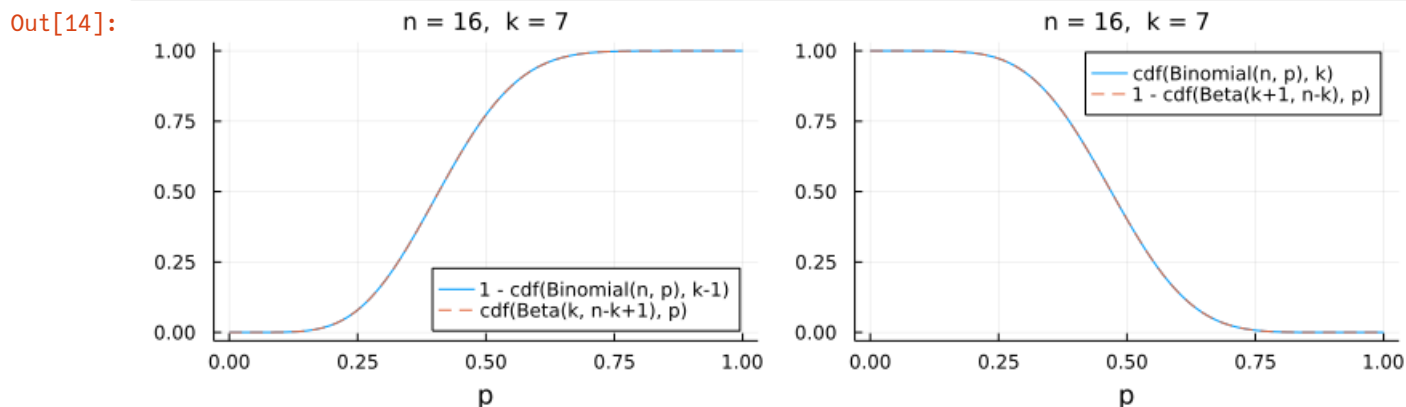
$$p_L = \text{beta_inc_inv}(k, n - k + 1, \alpha/2)[1], \quad p_U = \text{beta_inc_inv}(k + 1, n - k, 1 - \alpha/2)[1].$$

Julia言語の `SpecialFunctions.jl` では正則化された不完全ベータ関数とその逆関数はそれぞれ using `SpecialFunctions` した後に `p = beta_inc(α, β, θ)[1]` と `θ = beta_inc_inv(α, β, p)` で使える。

Wolfram言語では `BetaRegularized` (<https://reference.wolfram.com/language/ref/BetaRegularized.html>) と `InverseBetaRegularized` (<https://reference.wolfram.com/language/ref/InverseBetaRegularized.html>) を使う。

Clopper-Pearsonの信頼区間を使うことのメリットは、二項分布の累積分布関数を和で計算してからパラメータに関する方程式を解くという面倒な手続きを経由せずに、基本特殊関数の1つである正則化された不完全ベータ関数の逆関数に帰着して効率的に計算できることである。別の信頼区間の定義の仕方との比較でデメリットもある。Sternの信頼区間([Stern \(1954\)](https://www.jstor.org/stable/2333026) (<https://www.jstor.org/stable/2333026>))との相対比較では、Clopper-Pearsonの信頼区間の方が無駄に広がってしまう場合が多い。 np が大きな場合にはどちらを使っても実践的に意味のある差は出ない。

```
In [14]: 1 n, k = 16, 7
2
3 P1 = plot(p → 1 - cdf(Binomial(n, p), k-1), 0, 1;
4         label="1 - cdf(Binomial(n, p), k-1)", legend=:bottomright)
5 plot!(p → cdf(Beta(k, n-k+1), p), 0, 1;
6         label="cdf(Beta(k, n-k+1), p)", ls=:dash)
7 title!("n = $n, k = $k"; xlabel="p")
8
9 P2 = plot(p → cdf(Binomial(n, p), k), 0, 1;
10         label="cdf(Binomial(n, p), k)", legend=:topright)
11 plot!(p → 1 - cdf(Beta(k+1, n-k), p), 0, 1;
12         label="1 - cdf(Beta(k+1, n-k), p)", ls=:dash)
13 title!("n = $n, k = $k"; xlabel="p")
14
15 plot(P1, P2; size = (800, 250), bottommargin=4Plots.mm)
```



上のグラフでは $1 - \text{cdf}(\text{Binomial}(n, p), k - 1)$ と $\text{cdf}(\text{Beta}(k, n - k + 1), p)$ がぴったり一致し、 $\text{cdf}(\text{Binomial}(n, p), k)$ と $1 - \text{cdf}(\text{Beta}(k + 1, n - k), p)$ がぴったり一致し、区別がつかなくなっている。ここで $\text{cdf}(D, x)$ は分布 D の累積分布関数を表す:

$$X \sim D \implies \text{cdf}(D, x) = P(X \leq x).$$

1.9.4 信頼区間よりも情報量が必要なP値関数のプロット

二項分布モデルにおいて n, k が与えられたときに、パラメータ p に対してP値を対応させる関数を **P値関数** (p-value function) と呼ぶ。P値関数の値が有意水準 α 以上の p 全体の集合が信頼度 $1 - \alpha$ の信頼区間になる。この意味でP値関数はすべての信頼度に関する信頼区間の情報をすべて持っており、適当な条件の下ではすべての信頼度に関する信頼区間が与えられていればそこからP値関数を逆に作れる。この意味でP値関数と信頼区間達は表裏一体の関係になっている。

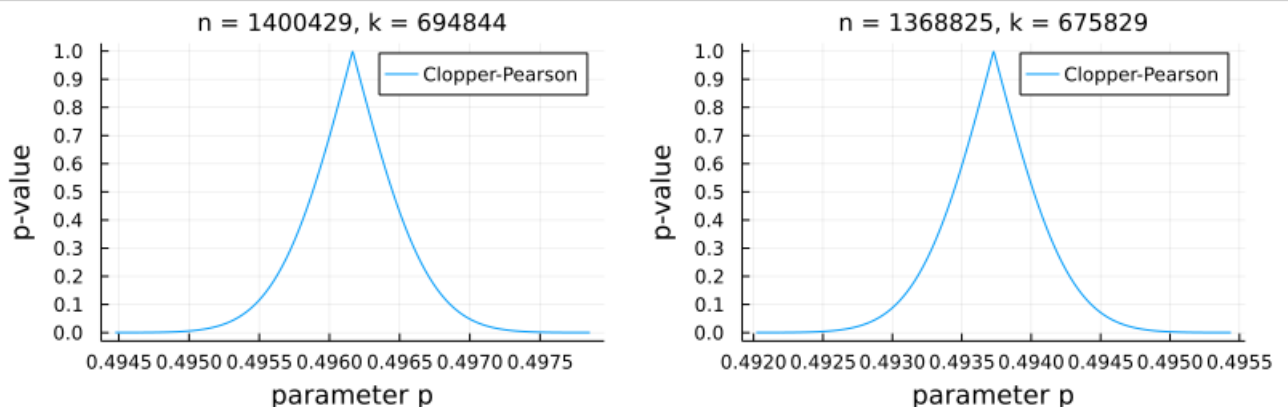
以下では上の問題の場合についてのP値関数をプロットしてみよう。

```

In [15]: 1 function plot_pvalue_function!(pvalue_func, n, k; label="", kwargs...)
2      $\hat{p} = k/n$ 
3      $SE = \sqrt{\hat{p}(1 - \hat{p})/n}$ 
4     ps = range( $\hat{p} - 4SE$ ,  $\hat{p} + 4SE$ , 1000)
5     plot!(ps, p  $\rightarrow$  pvalue_func(n, k, p); label, kwargs...)
6 end
7
8 function plot_pvalue_function(pvalue_func, n, k; label="", kwargs...)
9     plot()
10    plot_pvalue_function!(pvalue_func, n, k; label)
11    title!("n = $n, k = $k")
12    plot!(); ytick=0:0.1:1, xlabel="parameter p", ylabel="p-value")
13    plot!(); kwargs...
14 end
15
16 function pvalue_clopper_pearson(dist::DiscreteUnivariateDistribution, x)
17     min(1, 2cdf(dist, x), 2ccdf(dist, x-1))
18 end
19 pvalue_clopper_pearson(n, k, p) = pvalue_clopper_pearson(Binomial(n, p), k)
20
21 # (3)
22 P1 = plot_pvalue_function(pvalue_clopper_pearson, 694844 + 705585, 694844;
23     label="Clopper-Pearson", xtick=0:0.0005:1)
24
25 # (4)
26 P2 = plot_pvalue_function(pvalue_clopper_pearson, 675829 + 692996, 675829;
27     label="Clopper-Pearson", xtick=0:0.0005:1)
28
29 plot(P1, P2; size=(800, 250),
30     leftmargin=4Plots.mm, bottommargin=4Plots.mm, rightmargin=4Plots.mm)

```

Out[15]:



データ x を与えられているとき(上の場合には「 n 人中 k 人が賛成」というデータを与えられているとき), 統計モデルのパラメータ θ に(上の場合には p に)対してP値を対応させる関数をP値関数と呼ぶ. 上のグラフは大阪都構想に関する住民投票の結果データとし, 統計モデルを二項分布にしたときのP値関数のプロットである.

P値が高いパラメータ値ほど与えられたデータに統計モデルがよくフィット(fit, 適合)している. P値が低い場合にはその逆になる. あまりにも低いP値を持つパラメータ値については, 統計モデルとそのパラメータ値の組が与えられたデータに全然適合していないことになる.

そのとき重要なことは, P値が低いパラメータ値について, そのパラメータ値のみについてデータに適合していないと考えるのではなく, 統計モデルとそのパラメータ値の組がデータにフィットしていないと考えなければいけないことである.

例えば, 上のグラフを見れば, グラフに描いた部分の横軸の目盛りをはみ出しているパラメータ値 $p = 0.5$ のP値はおそらく低い. 二項分布モデル内で $p = 0.5$ は賛成派の比率がちょうど 50% であることを意味している. $p = 0.5$ のP値がおそらく低いことを理由に「賛成派の比率が 50% であるという仮説は否定された」のように単純に考えてはいけな. そうではなく, 「賛成派の比率が 50% であることを意味するパラメータ値の二項分布モデルの現実における妥当性は疑わしい」と考えなければいけない. 二項分布モデルも現実における妥当性を疑う対象に入れる必要がある.

この手のことが大学学部生向けの教科書に書かれていないせいで, 大学で統計学の講義を受講した人達の多くが統計学の最も基本的な部分をひどく誤解したまま単位を取得しているように思われる.

誤解を修正するためには次の講義動画がお勧めである:

- 京都大学大学院医学研究科 聴講コース 臨床研究者のための生物統計学「仮説検定とP値の誤解」佐藤 俊哉 医学研究科教授.
[YouTube (<https://youtu.be/vz9cZnB1d1c>)]

その講義で扱われているP値に関するASA声明の翻訳とその翻訳の経緯の解説を以下の場所で読める:

- 統計的有意性とP値に関するASA声明 [[pdf \(https://www.biometrics.gr.jp/news/all/ASA.pdf\)](https://www.biometrics.gr.jp/news/all/ASA.pdf)]
- 佐藤俊哉, ASA声明と疫学研究におけるP値, 計量生物学, 2018年38巻2号, pp. 109-115. [[link \(https://www.jstage.jst.go.jp/article/jjb/38/2/38_109/_article/-char/ja/\)](https://www.jstage.jst.go.jp/article/jjb/38/2/38_109/_article/-char/ja/)]

1.9.5 Sternの信頼区間とそれを与えるP値関数

Sternの信頼区間を与えるP値関数の定義は, 二項分布の確率質量関数を

$$P(k|n, p) = \binom{n}{k} p^k (1-p)^{n-k} \quad (k = 0, 1, \dots, n)$$

と書くとき,

$$\text{pvalue}_{\text{Stern}}(k|n, p) = \sum_{j \text{ with } P(j|n, p) \leq P(k|n, p)} P(j|n, p)$$

と $P(k|n, p)$ 以下となるような $P(j|n, p)$ 達の和として定義される. すなわち, 二項分布 $\text{Binomial}(n, p)$ においてその値が生じる確率がデータの数値 k が生じる確率以下になる確率としてSternの信頼区間を与えるP値は定義される.

そして, 与えられた n, k について, Sternの信頼区間はこのP値が α 以上になるパラメータ p の範囲として定義される. (実はその定義だと区間になるとは限らない場合が稀にあるので, その場合には適当に定義を訂正することになる.)

1.9.6 Sternの信頼区間を与えるP値関数の実装例

以下はSternの信頼区間を与えるP値関数の実装例である.

Clopper-Pearsonの信頼区間を与えるP値関数の実装(実質1行!)と比較すると相当に複雑になっている.

そして, 実装の仕方によって計算効率に大きな違いが生じていることにも注目せよ.

In [16]:

```
1 x ≲ y = x < y || x ≈ y
2
3 # Naive implementation is terribly slow.
4 function pvalue_stern_naive(dist::DiscreteUnivariateDistribution, x; xmax = 10^6)
5     Px = pdf(dist, x)
6     Px == 0 && return Px
7     ymin, maxdist = minimum(dist), maximum(dist)
8     ymax = maxdist == Inf ? xmax : maxdist
9     sum(pdf(dist, y) for y in ymin:ymax if 0 < pdf(dist, y) ≲ Px; init = 0.0)
10 end
11 pvalue_stern_naive(n, k, p) = pvalue_stern_naive(Binomial(n, p), k)
12
13 # Second implementation is very slow.
14 function pvalue_stern_old(dist::DiscreteUnivariateDistribution, x)
15     Px = pdf(dist, x)
16     Px == 0 && return Px
17     distmin, distmax = extrema(dist)
18     m = mode(dist)
19     Px ≈ pdf(dist, m) && return one(Px)
20     if x < m
21         y = m + 1
22         while !(pdf(dist, y) ≲ Px)
23             y += 1
24         end
25         cdf(dist, x) + ccdf(dist, y-1)
26     else # k > m
27         y = m - 1
28         while !(pdf(dist, y) ≲ Px)
29             y -= 1
30         end
31         cdf(dist, y) + ccdf(dist, x-1)
32     end
33 end
34 pvalue_stern_old(n, k, p) = pvalue_stern_old(Binomial(n, p), k)
35
36 ### The following implementation efficient.
37
38 _pdf_le(x, (dist, y)) = pdf(dist, x) ≲ y
39
40 function _search_boundary(f, x0, Δx, param)
41     x = x0
42     if f(x, param)
43         while f(x - Δx, param) x -= Δx end
44     else
45         x += Δx
46         while !f(x, param) x += Δx end
47     end
48     x
49 end
50
51 function pvalue_stern(dist::DiscreteUnivariateDistribution, x)
52     Px = pdf(dist, x)
53     Px == 0 && return Px
54     m = mode(dist)
55     Px ≈ pdf(dist, m) && return one(Px)
56     if x < m
57         y = _search_boundary(_pdf_le, 2m - x, 1, (dist, Px))
58         cdf(dist, x) + ccdf(dist, y-1)
59     else # x > m
60         y = _search_boundary(_pdf_le, 2m - x, -1, (dist, Px))
61         cdf(dist, y) + ccdf(dist, x-1)
62     end
63 end
64 pvalue_stern(n, k, p) = pvalue_stern(Binomial(n, p), k)
```

Out[16]: pvalue_stern (generic function with 2 methods)

```
In [17]: 1 n = 10
2 k = -1:11
3 p = 0.4
4 a = @time pvalue_stern_naive.(n, k, p)
5 b = @time pvalue_stern_old.(n, k, p)
6 c = @time pvalue_stern.(n, k, p)
7 d = @time pvalue_clopper_pearson.(n, k, p)
8 @show a ≈ b ≈ c
9 [a b c d]
```

```
0.190067 seconds (440.66 k allocations: 24.315 MiB, 99.55% compilation time)
0.127227 seconds (211.57 k allocations: 11.444 MiB, 99.35% compilation time)
0.128147 seconds (190.21 k allocations: 10.235 MiB, 99.38% compilation time)
0.061974 seconds (125.54 k allocations: 6.829 MiB, 98.75% compilation time)
a ≈ b ≈ c = true
```

```
Out[17]: 13x4 Matrix{Float64}:
0.0      0.0      0.0      0.0
0.00772434 0.00772434 0.00772434 0.0120932
0.058652  0.058652  0.058652  0.0927148
0.333528  0.333528  0.333528  0.33458
0.749177  0.749177  0.749177  0.764561
1.0      1.0      1.0      1.0
0.534186  0.534186  0.534186  0.733793
0.212596  0.212596  0.212596  0.332477
0.101119  0.101119  0.101119  0.109524
0.0183412 0.0183412  0.0183412  0.0245891
0.00167772 0.00167772 0.00167772 0.00335544
0.000104858 0.000104858 0.000104858 0.000209715
0.0      0.0      0.0      0.0
```

```
In [18]: 1 # (3)の場合に
2 # pvalue_stern_naive は pvalue_stern_old よりも数百倍遅く,
3 # pvalue_stern_old は pvalue_stern よりも数百倍遅く,
4 # pvalue_stern は pvalue_clopper_pearson よりも少し遅い.
5 n = 694844 + 705585
6 k = 694844
7 a = @btime pvalue_stern_naive($n, $k, 0.5)
8 b = @btime pvalue_stern_old($n, $k, 0.5)
9 c = @btime pvalue_stern($n, $k, 0.5)
10 d = @btime pvalue_clopper_pearson($n, $k, 0.5)
11 @show a ≈ b ≈ c ≈ d
12 a, b, c, d
```

```
196.339 ms (0 allocations: 0 bytes)
571.100 μs (0 allocations: 0 bytes)
1.580 μs (0 allocations: 0 bytes)
1.110 μs (0 allocations: 0 bytes)
a ≈ b ≈ c ≈ d = true
```

```
Out[18]: (1.1304225731963725e-19, 1.1304225734350622e-19, 1.1304225734350622e-19, 1.1304225734350622e-19)
```

```
In [19]: 1 # 極端な場合
2 n = 694844 + 705585
3 k = 600000
4 b = @btime pvalue_stern_old($n, $k, 0.5)
5 c = @btime pvalue_stern($n, $k, 0.5)
6 d = @btime pvalue_clopper_pearson($n, $k, 0.5)
7 b, c, d
```

```
105.043 ns (0 allocations: 0 bytes)
104.767 ns (0 allocations: 0 bytes)
325.225 ns (0 allocations: 0 bytes)
```

```
Out[19]: (0.0, 0.0, 0.0)
```



```
In [20]: 1 # この場合には pvalue_stern_naive はさらに遅い。
2 n = 100000
3 k = 49500:50500
4 a = @time pvalue_stern_naive.(n, k, 0.5)
5 b = @time pvalue_stern_old.(n, k, 0.5)
6 c = @time pvalue_stern.(n, k, 0.5)
7 d = @time pvalue_clopper_pearson.(n, k, 0.5)
8 @show a ≈ b ≈ c ≈ d;
```

```
11.727193 seconds (3 allocations: 8.094 KiB)
0.027948 seconds (3 allocations: 8.094 KiB)
0.001521 seconds (3 allocations: 8.094 KiB)
0.001150 seconds (3 allocations: 8.094 KiB)
a ≈ b ≈ c ≈ d = true
```

```
In [21]: 1 # 以上の実装は超幾何分布でも使える。
2 dist = Hypergeometric(9, 9, 9)
3 k = -1:10
4 a = @time pvalue_stern_naive.(dist, k)
5 b = @time pvalue_stern_old.(dist, k)
6 c = @time pvalue_stern.(dist, k)
7 d = @time pvalue_clopper_pearson.(dist, k)
8 @show a ≈ b ≈ c ≈ d
9 [a b c d]
```

```
0.098966 seconds (360.39 k allocations: 19.759 MiB, 13.03% gc time, 99.52% compilation time)
0.082584 seconds (166.45 k allocations: 8.989 MiB, 61.65% compilation time)
0.064714 seconds (150.20 k allocations: 8.065 MiB, 99.28% compilation time)
0.038726 seconds (171.76 k allocations: 8.866 MiB, 99.14% compilation time)
a ≈ b ≈ c ≈ d = true
```

```
Out[21]: 12×4 Matrix{Float64}:
0.0      0.0      0.0      0.0
4.11353e-5 4.11353e-5 4.11353e-5 4.11353e-5
0.0033731 0.0033731 0.0033731 0.0033731
0.0566845 0.0566845 0.0566845 0.0566845
0.346935  0.346935  0.346935  0.346935
1.0        1.0        1.0        1.0
1.0        1.0        1.0        1.0
0.346935  0.346935  0.346935  0.346935
0.0566845 0.0566845 0.0566845 0.0566845
0.0033731 0.0033731 0.0033731 0.0033731
4.11353e-5 4.11353e-5 4.11353e-5 4.11353e-5
0.0        0.0        0.0        0.0
```

```
In [22]: 1 # 以上の実装は無限離散分布のPoisson分布でも使える。
2 dist = Poisson(4)
3 k = -1:10
4 a = @time pvalue_stern_naive.(dist, k)
5 b = @time pvalue_stern_old.(dist, k)
6 c = @time pvalue_stern.(dist, k)
7 d = @time pvalue_clopper_pearson.(dist, k)
8 @show a ≈ b ≈ c
9 [a b c d]
```

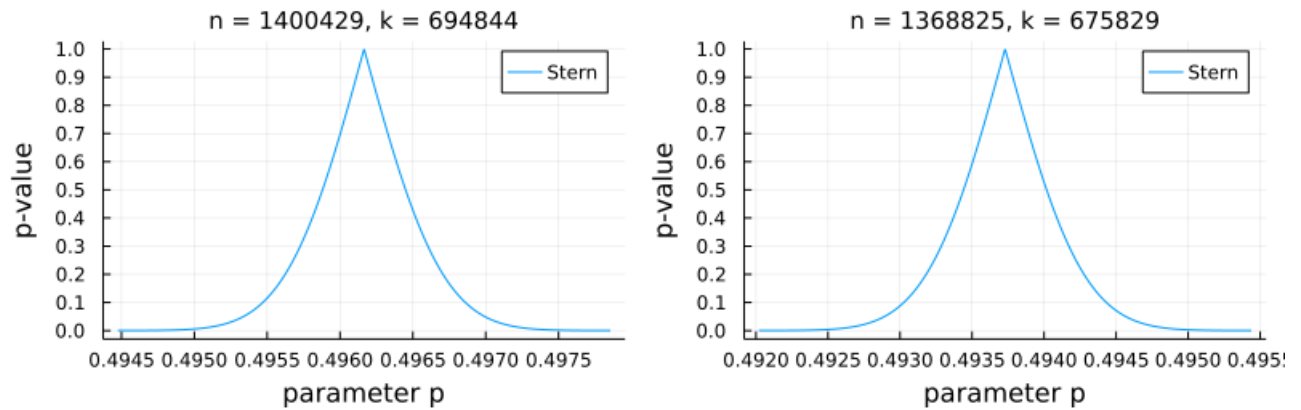
```
0.572695 seconds (433.78 k allocations: 24.200 MiB, 1.98% gc time, 19.17% compilation time)
0.045804 seconds (158.26 k allocations: 8.643 MiB, 98.82% compilation time)
0.070978 seconds (145.64 k allocations: 7.889 MiB, 71.23% compilation time)
0.039481 seconds (165.00 k allocations: 8.616 MiB, 99.15% compilation time)
a ≈ b ≈ c = true
```

```
Out[22]: 12×4 Matrix{Float64}:
0.0      0.0      0.0      0.0
0.0396791 0.0396791 0.0396791 0.0366313
0.202252 0.202252 0.202252 0.183156
0.452973 0.452973 0.452973 0.476207
1.0        1.0        1.0        0.86694
1.0        1.0        1.0        1.0
0.609266 0.609266 0.609266 0.742326
0.306448 0.306448 0.306448 0.429739
0.12899  0.12899  0.12899  0.221348
0.0694493 0.0694493 0.0694493 0.102267
0.0213634 0.0213634 0.0213634 0.0427269
0.00813224 0.00813224 0.00813224 0.0162645
```

In [23]:

```
1 # (3)
2 P1 = plot_pvalue_function(pvalue_stern, 694844 + 705585, 694844;
3   label="Stern", xtick=0:0.0005:1)
4
5 # (4)
6 P2 = plot_pvalue_function(pvalue_stern, 675829 + 692996, 675829;
7   label="Stern", xtick=0:0.0005:1)
8
9 plot(P1, P2; size=(800, 250), leftmargin=4Plots.mm, bottommargin=4Plots.mm)
```

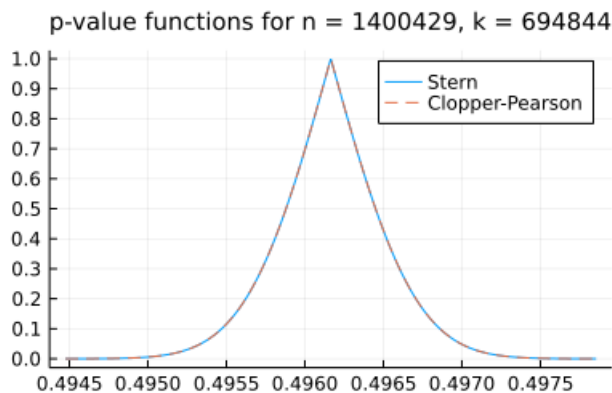
Out[23]:



In [24]:

```
1 # Clopper-Pearsonの信頼区間を与えるP値函数とSternの信頼区間を与えるP値函数の比較
2 # (3)の場合にはほぼびったり一致している。
3 n, k = 694844 + 705585, 694844
4 plot(title="p-value functions for n = $n, k = $k", ytick=0:0.1:1)
5 plot_pvalue_function!(pvalue_stern, n, k; label="Stern")
6 plot_pvalue_function!(pvalue_clopper_pearson, n, k; label="Clopper-Pearson", ls=:dash)
```

Out[24]:

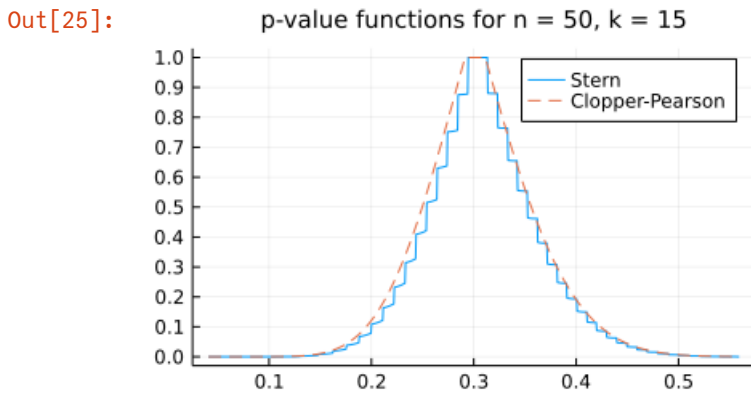


このように数値的にぴったり一致する場合にはClopper-Pearsonの信頼区間を与えるP値函数とSternの信頼区間のどちらを使うか悩む必要はないだろう。どちらを使っても(ほぼ)同じ結果が得られる。

```

In [25]: 1 # Clopper-Pearsonの信頼区間を与えるP値関数とSternの信頼区間を与えるP値関数の比較
2 # n = 50, k = 15 の場合には違いが見える。
3 n, k = 50, 15
4 plot(title="p-value functions for n = $n, k = $k", ytick=0:0.1:1)
5 plot_pvalue_function!(pvalue_stern, n, k; label="Stern")
6 plot_pvalue_function!(pvalue_clopper_pearson, n, k; label="Clopper-Pearson", ls=:dash)

```



Sternの信頼区間を与えるP値関数の値はClopper-Pearsonの信頼区間を与えるP値関数の値よりも小さくなりことが多い。(常にそうなるわけではない。) その結果, 対応する信頼区間もSternの側が狭くなってくることが多い。

1.10 対ごとに独立であっても全体が独立であるとは限らない

確率変数 X, Y, Z の互いに異なる任意の2つが独立であっても, X, Y, Z の全体が独立でない場合があることを具体的な例を作ることによって証明しよう。

確率変数 X, Y, Z の同時確率質量関数が $P(x, y, z)$ であるとき, X 単独の確率質量関数は $P(x) = \sum_{y,z} P(x, y, z)$ になり, (X, Y) の同時確率質量関数は $P(x, y) = \sum_z P(x, y, z)$ になる。他の場合も同様である。

注意: 変数名で異なる関数を区別するスタイルを採用したので記号法の運用時に注意すること。このスタイルでは $P(x, y)$ と $P(x, z)$ は異なる関数になる。値を代入する場合には $P(x = 1, y = 1)$ や $P(x = 1, z = 1)$ のように書いて区別できるようにする。

$P(x, y, z)$ で $P(x, y) = P(x)P(y)$, $P(x, z) = P(x)P(z)$, $P(y, z) = P(y)P(z)$ を満たすが $P(x, y, z) \neq P(x)P(y)P(z)$ となるものを具体的に構成すればよい。

例えば以下のような確率質量関数 $P(x, y, z)$ の例を作ることができる:

	$z = 1$		$z = 0$	
	$y = 1$	$y = 0$	$y = 1$	$y = 0$
$x = 1$	$P(1, 1, 1) = 1/4$	$P(1, 0, 1) = 0$	$P(1, 1, 0) = 0$	$P(1, 0, 0) = 1/4$
$x = 0$	$P(0, 1, 1) = 0$	$P(0, 0, 1) = 1/4$	$P(0, 1, 0) = 1/4$	$P(0, 0, 0) = 0$

x, y, z はそれぞれ 1, 0 を動くとする。このとき, $P(x, y)$, $P(x, z)$, $P(y, z)$ の値はすべて $1/4$ になり, $P(x)$, $P(y)$, $P(z)$ の値はすべて $1/2$ になることがわかる。たとえば, $P(x = 1, y = 1) = P(1, 1, 1) + P(1, 1, 0) = 1/4 + 0 = 1/4$ であり, $P(x = 1) = \sum_{y,z} P(1, y, z) = 1/4 + 0 + 0 + 1/4 = 1/2$ 。

だから, $P(x, y) = P(x)P(y)$, $P(x, z) = P(x)P(z)$, $P(y, z) = P(y)P(z)$ が成立するが, $P(x, y, z) \neq 1/8 = P(x)P(y)P(z)$ となっている。

このとき, X, Y, Z が同時確率質量関数 $P(x, y, z)$ を持つとすると, そのうちの任意の異なる2つは独立だが, X, Y, Z の全体は独立ではない。

1.11 確率変数の独立性の現実における解釈に関する重大な注意

上の例は具体的には次のような状況だと解釈可能である。

(1) $P(x) = 1/2$ の解釈: $X = 1$ は薬Aを与えたことを, $X = 0$ は薬Bを与えたことを意味する。全員に確率 $1/2$ で薬Aまたは薬Bを与えた。

$x = 1$ = 薬A	$P(x = 1) = 1/2$
$x = 0$ = 薬B	$P(x = 0) = 1/2$

(2) $P(y) = 1/2$ の解釈: $Y = 1$ は薬に効果があったことを, $Y = 0$ は薬に効果がなかったことを意味する。全体で見たら, $1/2$ の確率で薬には効果があった。

$y = 1 = \text{効果有}$	$y = 0 = \text{効果無}$
$P(y = 1) = 1/2$	$P(y = 0) = 1/2$

(3) $P(z) = 1/2$ の解釈: $Z = 1$ は女性であることを、 $Z = 0$ は男性であることを意味する。女性である確率と男性である確率は $1/2$ だった。

$z = 1 = \text{女性}$	$z = 0 = \text{男性}$
$P(z = 1) = 1/2$	$P(z = 0) = 1/2$

(4) 男女の区別をやめると、薬Aも薬Bも効果がある確率は半々であり、薬Aと薬Bのどちらを与えたかと効果があつたかどうかは独立である。男女を合わせた($z = 1, 0$ の場合の和を取って得られる)確率質量関数 $P(x, y)$ の表

	$y = 1 = \text{効果有}$	$y = 0 = \text{効果無}$
$x = 1 = \text{薬A}$	$P(x = 1, y = 1) = 1/4$	$P(x = 1, y = 0) = 1/4$
$x = 0 = \text{薬B}$	$P(x = 0, y = 1) = 1/4$	$P(x = 0, y = 0) = 1/4$

は $x = 1$ の薬Aの場合も $x = 0$ の薬Bの場合も男女の区別をやめると薬Aと薬Bで効果に変わりがないことを意味している。

(5) 薬Aと薬Bのどちらを与えたかと男女のどちらであるかは独立である。そのことは効果の有無を意味する $y = 1, 0$ について和を取って得られる確率質量関数 $P(x, z)$ の表

	$z = 1 = \text{女性}$	$z = 0 = \text{男性}$
$x = 1 = \text{薬A}$	$P(x = 1, z = 1) = 1/4$	$P(x = 1, z = 0) = 1/4$
$x = 0 = \text{薬B}$	$P(x = 0, z = 1) = 1/4$	$P(x = 0, z = 0) = 1/4$

からわかる。

(6) 薬Aと薬Bのどちらを与えたかを無視すると、男女のどちらであるかと薬の効果の有無は独立である。そのことは薬Aと薬Bのどちらを与えたかを意味する $x = 1, 0$ について和を取って得られる確率質量関数 $P(y, z)$ の表

	$y = 1 = \text{効果有}$	$y = 0 = \text{効果無}$
$z = 1 = \text{女性}$	$P(y = 1, z = 1) = 1/4$	$P(y = 0, z = 1) = 1/4$
$z = 0 = \text{男性}$	$P(y = 1, z = 0) = 1/4$	$P(y = 0, z = 0) = 1/4$

からわかる。

(7) しかし、男女を区別すると全然違う結果が見えて来る。薬Aは女性だけに効果があり、薬Bは男性だけに効果がある。 $z = 1$ の女性の場合に制限した確率質量関数 $P(x, y, z)$ の表

	$z = 1 = \text{女性}$
	$y = 1 = \text{効果有}$ $y = 0 = \text{効果無}$
$x = 1 = \text{薬A}$	$P(1, 1, 1) = 1/4$ $P(1, 0, 1) = 0$
$x = 0 = \text{薬B}$	$P(0, 1, 1) = 0$ $P(0, 0, 1) = 1/4$

より、 $x = 1$ の薬Aの場合には $y = 1$ の効果有の確率が正であるが、 $x = 0$ の薬Bの場合には $y = 1$ の効果有の確率が0になっている。 $z = 0$ の男性の場合に制限した確率質量関数 $P(x, y, z)$ の表

	$z = 0 = \text{男性}$
	$y = 1 = \text{効果有}$ $y = 0 = \text{効果無}$
$x = 1 = \text{薬A}$	$P(1, 1, 0) = 0$ $P(1, 0, 0) = 1/4$
$x = 0 = \text{薬B}$	$P(0, 1, 0) = 1/4$ $P(0, 0, 0) = 0$

より、 $x = 1$ の薬Aの場合には $y = 1$ の効果有の確率が0で効果がないが、 $x = 0$ の薬Bの場合には $y = 1$ の効果有の確率が正になっている。

このように現実世界において確率変数達が独立か否かは重大な意味を持ち得る。ある重要な条件(上の場合には女性か男性か)を無視して、「XとYは独立である」(どちらの薬を与えても効果は同じである)と結論すると大変なことになってしまうかもしれない。XとYも、XとZも、YとZも独立であっても、XとYとZの全体は独立でないかもしれない。

2 確率変数達の共分散と相関係数と無相関性

分散と共分散と相関係数と無相関性に関する理論は線形代数における有限実ベクトル空間における内積の理論と本質的に同じものであると考えられる。分散、共分散、相関係数、無相関性はそれぞれベクトルのノルムの2乗、内積、ベクトルのあいだの角度 θ に関する $\cos \theta$ 、ベクトルの直交に対応している。

確率変数	線形代数
分散	ノルムの2乗
共分散	内積
相関係数	$\cos \theta$
無相関	直交

2.1 確率変数達の共分散と相関係数の定義

確率関数 X, Y の期待値をそれぞれ $\mu_X = E[X]$, $\mu_Y = E[Y]$ と書くことにする.

確率変数 X, Y の **共分散** (covariance) $\sigma_{XY} = \text{cov}(X, Y)$ を次のように定義する:

$$\sigma_{XY} = \text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)].$$

$X = Y$ のときこれは X の分散になる. 分散の場合と同様に

$$\begin{aligned} \text{cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY] - \mu_X E[Y] - E[X]\mu_Y + \mu_X \mu_Y \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y]. \end{aligned}$$

$X = Y$ の場合から $\text{var}(X) = E[X^2] - E[X]^2$ が得られる.

線形代数の言葉を使えば, 共分散 $\sigma_{XY} = \text{cov}(X, Y)$ は内積に対応しており, 分散 $\sigma_X^2 = \text{var}(X)$, $\sigma_Y^2 = \text{var}(Y)$ はノルム(ベクトルの長さ)の二乗に対応しており, 標準偏差 $\sigma_X = \text{std}(X)$, $\sigma_Y = \text{std}(Y)$ はノルム(ベクトルの長さ)に対応している.

さらに確率変数 X, Y の **相関係数** (correlation coefficient) $\rho_{XY} = \text{cov}(X, Y)$ はベクトルのあいだのなす角度 θ に対する $\cos \theta$ の対応物として次のように定義される:

$$\rho_{XY} = \text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\text{std}(X) \text{std}(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

相関係数という言葉を見たら $\cos \theta$ を想像し, 分散, 標準偏差, 共分散という言葉を見たら, ベクトルの長さの2乗, ベクトルの長さ, ベクトルの内積を想像すればよい. $X - \mu_X$ や $Y - \mu_Y$ がベクトルに対応している.

非常に重要な注意・警告: 以上で定義したのは確率変数達の共分散と相関係数であり, 後で定義する標本(サンプル, データ)の共分散と相関係数とは異なる数学的対象になるので注意が必要である. この点は平均と分散の場合と同様である. 我々は複雑な現実世界に立ち向かうための道具として数学的に統計モデルを設定し, 統計モデル内の住人として確率変数を考える. そのとき, 確率変数達の平均(期待値), 分散, 共分散, 相関係数を計算することは, 統計モデル内の確率変数の特徴を調べていることになる. 統計モデルを数学的フィクションである現実世界のデータと比較することは以下のように行われる:

1. 数学的フィクションである統計モデル内に住んでる現実世界のデータのモデル化になっている確率変数の特徴を調べる.
2. 現実世界で取得したデータの特徴を調べる.
3. 統計モデル内の確率変数の特徴と現実世界で取得したデータの特徴を比較してみる.

上で定義したの確率変数達の共分散と相関係数は1の段階で使用される. 2の段階で使われる道具については後で説明する.

数学的な注意: 統計学的な位置付けとして, 確率変数達の平均, 分散, 共分散, 相関係数などと, 標本(サンプル, データ)平均, 分散, 共分散, 相関係数は異なる対象だとみなす必要があることはすぐ上で説明した. しかし, 純粋に数学的性質を調べる場合には, 後者は前者の特別な場合になっていることから, 前者の数学的性質を調べれば後者の数学的性質もわかるという仕組みになっている. 初学者が混乱しがちになる原因の1つは, 数学的に鋭い初学者が確率変数達の扱いとデータ(サンプル, 標本)の取り扱いが数学的に似ていることに気付いてしまうせいで, それらの位置付けを混同してしまいがちなことにもあるように思われる.

2.2 確率変数達の無相関の定義

X, Y の相関係数 $\rho_{XY} = \text{cor}(X, Y)$ が 0 になるとき, 確率変数達 X, Y は **無相関** であるという. X と Y が無相関であることは $X - \mu_X$ と $Y - \mu_Y$ が「直交している」と考えてもよい.

注意: X, Y の相関係数 $\rho_{XY} = \text{cor}(X, Y)$ が **近似的に 0 になっている場合** には($\cos \theta$ に対応する量が近似的に 0 になっている場合には), 確率変数達 X, Y は **近似的に無相関** であると考え. そのとき注意すべきことは共分散 $\sigma_{XY} = \text{cov}(X, Y)$ が近似的に 0 になっていても(内積に当たる量が近似的に 0 になっていても), 相関係数が 0 からほど遠い値になっていて, 近似的に無相関とは言えない場合があることである. 下にある関連問題の節の内容を参照せよ.

確率変数達 X_1, \dots, X_n が **無相関** であるとは, そのうちの互いに異なる任意の2つが無相関であることだと定める. $\mu_i = E[X_i]$, $\sigma_i^2 = \text{var}(X_i)$ のとき, X_1, \dots, X_n が無相関であることは,

$$\text{cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = \sigma_i^2 \delta_{ij}$$

と書けることと同値である。ここで δ_{ij} は $i = j$ の場合にのみ 1 になり、それ以外の場合に 0 になる Kronecker のデルタである。

2.3 問題: 確率変数の相関係数の計算例

確率変数の組 (X, Y) は確率 0.1, 0.2, 0.3, 0.4 でそれぞれ $(10, -50), (-4, -10), (0, 20), (2, 0)$ という値になると仮定する:

probability	0.1	0.2	0.3	0.4
X	10	-4	0	2
Y	-50	-10	20	0

確率変数 X, Y の相関係数を小数点以下第2桁まで求めよ。

解答例: まず X, Y の期待値を求める:

$$\begin{aligned}\mu_X &= E[X] = 10 \cdot 0.1 + (-4) \cdot 0.2 + 0 \cdot 0.3 + 2 \cdot 0.4 = 1, \\ \mu_Y &= E[Y] = (-50) \cdot 0.1 + (-10) \cdot 0.2 + 20 \cdot 0.3 + 0 \cdot 0.4 = -1.\end{aligned}$$

ゆえに期待値を X, Y から引いた値の分布は次の表ようになる:

probability	0.1	0.2	0.3	0.4
$X - \mu_X$	9	-5	-1	1
$Y - \mu_Y$	-49	-9	21	1

この表を用いて確率変数達としての X, Y の分散と共分散を求める:

$$\begin{aligned}\text{var}(X) &= 9^2 \cdot 0.1 + (-5)^2 \cdot 0.2 + (-1)^2 \cdot 0.3 + 1^2 \cdot 0.4 = 13.8, \\ \text{var}(Y) &= (-49)^2 \cdot 0.1 + (-9)^2 \cdot 0.2 + 21^2 \cdot 0.3 + 1^2 \cdot 0.4 = 389, \\ \text{cov}(X, Y) &= 9(-49) \cdot 0.1 + (-5)(-9) \cdot 0.2 + (-1)21 \cdot 0.3 + 1 \cdot 1 \cdot 0.4 = -41.\end{aligned}$$

相関係数を求めよう:

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} \approx -0.56.$$

解答終

```
In [26]: 1 @show E_X = 10*0.1 + (-4)*0.2 + 0*0.3 + 2*0.4
          2 @show E_Y = (-50)*0.1 + (-10)*0.2 + 20*0.3 + 0*0.4
```

```
E_X = 10 * 0.1 + -4 * 0.2 + 0 * 0.3 + 2 * 0.4 = 1.0
E_Y = -50 * 0.1 + -10 * 0.2 + 20 * 0.3 + 0 * 0.4 = -1.0
```

Out[26]: -1.0

```
In [27]: 1 @show var_X = 9^2*0.1 + (-5)^2*0.2 + (-1)^2*0.3 + 1^2*0.4
          2 @show var_Y = (-49)^2*0.1 + (-9)^2*0.2 + 21^2*0.3 + 1^2*0.4
          3 @show cov_XY = 9*(-49)*0.1 + (-5)*(-9)*0.2 + (-1)*21*0.3 + 1*1*0.4;
```

```
var_X = 9 ^ 2 * 0.1 + (-5) ^ 2 * 0.2 + (-1) ^ 2 * 0.3 + 1 ^ 2 * 0.4 = 13.8
var_Y = (-49) ^ 2 * 0.1 + (-9) ^ 2 * 0.2 + 21 ^ 2 * 0.3 + 1 ^ 2 * 0.4 = 389.0
cov_XY = 9 * -49 * 0.1 + -5 * -9 * 0.2 + -1 * 21 * 0.3 + 1 * 1 * 0.4 = -41.0
```

```
In [28]: 1 @show cor_XY = cov_XY/(sqrt(var_X) * sqrt(var_Y));
```

```
cor_XY = cov_XY / (sqrt(var_X) * sqrt(var_Y)) = -0.5595894432479921
```

In [29]:

```
1 X = [10, -4, 0, 2]
2 Y = [-50, -10, 20, 0]
3 P = [0.1, 0.2, 0.3, 0.4]
4 @show E_X = sum(X[i]*P[i] for i in eachindex(P))
5 @show E_Y = sum(Y[i]*P[i] for i in eachindex(P))
6 @show var_X = sum((X[i] - E_X)^2*P[i] for i in eachindex(P))
7 @show var_Y = sum((Y[i] - E_Y)^2*P[i] for i in eachindex(P))
8 @show cov_XY = sum((X[i] - E_X) * (Y[i] - E_Y) * P[i] for i in eachindex(P))
9 @show cor_XY = cov_XY/(sqrt(var_X) * sqrt(var_Y));
```

```
E_X = sum((X[i] * P[i] for i = eachindex(P))) = 1.0
E_Y = sum((Y[i] * P[i] for i = eachindex(P))) = -1.0
var_X = sum(((X[i] - E_X) ^ 2 * P[i] for i = eachindex(P))) = 13.8
var_Y = sum(((Y[i] - E_Y) ^ 2 * P[i] for i = eachindex(P))) = 389.0
cov_XY = sum(((X[i] - E_X) * (Y[i] - E_Y) * P[i] for i = eachindex(P))) = -41.0
cor_XY = cov_XY / (sqrt(var_X) * sqrt(var_Y)) = -0.5595894432479921
```

2.4 問題: Cauchy-Schwarzの不等式

確率変数達 X, Y について、以下の不等式を示せ:

$$\text{cov}(X, Y) \leq \text{std}(X) \text{std}(Y).$$

簡単のため X, Y が同時確率質量関数を持つと仮定してよい。

注意: Cauchy-Schwarzの不等式より、 $\text{std}(X) \text{std}(Y) \neq 0$ のとき、相関係数

$$\rho_{XY} = \text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\text{std}(X) \text{std}(Y)}$$

が -1 以上 1 以下になることがわかる。これより、 0 以上 π 以下の θ を

$$\rho_{XY} = \text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\text{std}(X) \text{std}(Y)} = \cos \theta$$

によって定めることができる。この θ はベクトルのあいだの角度の確率変数の場合の類似物になっている。

解答例1: 右辺の2乗から左辺の2乗を引いた結果が 0 以上になること、すなわち、次を示せばよい:

$$\text{var}(X) \text{var}(Y) - \text{cov}(X, Y)^2 \geq 0. \quad (*)$$

確率変数の分散や共分散は確率変数から定数を引いても変わらないので、 X, Y からそれらの期待値を引いたものを考えることによって、 X, Y の期待値は 0 であると仮定してよい。このとき、

$$\text{var}(X) = E[X^2], \quad \text{var}(Y) = E[Y^2], \quad \text{cov}(X, Y) = E[XY].$$

X, Y の同時確率質量関数を $P(x, y)$ と書く:

$$E[f(X, Y)] = \sum_{x, y} f(x, y) P(x, y).$$

このとき、

$$E[f(X, Y)]E[g(X, Y)] = \sum_{x, y, x', y'} f(x, y)g(x', y')P(x, y)P(x', y').$$

以上の準備のもとで、

$$\begin{aligned} 2 \text{var}(X) \text{var}(Y) &= E[X^2]E[Y^2] + E[Y^2]E[X^2] \\ &= \sum_{x, y, x', y'} (x^2 y'^2 + y^2 x'^2) P(x, y) P(x', y'), \\ 2 \text{cov}(X, Y)^2 &= 2 E[XY]E[XY] = \sum_{x, y, x', y'} 2xyx'y' P(x, y) P(x', y') \\ &= \sum_{x, y, x', y'} 2xy'yx' P(x, y) P(x', y'). \end{aligned}$$

ゆえに、

$$\begin{aligned} 2 \operatorname{var}(X) \operatorname{var}(Y) - 2 \operatorname{cov}(X, Y)^2 &= \sum_{x, y, x', y'} (x^2 y'^2 + y^2 x'^2 - 2xy' yx') P(x, y) P(x', y') \\ &= \sum_{x, y, x', y'} (xy' - yx')^2 P(x, y) P(x', y') \geq 0. \end{aligned}$$

これより(*)が成立することがわかる.

解答終

上の解答例でやっていることは本質的に (X, Y) と独立で同分布な確率変数 (X', Y') に対応する離散変数 x', y' を導入したことである. そのことに気付けば, 同時確率質量関数の存在に頼らない一般的な証明法も得られる.

解答例2(より一般的な場合): 右辺の2乗から左辺の2乗を引いた結果が 0 以上になること, すなわち, 次を示せばよい:

$$\operatorname{var}(X) \operatorname{var}(Y) - \operatorname{cov}(X, Y)^2 \geq 0. \quad (*)$$

確率変数の分散や共分散は確率変数から定数を引いても変わらないので, X, Y からそれらの期待値を引いたものを考えることによって, X, Y の期待値は 0 であると仮定してよい. このとき,

$$\operatorname{var}(X) = E[X^2], \quad \operatorname{var}(Y) = E[Y^2], \quad \operatorname{cov}(X, Y) = E[XY].$$

さらに, (X', Y') は (X, Y) と同じ分布を持つ (X, Y) と独立な確率変数であるとする. このとき, 次の公式を使用できる:

$$E[f(X, Y)] = E[f(X', Y')], \quad E[f(X, Y)g(X', Y')] = E[f(X, Y)]E[g(X', Y')].$$

以上の準備のもとで,

$$\begin{aligned} 2 \operatorname{var}(X) \operatorname{var}(Y) &= 2 E[X^2] E[Y^2] = E[X^2] E[Y'^2] + E[Y^2] E[X'^2] = E[X^2 Y'^2 + Y^2 X'^2], \\ 2 \operatorname{cov}(X, Y)^2 &= 2 E[XY]^2 = 2 E[XY] E[X'Y'] = E[2XY'YX']. \end{aligned}$$

ゆえに,

$$\begin{aligned} 2 \operatorname{var}(X) \operatorname{var}(Y) - 2 \operatorname{cov}(X, Y)^2 &= E[X^2 Y'^2 + Y^2 X'^2] - E[2XY'YX'] \\ &= E[X^2 Y'^2 + Y^2 X'^2 - 2XY'YX'] \\ &= E[(XY' - X'Y)^2] \geq 0. \end{aligned}$$

これより(*)が成立することがわかる.

解答終

注意: 線形代数の教科書におけるCauchy-Schwarzの不等式の証明と以上で紹介した証明を比較してみよ. もしも比較した線形代数の教科書の証明が以上で紹介した証明と全く違っているならば, その教科書の文脈において上で紹介した証明の方法が使えないかどうかを考えてみよ. 参照にした教科書に上で紹介した証明と全く異なる証明が載っているならば, 上で証明した確率変数に関するCauchy-Schwarzの不等式をその教科書の方法で証明してみよ.

2.5 問題: 等確率有限離散分布の相関係数と $\cos \theta$ の関係

$a_i, b_i \in \mathbb{R}$ であるとし, 確率変数の組 (X, Y) は同じ確率 $1/n$ で $(a_1, b_1), \dots, (a_n, b_n)$ のどれかの値になるものと仮定する. この仮定は

$$E[f(X, Y)] = \frac{1}{n} \sum_{i=1}^n f(a_i, b_i)$$

の成立と同値である(確率変数の組が定義されていることはその関数の期待値が定義されていることと同じことだと考えてよい). さらに,

$$\mu_X = E[X] = \frac{1}{n} \sum_{i=1}^n a_i, \quad \mu_Y = E[Y] = \frac{1}{n} \sum_{i=1}^n b_i$$

とおき, ベクトル v_X, v_Y を次のように定める:

$$v_X = \frac{1}{\sqrt{n}} \begin{bmatrix} a_1 - \mu_X \\ \vdots \\ a_n - \mu_X \end{bmatrix}, \quad v_Y = \frac{1}{\sqrt{n}} \begin{bmatrix} b_1 - \mu_Y \\ \vdots \\ b_n - \mu_Y \end{bmatrix}$$

これらの内積(対応する成分の積の和)を (v_X, v_Y) と書き, ノルム(成分の2乗の和の平方根)を $\|v_X\|, \|v_Y\|$ と書き, v_X, v_Y のあいだの角度を θ と書くことにし, $\|v_X\| > 0, \|v_Y\| > 0$ であると仮定する. このとき, X, Y の相関係数 $\operatorname{cor}(X, Y)$ について次が成立することを示せ:

$$\text{cor}(X, Y) = \frac{(v_X, v_Y)}{\|v_X\| \|v_Y\|} = \cos \theta.$$

注意: 相関係数とベクトルのあいだの角度の関係についてはこの場合をイメージしておけばよい。

解答例: 内積とノルムの定義より、

$$\begin{aligned}\|v_X\|^2 &= \frac{1}{n} \sum_{i=1}^n (a_i - \mu_X)^2 = E[(X - \mu_X)^2] = \text{var}(X), \\ \|v_Y\|^2 &= \frac{1}{n} \sum_{i=1}^n (b_i - \mu_Y)^2 = E[(Y - \mu_Y)^2] = \text{var}(Y), \\ (v_X, v_Y) &= \frac{1}{n} \sum_{i=1}^n (a_i - \mu_X)(b_i - \mu_Y) = E[(X - \mu_X)(Y - \mu_Y)] = \text{cov}(X, Y).\end{aligned}$$

ゆえに、

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} = \frac{(v_X, v_Y)}{\|v_X\| \|v_Y\|} = \cos \theta.$$

解答終

2.6 問題: 相関係数の計算

確率変数達 X, Y は無相関であるとし、それぞれの期待値と分散はどちらも 0 と 1 であるとし、 $a, b, c, d \in \mathbb{R}$, $a^2 + b^2 > 0$, $c^2 + d^2 > 0$ であると仮定する。確率変数 A, B を次のように定める:

$$A = aX + bY, \quad B = cX + dY.$$

確率変数達 A, B の相関係数を求めよ。

解答例: X, Y の期待値はどちらも 0 なのでそれらの一次結合である A, B の期待値も 0 になる。さらに、 X, Y の分散はどちらも 1 でそれらは無相関だと仮定したので、

$$\begin{aligned}\text{var}(A) &= a^2 \text{var}(X) + b^2 \text{var}(Y) = a^2 + b^2, \\ \text{var}(B) &= c^2 \text{var}(X) + d^2 \text{var}(Y) = c^2 + d^2, \\ \text{cov}(A, B) &= E[AB] = E[acX^2 + (ad + bc)XY + bdY^2] \\ &= ac \text{var}(X) + (ad + bc) \text{cov}(X, Y) + bd \text{var}(Y) = ac + bd.\end{aligned}$$

ゆえに

$$\text{cor}(A, B) = \frac{\text{cov}(A, B)}{\sqrt{\text{var}(A)}\sqrt{\text{var}(B)}} = \frac{ac + bd}{\sqrt{a^2 + b^2}\sqrt{c^2 + d^2}}.$$

解答終

注意: ベクトル v_A, v_B を

$$v_A = \begin{bmatrix} a \\ b \end{bmatrix}, \quad v_B = \begin{bmatrix} c \\ d \end{bmatrix}$$

と定め、これらの標準内積(対応する成分の積の和)を (v_A, v_B) と書き、それらのノルム(すべての成分の2乗の和の平方根)を $\|v_A\|, \|v_B\|$ と書き、それらの間の角度を θ と書くとき、

$$\text{cor}(A, B) = \frac{(v_A, v_B)}{\|v_A\| \|v_B\|} = \cos \theta$$

が成立している。上の問題の状況において、確率変数 A, B の相関係数が 1 に近いことはベクトル v_A, v_B の向きがほぼ同じであることと同値であり、 A, B の相関係数が -1 に近いことは v_A, v_B の向きがほぼ正反対であることに同値であり、 A, B が無相関に近いことは v_A, v_B がほぼ直交していることと同値である。

2.7 問題: 共分散が 0 に近くても相関係数が 0 から遠い場合がある

平均が 0 の具体的な確率変数達 X, Y の組で、共分散 $\sigma_{XY} = \text{cov}(X, Y) = E[XY]$ は 0 に近い値だが、相関係数 $\rho_{XY} = \text{cov}(X, Y) / (\text{std}(X) \text{std}(Y))$ が 0 から遠い値になっているものを構成せよ。

解答例: $\varepsilon, \delta > 0$ であるとし、確率変数の組 (X, Y) は同じ確率 $1/3$ で $(\varepsilon + \delta, \varepsilon)$, $(-\varepsilon, \delta)$, $(-\delta, -\varepsilon - \delta)$ のどれかの値になるものとする。このとき、 X, Y の平均はそれぞれ

$$E[X] = \frac{(\varepsilon + \delta) + (-\varepsilon) + (-\delta)}{3} = 0, \quad E[Y] = \frac{\varepsilon + \delta + (-\varepsilon - \delta)}{3} = 0$$

となる. X と Y の分散と共分散と相関係数は以下ようになる:

$$\begin{aligned} \text{var}(X) &= E[X^2] = \frac{(\varepsilon + \delta)^2 + (-\varepsilon)^2 + (-\delta)^2}{3} = \frac{2(\varepsilon^2 + \varepsilon\delta + \delta^2)}{3}, \\ \text{var}(Y) &= E[Y^2] = \frac{\varepsilon^2 + \delta^2 + (-\varepsilon - \delta)^2}{3} = \frac{2(\varepsilon^2 + \varepsilon\delta + \delta^2)}{3}, \\ \text{cov}(X, Y) &= E[XY] = \frac{(\varepsilon + \delta)\varepsilon + (-\varepsilon)\delta + (-\delta)(-\varepsilon - \delta)}{3} = \frac{(\varepsilon^2 + \varepsilon\delta + \delta^2)}{3}, \\ \text{cor}(X, Y) &= \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} = \frac{1}{2}. \end{aligned}$$

$\varepsilon, \delta > 0$ の値が小さいとき, X, Y の共分散は小さくなるが, 相関係数は $1/2$ で全然 0 に近くない. **解答終**

注意: 3次元ベクトル v_X, v_Y を以下のように定める:

$$v_X = \frac{1}{\sqrt{3}} \begin{bmatrix} \varepsilon + \delta \\ -\varepsilon \\ -\delta \end{bmatrix}, \quad v_Y = \frac{1}{\sqrt{3}} \begin{bmatrix} \varepsilon \\ \delta \\ -\varepsilon - \delta \end{bmatrix}.$$

このとき, これらの標準内積(対応する成分の積の和)を (v_X, v_Y) と書き, それらのノルム(すべての成分の2乗の和の平方根)を $\|v_X\|, \|v_Y\|$ と書き, それらの間の角度を θ と書くとき,

$$\begin{aligned} \text{var}(X) &= \|v_X\|^2, \quad \text{var}(Y) = \|v_Y\|^2, \quad \text{cov}(X, Y) = (v_X, v_Y), \\ \text{cor}(X, Y) &= \frac{(v_X, v_Y)}{\|v_X\|\|v_Y\|} = \cos \theta \end{aligned}$$

が成立していることに注意せよ.

2.8 問題: 独立ならば無相関である (実質1行で解ける)

確率変数達 X, Y が独立ならば確率変数達 X, Y は無相関であることを示せ.

解答例: X, Y は独立なので $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$ となる. ゆえに, $\mu_X = E[X], \mu_Y = E[Y]$ とおくと,

$$E[(X - \mu_X)(Y - \mu_Y)] = E[X - \mu_X]E[Y - \mu_Y] = (E[X] - \mu_X)(E[Y] - \mu_Y) = 0 \cdot 0 = 0.$$

解答終

2.9 問題: 無相関でも独立とは限らない

確率変数 X, Y で無相関だが独立でないものを具体的に構成せよ.

解答例1: 確率質量関数 $P(x, y)$ を次の表の通りに定める:

	$y = 1$	$y = 2$	$y = 3$	
$x = 1$	$P(1, 1) = 0$	$P(1, 2) = 1/4$	$P(1, 3) = 0$	$P(x = 1) = 1/4$
$x = 2$	$P(2, 1) = 1/4$	$P(2, 2) = 0$	$P(2, 3) = 1/4$	$P(x = 2) = 2/4$
$x = 3$	$P(3, 1) = 0$	$P(3, 2) = 1/4$	$P(3, 3) = 0$	$P(x = 3) = 1/4$
	$P(y = 1) = 1/4$	$P(y = 2) = 2/4$	$P(y = 3) = 1/4$	

確率変数 X, Y は同時確率質量関数 $P(x, y)$ を持つとする. このとき,

$$E[X] = E[Y] = 1 \cdot (1/4) + 2 \cdot (2/4) + 3 \cdot (1/4) = 2.$$

であり, 確率が 0 でない (x, y) については x または y が X と Y の期待値 2 になるので, 共分散は 0 になる:

$$\begin{aligned} \text{cov}(X, Y) &= E[(X - 2)(Y - 2)] = \sum_{x, y} (x - 2)(y - 2)P(x, y) \\ &= (2 - 2)(1 - 2)(1/4) + (1 - 2)(2 - 2)(1/4) \\ &\quad + (3 - 2)(2 - 2)(1/4) + (2 - 2)(3 - 2)(1/4) = 0. \end{aligned}$$

これで X, Y は無相関であることがわかった.

しかし, $P(1, 1) = 0$ と $P(x = 1)P(y = 1) = (1/4)(1/4)$ は等しくないので X, Y は独立ではない.

解答終

解答例2: 確率質量函数 $P(x, y)$ を次の表の通りに定めても上と同様に, X, Y は無相関だが独立ではないことを示せる:

	$y = 1$	$y = 2$	$y = 3$	
$x = 1$	$P(1, 1) = 1/8$	$P(1, 2) = 1/8$	$P(1, 3) = 1/8$	$P(x = 1) = 3/8$
$x = 2$	$P(2, 1) = 1/8$	$P(2, 2) = 0$	$P(2, 3) = 1/8$	$P(x = 2) = 2/8$
$x = 3$	$P(3, 1) = 1/8$	$P(3, 2) = 1/8$	$P(3, 3) = 1/8$	$P(x = 3) = 3/8$
	$P(y = 1) = 3/8$	$P(y = 2) = 2/8$	$P(y = 3) = 3/8$	

解答終

解答例3: \mathbb{R}^2 における単位円盤上の一様分布の確率密度函数を次のように定める:

$$p(x, y) = \begin{cases} 1/\pi & (x^2 + y^2 \leq 1) \\ 0 & (x^2 + y^2 > 1) \end{cases}$$

これを同時確率密度函数として持つ確率変数 X, Y を考える:

$$E[f(X, Y)] = \iint_{\mathbb{R}^2} f(x, y)p(x, y) dx dy.$$

単位円盤の対称性から $E[X] = E[Y] = 0$ となることがわかる. (具体的に積分を計算しても易しい.) それらの共分散は

$$\text{cov}(X, Y) = E[XY] = \frac{1}{\pi} \iint_{x^2+y^2 \leq 1} xy dx dy$$

と書けるが, $xy \geq 0$ の部分の積分と $xy \leq 0$ の部分の積分が単位円盤の対称性より互いに打ち消しあって $\text{cov}(X, Y) = 0$ となることがわかる. X 単独の密度函数は

$$p(x) = \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = \frac{2}{\pi} \sqrt{1-x^2}$$

となり, 同様にして $p(y) = (2/\pi)\sqrt{1-y^2}$ となるが, $-1 < x, y < 1$ かつ $x^2 + y^2 > 1$ のとき, $p(x, y) = 0$ となるが, $p(x)p(y) \neq 0$ となるので, それらは等しくない. ゆえに X, Y は独立ではない.

解答終

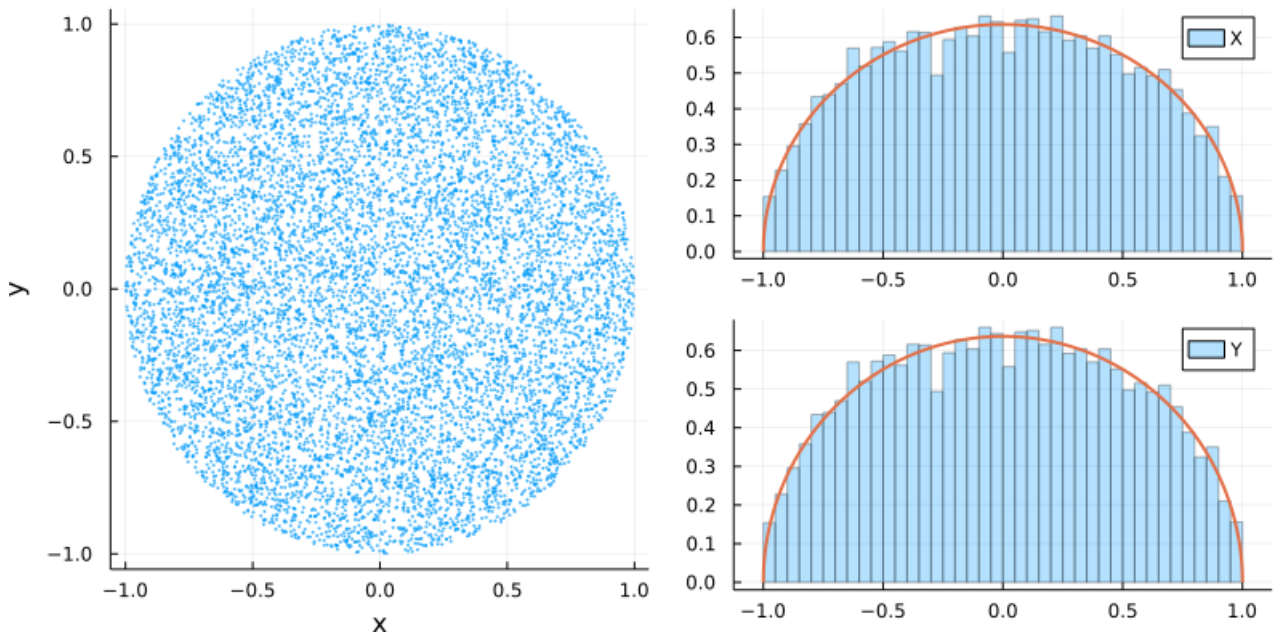
この解答例3については, 下の方でプロットした (X, Y) の分布のサンプルの散布図も参照して欲しい. 散布図を見れば直観的にどうなっているかを把握しやすい.

In [30]:

```
1 n = 10^4
2 XY = [(r = sqrt(rand()); t = 2*pi*rand()); (r*cos(t), r*sin(t))] for _ in 1:n]
3 X, Y = first.(XY), last.(XY)
4 @show cov(X, Y) cor(X, Y)
5 P1 = scatter(X, Y; msc=:auto, ms=2, alpha=0.7, label="", xlabel="x", ylabel="y")
6 P2 = histogram(X; norm=true, alpha=0.3, bin=41, label="X")
7 plot!(x -> 2/pi*sqrt(1 - x^2), -1, 1; label="", lw=2)
8 P3 = histogram(Y; norm=true, alpha=0.3, bin=41, label="Y")
9 plot!(y -> 2/pi*sqrt(1 - y^2), -1, 1; label="", lw=2)
10 plot(P1, P2, P3; size=(800, 400), layout=@layout [a [b; c]])
```

```
cov(X, Y) = -0.0005083072927225841
cor(X, Y) = -0.0020186738360558273
```

Out[30]:



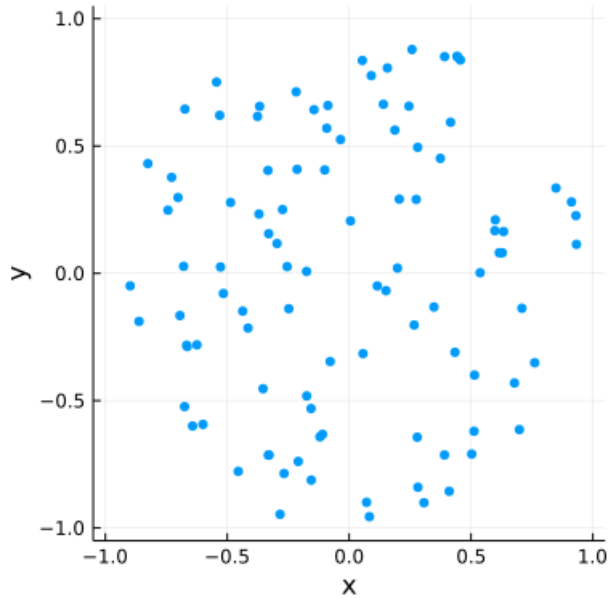
単位円盤上の一様分布は「無相関だが独立ではない場合」の例になっている。

この「無相関だが独立ではない場合」の例のサンプルに含まれる点の個数を減らすと以下のように見える。

```
In [31]: 1 n = 100
2 XY = [(r = √rand()); t = 2π*rand()); (r*cos(t), r*sin(t))] for _ in 1:n]
3 X, Y = first.(XY), last.(XY)
4 @show cov(X, Y) cor(X, Y)
5 P1 = scatter(X, Y; msc=:auto, label="", xlabel="x", ylabel="y")
6 plot!(); xlim=(-1.05, 1.05), ylim=(-1.05, 1.05))
7 plot!(); size=(400, 400))
```

```
cov(X, Y) = 0.011653127995250636
cor(X, Y) = 0.046497931398409825
```

Out[31]:



上の場合において X, Y はどちらも **半円分布** (semicircle distribution)に従っているが, 独立ではない. X, Y の分布が以上と同じ半円分布のときと X, Y が独立になるならば, (X, Y) の分布のサンプルは以下ようになる.

```

In [32]: 1 n = 10^4
2 sc = Semicircle(1)
3 X, Y = rand(sc, n), rand(sc, n)
4 @show cov(X, Y) cor(X, Y)
5 P1 = scatter(X, Y; msc=:auto, ms=2, alpha=0.7, label="", xlabel="x", ylabel="y")
6 P2 = histogram(X; norm=true, alpha=0.3, bin=41, label="X")
7 plot!(x → 2/π*√(1 - x^2), -1, 1; label="", lw=2)
8 P3 = histogram(Y; norm=true, alpha=0.3, bin=41, label="Y")
9 plot!(y → 2/π*√(1 - y^2), -1, 1; label="", lw=2)
10 plot(P1, P2, P3; size=(800, 400), layout=@layout [a [b; c]])

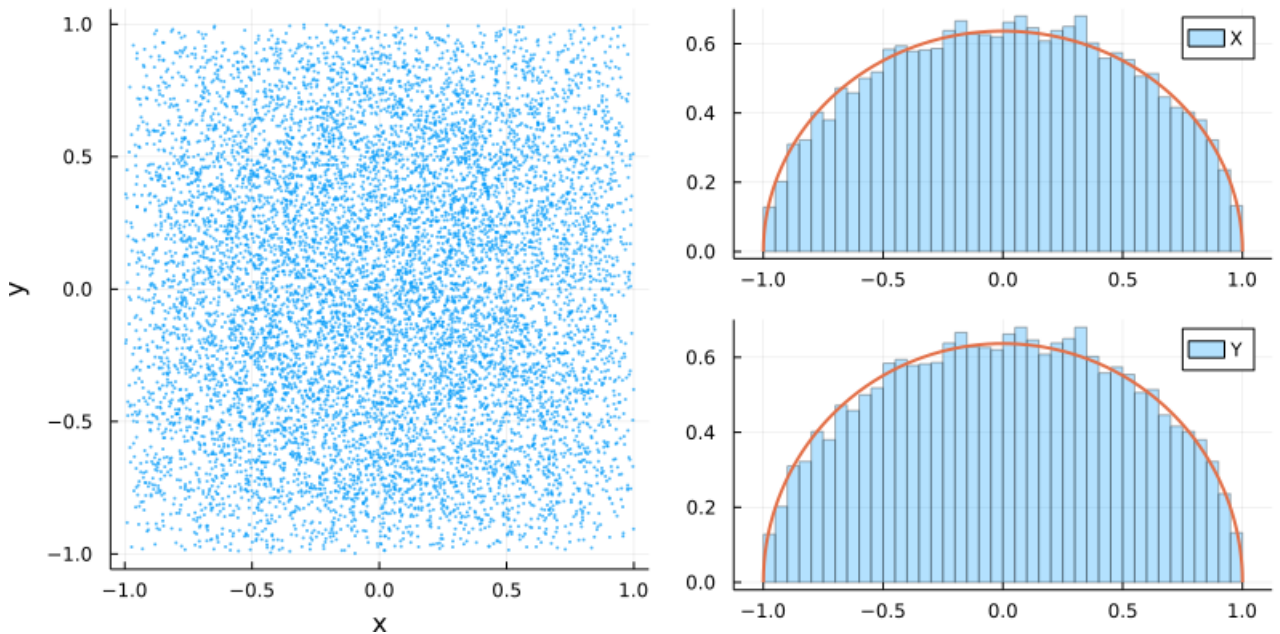
```

```

cov(X, Y) = 0.004473060988369731
cor(X, Y) = 0.0179942903137646

```

Out[32]:



これは X, Y が独立な場合であり、ゆえに無相関にもなっている。

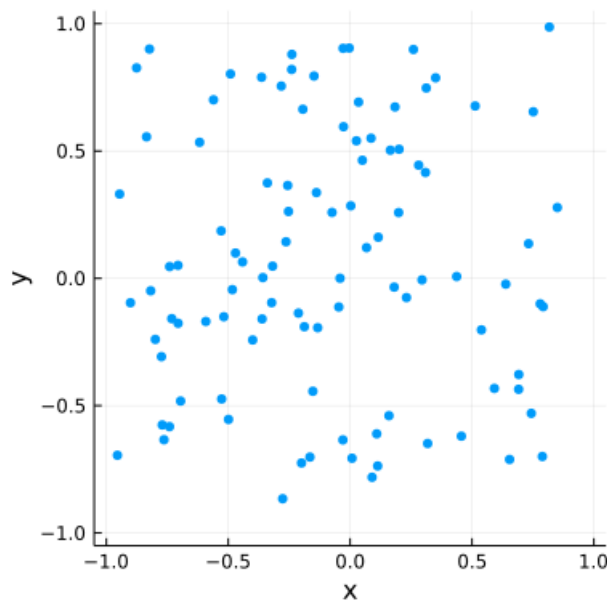
サンプルが含む点の個数を減らすと以下のように見える。

In [33]:

```
1 n = 100
2 sc = Semicircle(1)
3 X, Y = rand(sc, n), rand(sc, n)
4 @show cov(X, Y)
5 P1 = scatter(X, Y; msc=:auto, label="", xlabel="x", ylabel="y")
6 plot!(); xlim=(-1.05, 1.05), ylim=(-1.05, 1.05))
7 plot!(); size=(400, 400))
```

cov(X, Y) = 0.0030840296722668315

Out[33]:



2.10 問題: 無相関な確率変数達の和の分散はそれぞれの分散の和になる

X_1, \dots, X_n は独立でも無相関とも限らない確率変数達であるとする. このとき, 期待値を取る操作の線形性より,

$$E[X_1 + \dots + X_n] = E[X_1] + \dots + E[X_n]$$

となる. 確率変数達の和の期待値は, 独立性や無相関性が成立していなくても, それぞれの期待値の和になる.

無相関性を仮定すると, 分散についても簡明な結果を得ることができる.

X_1, \dots, X_n が **無相関な確率変数達** ならば(特に独立な確率変数ならば), 次が成立することを示せ:

$$\text{var}(X_1 + \dots + X_n) = \text{var}(X_1) + \dots + \text{var}(X_n).$$

この結果は今後空気のごとく使われる.

ヒント: 互いに直交するベクトル達 v_1, \dots, v_n について, 内積を (\cdot, \cdot) と書くとき, $(v_i, v_j) = \delta_{ij} \|v_i\|^2$ が成立することを使えば, $\|v_1 + \dots + v_n\|^2 = \|v_1\|^2 + \dots + \|v_n\|^2$ を示せることと本質的に同じ. この結果はPythagorasの定理(平面の場合は三平方の定理)そのものである.

解答例: $\mu_i = E[X_i]$ とおくと,

$$E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n \mu_i$$

が成立しており, さらに, $\sigma_i^2 = \text{var}(X_i) = E[(X_i - \mu_i)^2]$ とおくと, X_1, \dots, X_n が無相関であることより,

$$E[(X_i - \mu_i)(X_j - \mu_j)] = \text{cov}(X_i, X_j) = \sigma_i^2 \delta_{ij}$$

が成立しているので(δ_{ij} は $i = j$ の場合にのみ 1 でそれ以外るとき 0), 一般に

$$\left(\sum_{i=1}^n a_i\right)^2 = \sum_{i=1}^n a_i \sum_{j=1}^n a_j = \sum_{i,j=1}^n a_i a_j$$

と計算できることを使うと,

$$\begin{aligned}\text{var}\left(\sum_{i=1}^n X_i\right) &= E\left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i\right)^2\right] = E\left[\left(\sum_{i=1}^n (X_i - \mu_i)\right)^2\right] \\ &= E\left[\sum_{i,j=1}^n (X_i - \mu_i)(X_j - \mu_j)\right] = \sum_{i,j=1}^n E[(X_i - \mu_i)(X_j - \mu_j)] \\ &= \sum_{i,j=1}^n \sigma_i^2 \delta_{ij} = \sum_{i=1}^n \sigma_i^2 = \sum_{i=1}^n \text{var}(X_i).\end{aligned}$$

解答終

2.11 問題: 二項分布と負の二項分布の平均と分散のBernoulli分布と幾何分布の場合への帰着

Bernoulli分布 $\text{Bernoulli}(p)$ の平均と分散がそれぞれ $p, p(1-p)$ であることと、幾何分布 $\text{Geometric}(p)$ の平均と分散がそれぞれ $(1-p)/p, (1-p)/p^2$ であることを認めて、二項分布 $\text{Binomial}(n, p)$ と負の二項分布 $\text{NegativeBinomial}(k, p)$ の平均と分散を平易な計算で求めてみよ。以下を示せ:

(1) $K \sim \text{Binomial}(n, p)$ ならば $E[K] = np, \text{var}(K) = np(1-p)$.

(2) $M \sim \text{NegativeBinomial}(k, p)$ ならば $E[M] = k(1-p)/p, \text{var}(M) = k(1-p)/p^2$.

解答例:

二項分布は試行回数 n の成功確率 p のBernoulli試行で生成された 1 と 0 からなる長さ n の数列中に含まれる 1 の個数の分布であった。Bernoulli試行の確率質量関数は

$$P(x_1, \dots, x_n) = P(x_1) \cdots P(x_n), \quad P(x_i) = p^{x_i} (1-p)^{1-x_i} \quad (x_i = 1, 0)$$

とBernoulli分布の確率質量関数 $P(x_i)$ の積で書けるのであった。この事実はBernoulli分布に従う確率変数 X_1, \dots, X_n が独立であることを意味する。そして、Bernoulli試行で生成された 1 と 0 からなる長さ n の数列中に含まれる 1 の個数を意味する確率変数は $K = \sum_{i=1}^n X_i$ と書ける。このことから、

$$\begin{aligned}E[K] &= \sum_{i=1}^n E[X_i] = \sum_{i=1}^n p = np, \\ \text{var}(K) &= \sum_{i=1}^n \text{var}(X_i) = \sum_{i=1}^n p(1-p) = np(1-p).\end{aligned}$$

となることがわかる。

幾何分布は成功確率 p のBernoulli試行を 1 が1つ出るまで続けたときに出了る 0 の個数の分布であった。 M_1, \dots, M_k はそれぞれが成功確率 p の幾何分布に従う独立な確率変数であるとする。このとき、 $M = \sum_{i=1}^k M_i$ はBernoulli試行を 1 が k 回出るまで続けたときに 0 が出た個数に等しい。 M_i は $i-1$ 番目の 1 から i 番目の 1 が出るまでに出了る 0 の個数を意味する確率変数だと解釈される。このことは、 M が負の二項分布 $\text{NegativeBinomial}(k, p)$ に従う確率変数になることを意味する。このことから、

$$\begin{aligned}E[M] &= \sum_{i=1}^k E[M_i] = \sum_{i=1}^k \frac{1-p}{p} = \frac{k(1-p)}{p}, \\ \text{var}(M) &= \sum_{i=1}^k \text{var}(M_i) = \sum_{i=1}^k \frac{1-p}{p^2} = \frac{k(1-p)}{p^2}.\end{aligned}$$

となることがわかる。

解答終

注意: 計算が大幅に簡単になった!

2.12 問題: 番号が異なる確率変数達が無相関なときの確率変数の和の共分散

確率変数達 $X_1, Y_1, \dots, X_n, Y_n$ について次が成立していると仮定する:

$$\text{cov}(X_i, Y_j) = \delta_{ij} \text{cov}(X_i, Y_i).$$

このとき、次が成立することを示せ:

$$\text{cov}(X_1 + \cdots + X_n, Y_1 + \cdots + Y_n) = \text{cov}(X_1, Y_1) + \cdots + \text{cov}(X_n, Y_n).$$

ヒント: ベクトル達 $u_1, v_2, \dots, u_n, v_n$ の中の2つの異なる添え字を持つ u_i と v_j が互いに直交するならば、内積 (\cdot, \cdot) について

$$(u_1 + \cdots + u_n, v_1 + \cdots + v_n) = (u_1, v_1) + \cdots + (u_n, v_n)$$

が成立することと本質的に同じことである。

解答例:

記号の簡単のため $A_i = X_i - E[X_i]$, $B_i = Y_i - E[Y_i]$ とおく。このとき, $E[A_i] = E[B_i] = 0$ より,

$$\text{var}(A_i) = E[A_i^2], \quad \text{var}(B_i) = E[B_i^2], \quad \text{cov}(A_i, B_j) = E[A_i B_j].$$

A_i, B_i 達について上の問題を解けばよい。問題の仮定より, A_i, B_i 達について次が成立している:

$$E[A_i B_j] = \delta_{ij} E[A_i B_i].$$

ゆえに,

$$E \left[\left(\sum_{i=1}^n A_i \right) \left(\sum_{j=1}^n B_j \right) \right] = \sum_{i,j=1}^n E[A_i B_j] = \sum_{i,j=1}^n \delta_{ij} E[A_i B_i] = \sum_{i=1}^n E[A_i B_i].$$

これは

$$\text{cov}(A_1 + \cdots + A_n, B_1 + \cdots + B_n) = \text{cov}(A_1, B_1) + \cdots + \text{cov}(A_n, B_n).$$

の成立を意味する。

解答終

2.13 分散共分散行列とその半正定値性

n 個の確率変数を縦に並べてできるベクトル X と実数を成分に持つベクトル a を考える:

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, \quad a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^n.$$

このとき, X, a の転置をそれぞれ X, a^T と書くと,

$$a^T X = X^T a = \sum_{i=1}^n a_i X_i$$

なので

$$a^T X X^T a = \left(\sum_{i=1}^n a_i X_i \right)^2 \geq 0.$$

ゆえに, 期待値を取る操作の線形性と単調性より,

$$0 \leq E[a^T X X^T a] = a^T E[X X^T] a$$

でかつ, $X X^T$ は $X_i X_j$ を (i, j) 成分とする $n \times n$ 行列になるので, $E[X X^T]$ は $E[X_i X_j]$ を (i, j) 成分とする $n \times n$ の実対称行列になる。上の計算より, その実対称行列の固有値はすべて 0 以上になることがわかる。

ここで $E[X_i] = 0$ と仮定する。このとき, $E[X X^T]$ は i 番目の対角成分が X_i の分散 $\text{var}(X_i)$ で, $i \neq j$ のときの (i, j) 成分が共分散 $\text{cov}(X_i, X_j)$ であるような $n \times n$ の実対称行列になる。

このとき, 行列 $E[X X^T]$ を **分散共分散行列** と呼ぶ。

分散共分散行列の固有値はすべて 0 以上になる。この結果を分散共分散行列の **半正定値性** と呼ぶことにする。特に分散共分散行列式は 0 以上になる。

3 標本(サンプル, データ)の平均と分散と共分散と相関係数

3.1 標本平均の定義

n 個の数値 x_1, \dots, x_n で構成されたデータ(標本, サンプル)に対して,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + \cdots + x_n}{n}$$

をその **標本平均** (sample mean)と呼ぶ。

統計モデルを設定するとは現実世界で得られるデータ(標本, サンプル)の生じ方を確率分布の言葉でモデル化することである。そのとき、統計モデル内部では、データ(標本, サンプル)を、 n 個の数値ではなく、 n 個の確率変数達 X_1, \dots, X_n でモデル化することになる。その場合には、確率変数達の加法平均

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{X_1 + \cdots + X_n}{n}$$

をその **標本平均** (sample mean)と呼ぶ。 X_i 達が確率変数であることより、それらの標本平均 \bar{X} も確率変数になることに注意せよ。この文脈での標本平均 \bar{X} は確率モデル内部における確率変数になる。

統計学入門の文脈では、標本のモデル化とみなされる n 個の確率変数達 X_1, \dots, X_n は独立同分布だと仮定することが多い。

以下では、 X_i 達は同じ期待値と分散を持ち、無相関であると仮定する(この条件は独立同分布よりも弱い)。

注意・警告: 標本のモデル化としての確率変数達の **標本平均と確率変数の平均(期待値)**を明瞭に区別しなければならない。各確率変数 X_i の期待値 $E[X_i]$ や確率変数としての標本平均 \bar{X} の期待値

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] = \mu$$

と標本平均 \bar{X} を明瞭に区別することが必要になる。

統計モデル内での標本のモデル化としての確率変数達 X_1, \dots, X_n の標本平均 \bar{X} は確率変数なので、数学的には関数とみなされ、直観的にはランダムに値が決まる変数だとみなされるものになる。それに対して、標本平均 \bar{X} の期待値 $E[\bar{X}]$ は単なる数値になる。このように、確率変数としての標本平均 \bar{X} とその期待値 $E[\bar{X}]$ は全く異なる数学的対象になる。

注意: 上で期待値を取る操作 $E[\]$ の線形性を用いて、標本平均の期待値が X_i 達の共通の期待値 μ に一致すること

$$E[\bar{X}] = \mu$$

も示されていることに注意せよ。これは標本平均を取る操作が平均の **不偏推定量** を作る操作になっていることも意味している。この点については後で詳しく説明する。

3.2 問題: 無相関な確率変数達の標本平均の分散

同じ期待値 μ と分散 σ^2 を持つ無相関な n 個の確率変数達 X_1, \dots, X_n の標本分散 \bar{X} について次が成立していることを示せ:

$$\text{var}(\bar{X}) = \frac{\sigma^2}{n}$$

注意: この問題で扱っている確率変数としての \bar{X} の分散と次の節で説明する標本のモデル化 X_1, \dots, X_n の分散は異なる数学的対象であることに注意せよ。

注意: $\text{std}(\bar{X}) = \sigma/\sqrt{n}$ は **標準誤差** (standard error)と呼ばれ、SE と書かれることがある。未知の真の標準誤差の推定量をも標準誤差と呼ぶことがあるが、このノートではそのスタイルを採用せずに、未知の真の標準誤差の推定量を「標準誤差の推定量」と丁寧に呼ぶことにする。

解答例: 無相関な確率変数達の和の分散はそれぞれの確率変数の分散の和になるのであった。 X_i 達が無相関であるという仮定より、 X_i/n 達も無相関になる。そして、

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$$

であることより、

$$\text{var}(\bar{X}) = \sum_{i=1}^n \text{var}\left(\frac{X_i}{n}\right) = \sum_{i=1}^n \frac{\text{var}(X_i)}{n^2} = \sum_{i=1}^n \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

解答終

3.3 標本分散と不偏分散の定義

n 個の数値 x_1, \dots, x_n で構成されたデータ(標本, サンプル)の標本平均を

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + \cdots + x_n}{n}$$

と書くとき、

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

を **標本分散** (sample variance, 標本の無補正分散) と呼ぶ。慣習的には補正無し標本分散が使用されることは少なく、次のように定義される **不偏分散** (unbiased variance, 標本の不偏補正分散) を使うことが多い:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

どうして n ではなく $n-1$ で割るかについては後で説明する。このとき、

$$\overline{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2$$

とおくと、

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

なので、

$$s^2 = \frac{n}{n-1} (\overline{x^2} - \bar{x}^2).$$

X_1, \dots, X_n は標本のモデル化とみなされる n 個の確率変数達であるとする。 X_i 達は同じ期待値と分散を持ち、無相関であると仮定し、それらの標本平均を

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{X_1 + \cdots + X_n}{n}$$

と書く。このとき、 X_1, \dots, X_n の **不偏分散** S^2 を次のように定義する:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

このとき、

$$\overline{X^2} = \frac{1}{n} \sum_{i=1}^n x_i^2$$

とおくと、

$$\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - 2\bar{X} \sum_{i=1}^n X_i + n\bar{X}^2 = n\overline{X^2} - 2n\bar{X}^2 + n\bar{X}^2 = n(\overline{X^2} - \bar{X}^2)$$

なので、

$$S^2 = \frac{n}{n-1} (\overline{X^2} - \bar{X}^2).$$

この場合には、不偏分散 S^2 は確率変数になることに注意せよ。この S^2 と単なる数値になる \bar{X} の確率変数としての分散 $\text{var}(\bar{X})$ は異なる数学的対象になるので混乱しないようにして欲しい。

3.4 不偏推定量について: 不偏分散の定義ではどうして n ではなく $n-1$ で割るか

X_1, \dots, X_n に共通の期待値と分散をそれぞれ μ, σ^2 と書くことにする。 X_1, \dots, X_n を現実で得られる標本のモデル化とみなす。そのとき、統計モデル内での μ, σ^2 の値は未知の真の平均と分散の値であって欲しい。この設定では未知の μ と σ^2 を標本から推定する方法が欲しくなる。

未知の μ の推定量として優れているのは標本平均である。標本のモデル化とみなされる X_1, \dots, X_n の標本平均 \bar{X} についてはその期待値 $E[\bar{X}]$ が未知の真の平均値 μ に等しくなる。この性質を標本平均の **不偏性** という。未知の値 θ の推定量を $E[T] = \theta$ を満たす確率変数 T とするとき、 T は θ の **不偏推定量** (unbiased estimator) と呼ぶ。

次に、不偏分散の定義でどうして n ではなく $n - 1$ で割るかを理解するために、 $n - 1$ で割る前の量の期待値を計算してみよう。

X_1, \dots, X_n 達の標本平均 \bar{X} の期待値と分散がそれぞれ次のようになることはわかっている:

$$E[\bar{X}] = \mu, \quad \text{var}(\bar{X}) = E[(\bar{X} - \mu)^2] = \frac{\sigma^2}{n}.$$

そして、以下が成立している:

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})^2 &= \sum_{i=1}^n ((X_i - \mu) - (\bar{X} - \mu))^2 \\ &= \sum_{i=1}^n ((X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2) \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2 \left(\sum_{i=1}^n (X_i - \mu) \right) (\bar{X} - \mu) + n(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - 2n(\bar{X} - \mu)^2 + n(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2. \end{aligned}$$

4つめの等号で $\sum_{i=1}^n X_i = n\bar{X}$ を用いた。ゆえに、

$$E \left[\sum_{i=1}^n (X_i - \bar{X})^2 \right] = \sum_{i=1}^n E[(X_i - \mu)^2] - nE[(\bar{X} - \mu)^2] = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2.$$

これより、 σ^2 の不偏推定量を作るためには $\sum_{i=1}^n (X_i - \bar{X})^2$ を $n - 1$ で割らなければいけないことがわかる。実際にそのように定義した σ^2 の不偏推定量が不偏分散 S^2 である:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, \quad E[S^2] = \sigma^2.$$

このように、 $\sum_{i=1}^n (X_i - \bar{X})^2$ の確率変数としての分散を計算して $(n-1)\sigma^2$ になることを確認すれば、 n ではなく $n - 1$ で割る理由は明瞭である。

さらに

$$\widehat{\text{SE}} = \sqrt{\frac{S^2}{n}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2}$$

を **標準誤差の推定量** として採用する。このとき、

$$E[\widehat{\text{SE}}^2] = \frac{E[S^2]}{n} = \frac{\sigma^2}{n} = \text{var}(\bar{X}) = \text{SE}^2$$

なので、 $\widehat{\text{SE}}^2$ は標準誤差の2乗 $\text{SE}^2 = \text{var}(\bar{X}) = \sigma^2/n$ の不偏推定量になっている。

注意: 不偏推定量の方が不偏でない推定量より常に優れていると考えるのは誤りである。推定量に不偏性の条件を課すと平均二乗誤差が大きくなってしまうことが多い(不偏分散についても実際にそうなることを後で示す)。不偏推定量を使うか否かはそうすることのメリットとデメリットの両方を考慮に入れて決定するべきである。ただし、上で紹介した不偏分散については害は相当に小さく、慣習的に非常によく使われている。

注意: X_1, \dots, X_n は標本のモデル化とみなされる独立同分布な確率変数達であるとし、 X_i 達に共通の期待値と分散をそれぞれ μ , σ^2 と書くことにする。このとき、その標本平均 $\bar{X} = (1/n) \sum_{i=1}^n X_i$ の期待値と分散がそれぞれ

$$E[\bar{X}] = \mu, \quad \text{var}(\bar{X}) = \frac{\sigma^2}{n}$$

となり、不偏分散 $S^2 = (1/(n-1)) \sum_{i=1}^n (X_i - \bar{X})^2$ の期待値が

$$E[S^2] = \sigma^2$$

となることまではすでにわかっている。この一連のノートではほぼ常に与えられた確率変数の期待値と分散を同時に計算して来た。ここでも、確率変数としての不偏分散の分散も計算するべきである。そして、さらに標本平均と不偏分散の共分散も計算するべきである。しかし、そのためには X_i 達共通の分布の **歪度** (わいど, skewness) と **尖度** (せんど, kurtosis) の定義を先にしておいた方がよい。次の節で歪度と尖度を定義する。

3.5 データの共分散の定義

n 個の数値の対 $(x_1, y_1), \dots, (x_n, y_n)$ で構成されたデータ(標本, サンプル)について, x_i, y_i 達各々の標本平均を

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

と書くとき,

$$\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

を **標本共分散** (sample covariance, 標本の無補正共分散)と呼ぶ. **慣習的には** 補正無しの標本共分散が使用されることは少なく, 次のように定義される **不偏共分散** (unbiased variance, 標本の不偏補正共分散)を使うことが多い:

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}).$$

n ではなく $n-1$ で割る理由は不偏分散の場合と同様である. 後で詳しく説明する. このとき,

$$\overline{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i$$

とおくと,

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^n x_i y_i - \bar{x} \sum_{i=1}^n y_i - \sum_{i=1}^n x_i \bar{y} + n\bar{x}\bar{y} \\ &= n\overline{xy} - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y} = n(\overline{xy} - \bar{x}\bar{y}) \end{aligned}$$

なので,

$$s_{xy} = \frac{n}{n-1} (\overline{xy} - \bar{x}\bar{y}).$$

$(X_1, Y_1), \dots, (X_n, Y_n)$ は上の型のデータのモデル化とみなされる n 個の確率変数対達であるとする. X_i 達は無相関で共通の同じ期待値 μ_X と分散 σ_X^2 を持ち, Y_i 達も無相関で共通の同じ期待値 μ_Y と分散 σ_Y^2 を持ち, X_i と Y_i の共分散は i によらない共通の値 σ_{XY} になっているとし, 異なる番号 i, j について X_i と Y_j は無相関であると仮定する. これらの仮定は次のように式で書ける:

$$\begin{aligned} E[X_i] &= \mu_X, & E[Y_i] &= \mu_Y \\ E[(X_i - \mu_X)(X_j - \mu_X)] &= \delta_{ij}\sigma_X^2, \\ E[(Y_i - \mu_Y)(Y_j - \mu_Y)] &= \delta_{ij}\sigma_Y^2, \\ E[(X_i - \mu_X)(Y_j - \mu_Y)] &= \delta_{ij}\sigma_{XY}. \end{aligned}$$

それらの X_i 達と Y_i 達それぞれの標本平均を次のように書くことにする:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i.$$

このとき, $(X_1, Y_1), \dots, (X_n, Y_n)$ の **不偏共分散** S_{XY} を次のように定義する:

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

このとき,

$$\overline{XY} = \frac{1}{n} \sum_{i=1}^n X_i Y_i$$

とおくと,

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) &= \sum_{i=1}^n X_i Y_i - \bar{X} \sum_{i=1}^n Y_i - \sum_{i=1}^n X_i \bar{Y} + n\bar{X}\bar{Y} \\ &= n\overline{XY} - n\bar{X}\bar{Y} - n\bar{X}\bar{Y} + n\bar{X}\bar{Y} = n(\overline{XY} - \bar{X}\bar{Y}) \end{aligned}$$

なので.

$$s_{XY} = \frac{n}{n-1} (\overline{XY} - n\bar{X}\bar{Y}).$$

この場合には、不偏分散 S_{XY} は確率変数になることに注意せよ。この S_{XY} と単なる数値になる \bar{X}, \bar{Y} の確率変数としての共分散 $\text{cov}(\bar{X}, \bar{Y})$ は異なる数学的対象になるので混乱しないようにして欲しい。

3.6 問題: 標本平均達の共分散

前節の設定のもとで次が成立することを示せ:

$$\text{cov}(\bar{X}, \bar{Y}) = \frac{\sigma_{XY}}{n}.$$

解答例: 標本平均の分散の計算とほぼ同じになる。

番号が異なる確率変数達が無相関なときの確率変数達の和の共分散は各々の対の共分散の和になるのであった。これを確率変数の和達

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n}, \quad \bar{Y} = \sum_{i=1}^n \frac{Y_i}{n}$$

に適用すると,

$$\text{cov}(\bar{X}, \bar{Y}) = \sum_{i=1}^n \text{cov}\left(\frac{X_i}{n}, \frac{Y_i}{n}\right) = \sum_{i=1}^n \frac{\text{cov}(X_i, Y_i)}{n^2} = \sum_{i=1}^n \frac{\sigma_{XY}}{n^2} = \frac{\sigma_{XY}}{n}.$$

解答終

3.7 問題: 不偏共分散の定義ではどうして n ではなく $n-1$ で割るか

前々の設定のもとで、不偏共分散 S_{XY} が X_i, Y_i の共分散 σ_{XY} の不偏推定量になっていることを示せ。すなわち次が成立することを示せ:

$$E[S_{XY}] = \sigma_{XY}.$$

注意: これが不偏共分散の定義で n ではなく $n-1$ で割る理由である。

解答例: 不偏分散の場合の議論とほぼ同じ。

\bar{X}, \bar{Y} の期待値と共分散がそれぞれ次のようになることはわかっている:

$$E[\bar{X}] = \mu_X, \quad E[\bar{Y}] = \mu_Y, \quad \text{cov}(\bar{X}, \bar{Y}) = E[(\bar{X} - \mu_X)(\bar{Y} - \mu_Y)] = \frac{\sigma_{XY}}{n}.$$

そして、以下が成立している:

$$\begin{aligned} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) &= \sum_{i=1}^n ((X_i - \mu_X) - (\bar{X} - \mu_X))((Y_i - \mu_Y) - (\bar{Y} - \mu_Y)) \\ &= \sum_{i=1}^n ((X_i - \mu_X)(Y_i - \mu_Y) - (X_i - \mu_X)(\bar{Y} - \mu_Y) \\ &\quad - (\bar{X} - \mu_X)(Y_i - \mu_Y) + (\bar{X} - \mu_X)(\bar{Y} - \mu_Y)) \\ &= \sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y) - \left(\sum_{i=1}^n (X_i - \mu_X)\right)(\bar{Y} - \mu_Y) \\ &\quad - (\bar{X} - \mu_X)\left(\sum_{i=1}^n (Y_i - \mu_Y)\right) + n(\bar{X} - \mu_X)(\bar{Y} - \mu_Y) \\ &= \sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y) - n(\bar{X} - \mu_X)(\bar{Y} - \mu_Y) \\ &\quad - n(\bar{X} - \mu_X)(\bar{Y} - \mu_Y) + n(\bar{X} - \mu_X)(\bar{Y} - \mu_Y) \\ &= \sum_{i=1}^n (X_i - \mu_X)(Y_i - \mu_Y) - n(\bar{X} - \mu_X)(\bar{Y} - \mu_Y). \end{aligned}$$

4つめの等号で $\sum_{i=1}^n X_i = n\bar{X}$, $\sum_{i=1}^n Y_i = n\bar{Y}$ を用いた。ゆえに,

$$\begin{aligned} E \left[\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \right] &= \sum_{i=1}^n E[(X_i - \mu_X)(Y_i - \mu_Y)] - nE[(\bar{X} - \mu_X)(\bar{Y} - \mu_Y)] \\ &= n\sigma_{XY} - n\frac{\sigma_{XY}}{n} = (n-1)\sigma_{XY}. \end{aligned}$$

ゆえに

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

より

$$E[S_{XY}] = \sigma_{XY}.$$

解答終

3.8 データの相関係数の定義 (以上の定義のまとめにもなっている)

数値データの場合の共分散の定義の設定のもとで, x_i, y_i 達の標本平均, 不偏分散達と不偏共分散を以下のように書くことにする:

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i, & \bar{y} &= \frac{1}{n} \sum_{i=1}^n y_i, \\ s_x^2 &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, & s_y^2 &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2, \\ s_{xy} &= \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}). \end{aligned}$$

s_x^2, s_y^2 の平方根をそれぞれ s_x, s_y と書く. x_i 達と y_i 達の相関係数 r_{xy} を次のように定める:

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$

この r_{xy} は数値になる.

データのモデル化である確率変数達の共分散の定義の設定のもとで, X_i, Y_i 達の標本平均, 不偏分散達と不偏共分散を以下のように書くことにする:

$$\begin{aligned} \bar{X} &= \frac{1}{n} \sum_{i=1}^n X_i, & \bar{Y} &= \frac{1}{n} \sum_{i=1}^n Y_i, \\ S_X^2 &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2, & S_Y^2 &= \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2, \\ S_{XY} &= \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}). \end{aligned}$$

S_X^2, S_Y^2 の平方根をそれぞれ S_X, S_Y と書く. X_i 達と Y_i 達の相関係数 R_{XY} を次のように定める:

$$R_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}.$$

この R_{XY} は確率変数になる. (確率変数とは数学的には「その関数の期待値が定義されている変数」のことであり, 直観的には値がランダムに決まる変数だと思える.)

3.9 問題: 最小二乗法

3.10 問題: 計算例

4 モーメントとキュムラントと歪度と尖度

4.1 モーメントとその母関数と特性関数とキュムラント母関数の定義

確率変数 X と $m = 0, 1, 2, \dots$ について

$$\mu_m(X) = E[X^m]$$

を X の m 次の **モーメント**(moment, 積率) と呼び、

$$M_X(t) = E[e^{tX}] = E\left[\sum_{m=0}^{\infty} X^m \frac{t^m}{m!}\right] = \sum_{m=0}^{\infty} E[X^m] \frac{t^m}{m!} = \sum_{m=0}^{\infty} \mu_m(X) \frac{t^m}{m!}$$

を **モーメント母関数** (moment generating function, mgf) と呼ぶ。

X が従う確率分布の名前が Dist のとき、これらを **分布 Dist のモーメントとモーメント母関数** と呼ぶ。以下も同様である。

モーメント母関数の定義で t を $it = \sqrt{-1} t$ で置き換えたもの

$$\begin{aligned}\varphi_X(t) &= E[e^{itX}] = E\left[\sum_{m=0}^{\infty} i^m X^m \frac{t^m}{m!}\right] \\ &= \sum_{m=0}^{\infty} i^m E[X^m] \frac{t^m}{m!} = \sum_{m=0}^{\infty} i^m \mu_m(X) \frac{t^m}{m!}\end{aligned}$$

を **特性関数** (characteristic function) と呼ぶ。特性関数を扱う場合には $i = \sqrt{-1}$ としたいので、 i を番号の意味で使わないように気を付ける必要がある。

モーメント母関数だけではなく、特性関数もモーメント達の母関数になっている。

モーメント母関数の対数

$$K_X(t) = \log M(t) = \log E[e^{tX}] = \sum_{m=1}^{\infty} \kappa_m(X) \frac{t^m}{m!} = \sum_{m=1}^{\infty} \kappa_m \frac{t^m}{m!}$$

を **キュムラント母関数** (cumulant generating function, cgf) と呼び、その展開係数 $\kappa_m = \kappa_m(X)$ を X の m 次のキュムラントと呼ぶ。

注意: 取り得る値が実数になる確率変数 X について $|e^{itX}| = 1$ となるので、 $E[e^{itX}]$ は常に絶対収束しており、特性関数は常にうまく定義されている。それに対して e^{tX} の値は巨大になる可能性があり、 $E[e^{tX}]$ が収束しない場合が出て来る。モーメント母関数やキュムラント母関数の取り扱いではこの点に注意する必要がある。

注意: モーメント母関数とキュムラント母関数はそれぞれ物理での統計力学における **分配関数** と **自由エネルギー** (もしくは **Massieu函数**) の確率論的な類似物になっている。ただし、逆温度 β について $t = -\beta$ とおく必要がある。逆に言えば、モーメント母関数とキュムラントの表示における $\beta = -t$ の逆数は絶対温度の確率論的類似物になっていることになる。

4.2 特性関数による期待値の表示

X が確率密度函数 $p(x)$ を持つとき、函数 $f(x)$ のFourier変換を

$$\hat{f}(t) = \int_{-\infty}^{\infty} f(x) e^{-itx} dx$$

と書くと、 $f(t)$ がそう悪くない函数ならば逆Fourier変換によってもとの函数に戻せる:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(t) e^{itx} dt.$$

Fourier解析の基礎については次のリンク先を参照せよ(逆Fourier変換に関する結果はこのノート内では認めて使ってよい):

- [12 Fourier解析 \(https://nbviewer.org/github/genkuroki/Calculus/blob/master/12%20Fourier%20analysis.ipynb\)](https://nbviewer.org/github/genkuroki/Calculus/blob/master/12%20Fourier%20analysis.ipynb)

ゆえに、 x に確率変数 X を代入して両辺の期待値を取り、期待値を取る操作と積分を交換すると、

$$E[f(X)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(t) E[e^{itX}] dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(t) \varphi_X(t) dt.$$

ここで $\varphi_X(t) = E[e^{itX}]$ は X の特性関数である。

確率変数 X が従う分布は様々な函数 $f(x)$ に関する期待値 $E[f(X)]$ から決まるので、 $E[f(X)]$ が X の特性関数 $\varphi_X(t)$ を用いて表せたということは、 **X の特性関数 $\varphi_X(t)$ から X が従う分布が唯一に決まる** ことを意味している。

さらに、分布Aの特性関数が分布Bの特性関数で近似されていれば、分布Aにおける期待値が分布Bにおける期待値で近似されることもわかる。これは **分布の近似を特性関数の近似で確認できる** ことを意味する。

モーメント母関数 $M_X(t) = E[e^{tX}]$ が t の関数として、定義域を自然に複素数まで拡張できているとき(正確には解析接続できていれば)、 $\varphi_X(t) = M_X(it)$ が成立する。キウムラント母関数はモーメント母関数の対数である。これらより、**モーメント母関数やキウムラント母関数からも分布が唯一に決まる** ことがわかる。

4.3 問題: 分布のアフィン変換のキウムラント

確率変数 X と $a, b \in \mathbb{R}, a \neq 0$ について、 $aX + b$ のモーメント母関数とキウムラント母関数とキウムラントが X のそれらで次のように表されることを示せ:

$$M_{aX+b}(t) = e^{bt} M_X(at), \quad K_{aX+b}(t) = K_X(at) + bt, \\ \kappa_1(aX + b) = a\kappa_1(X) + b, \quad \kappa_m(aX + b) = a^m \kappa_m(X) \quad (m = 2, 3, 4, \dots)$$

注意: キウムラントの変換公式は非常に単純な形になる。 $\kappa_1(aX + b) = a\kappa_1(X) + b$ は $\kappa_1(X) = E[X]$ だったので当然である。2 次以上のキウムラントは a^m 倍されるだけになる。モーメント母関数の対数を取ってキウムラント母関数を定義し、その展開によってキウムラントを定義することにはこのような利点がある。キウムラント母関数を定義することには、物理の統計力学で分配関数の対数を取って自由エネルギーを定義することと同様の利点がある。

注意: この結果は空気のごとく使われる。

解答例: $aX + b$ のモーメント母関数を X のモーメント母関数であらわそう:

$$M_{aX+b}(t) = E[e^{t(aX+b)}] = e^{bt} E[e^{atX}] = e^{bt} M_X(at).$$

ゆえに、 $aX + b$ のキウムラントは次の形になる:

$$K_{aX+b}(t) = \log M_{aX+b}(t) = \log(e^{bt} M_X(at)) = K_X(at) + bt.$$

$X \mapsto aX + b$ によってキウムラント母関数は $K_X(t) \mapsto K_X(at) + bt$ に似た形式で変換される。

X のキウムラント $\kappa_m(X)$ は次のようにキウムラント母関数を展開することによって定義されるのであった:

$$K_X(t) = \sum_{m=1}^{\infty} \kappa_m(X) \frac{t^m}{m!}.$$

$K_{aX+b}(t)$ の展開結果は

$$K_{aX+b}(t) = K_X(at) + bt = \sum_{m=0}^{\infty} \kappa_m(X) \frac{(at)^m}{m!} + bt \\ = (a\kappa_1(X) + b)t + \sum_{m=2}^{\infty} a^m \kappa_m(X) \frac{t^m}{m!}$$

になるので、

$$\kappa_1(aX + b) = a\kappa_1(X) + b, \quad \kappa_m(aX + b) = a^m \kappa_m(X) \quad (m = 2, 3, 4, \dots)$$

となることがわかる。

解答終

4.4 問題: 標準正規分布のモーメント母関数と特性関数とキウムラント母関数

標準正規分布に従う確率変数 Z のモーメント母関数と特性関数とキウムラント母関数が次のようになることを示せ:

$$M_Z(t) = e^{t^2/2}, \quad \varphi_Z(t) = M_Z(it) = e^{-t^2/2}, \quad K_Z(t) = \log M_Z(t) = \frac{t^2}{2}.$$

この結果は中心極限定理の証明で使われる。

解答例: $Z \sim \text{Normal}(0, 1)$ と仮定する。このとき、

$$tz - \frac{z^2}{2} = -\frac{z^2 - 2tz}{2} = -\frac{(z-t)^2 - t^2}{2} = -\frac{(z-t)^2}{2} + \frac{t^2}{2}$$

なので、

$$\begin{aligned}
M_Z(t) &= E[e^{tZ}] = \int_{-\infty}^{\infty} e^{tz} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(z-t)^2/2+t^2/2} dz \\
&= \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-w^2/2} dw = e^{t^2/2}.
\end{aligned}$$

4つめの等号で $z = w + t$ とおいた。ゆえに、

$$\varphi_Z(t) = E[e^{itZ}] = M_Z(it) = e^{-t^2/2}, \quad K_Z(t) = \log M_Z(t) = \frac{t^2}{2}.$$

解答終

注意: 標準正規分布のキュムラント母関数は $t^2/2$ というたったの一項だけになってしまう。標準でない正規分布のキュムラント母関数は t について1次と2次の項だけになる。1次の項の係数は分布の期待値で、 $t^2/2$ の係数は分散になる。実際、平均 μ 、分散 σ^2 を持つ確率変数 X について、

$$\begin{aligned}
K_X(t) &= \log E[e^{tX}] = \log E[e^{t(X-\mu)+t\mu}] \\
&= t\mu + \log \left(1 + E[X-\mu]t + E[(X-\mu)^2] \frac{t^2}{2} + O(t^3) \right) \\
&= t\mu + \log \left(1 + \sigma^2 \frac{t^2}{2} + O(t^3) \right) = t\mu + \sigma^2 \frac{t^2}{2} + O(t^3).
\end{aligned}$$

そして、 $\sigma Z + \mu \sim \text{Normal}(\mu, \sigma)$ であり、

$$M_{\sigma Z + \mu}(t) = E[e^{t(\sigma Z + \mu)}] = e^{t\mu} E[e^{t\sigma Z}] = e^{t\mu} M_Z(\sigma t) = e^{t\mu + \sigma^2 t^2/2}.$$

ゆえに

$$K_{\sigma Z + \mu}(t) = \log M_{\sigma Z + \mu}(t) = t\mu + \sigma^2 \frac{t^2}{2}.$$

他の分布のキュムラント母関数を計算したときに出て来る t について3次以上の項はその分布が正規分布からどれだけ離れているかを表している。

4.5 確率変数の標準化と標準化キュムラントと歪度と尖度

確率変数 X は確率変数であるとし、 $\mu = E[X]$ 、 $\sigma = \sqrt{E[(X-\mu)^2]}$ とおく。このとき、

$$Z = \frac{X - \mu}{\sigma}$$

を確率変数の **標準化** (standardization) と呼ぶ。 Z の期待値と分散はそれぞれ 0 と 1 になる。

X の標準化のモーメントやキュムラントおよびそれらの母関数をそれぞれ **標準化モーメント**、**標準化キュムラント** と呼び、それぞれを $\bar{\mu}_m(X)$ 、 $\bar{\kappa}_m(X)$ と表す。詳しくは以下の通り:

$$\begin{aligned}
\bar{\mu}_m(X) &= \mu_m(Z) = E \left[\left(\frac{X - \mu}{\sigma} \right)^m \right], \\
M_Z(t) &= E \left[\exp \left(t \frac{X - \mu}{\sigma} \right) \right] = \sum_{m=0}^{\infty} \bar{\mu}_m(X) \frac{t^m}{m!} = 1 + \frac{t^2}{2} + \bar{\mu}_3(X) \frac{t^3}{3!} + \bar{\mu}_4(X) \frac{t^4}{4!} + \dots, \\
K_Z(t) &= \log M_Z(t) = \sum_{m=1}^{\infty} \bar{\kappa}_m(X) \frac{t^m}{m!} = \frac{t^2}{2} + \bar{\kappa}_3(X) \frac{t^3}{3!} + \bar{\kappa}_4(X) \frac{t^4}{4!} + \dots.
\end{aligned}$$

$\bar{\kappa}_3(X)$ と $\bar{\kappa}_4(X)$ は次のように表される:

$$\bar{\kappa}_3(X) = \bar{\mu}_3(X) = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right], \quad \bar{\kappa}_4(X) = \bar{\mu}_4(X) - 3 = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] - 3.$$

このことは、 $\log(1+a) = a - a^2/2 + O(a^3)$ を使って以下のようにして確認される:

$$\begin{aligned}
\log \left(1 + \frac{t^2}{2} + \bar{\mu}_3(X) \frac{t^3}{3!} + \bar{\mu}_4(X) \frac{t^4}{4!} + O(t^5) \right) &= \frac{t^2}{2} + \bar{\mu}_3(X) \frac{t^3}{3!} + \bar{\mu}_4(X) \frac{t^4}{4!} - \frac{1}{2} \left(\frac{t^2}{2} \right)^2 + O(t^5) \\
&= \frac{t^2}{2} + \bar{\mu}_3(X) \frac{t^3}{3!} + (\bar{\mu}_4(X) - 3) \frac{t^4}{4!} + O(t^5).
\end{aligned}$$

$\bar{\kappa}_3(X)$ を X もしくは X が従う分布の **歪度** (わいど, skewness) と呼び, $\bar{\kappa}_4(X)$ を **尖度** (せんど, kurtosis) と呼び, 次のようにも書くことにする:

$$\text{skewness}(X) = \bar{\kappa}_3(X) = E \left[\left(\frac{X - \mu}{\sigma} \right)^3 \right], \quad \text{kurtosis}(X) = \bar{\kappa}_4(X) = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] - 3.$$

歪度は左右の非対称性の尺度であり, 尖度は分布の尖り具合が正規分布とどれだけ違うかの尺度になっている.

X が正規分布に従う確率変数の場合にはその標準化 $Z = (X - \mu)/\sigma$ は標準正規分布に従う確率変数になるので, その標準化キウムラント達は $\bar{\kappa}_2(X) = 1, \bar{\kappa}_m(X) = 0 (m \neq 0)$ となる. 2次の標準化キウムラントは常に 1 になるが, 3 次以上の標準化キウムラントは X が正規分布でなければ 0 でなくなる.

このことから, 3 次以上の標準化キウムラントは分布が正規分布からどれだけ離れているかを表していると考えられる. それらのうち最初の2つが上で定義した歪度 $\bar{\kappa}_3(X)$ と尖度 $\bar{\kappa}_4(X)$ になっている.

分布の歪度 $\bar{\kappa}_3(X)$ と尖度 $\bar{\kappa}_4(X)$ は分布がどれだけ正規分布から離れているかを表す最も基本的な量である.

注意: $\bar{\kappa}_4(X) = \bar{\mu}_4(X) - 3$ ではなく, 3 を引く前の $\bar{\mu}_4(X)$ を尖度と定義する流儀もあるが, このノートでは正規分布の扱いでキウムラントが非常に便利なことを重視したいので, 3 を引いた側の $\bar{\kappa}_4$ を尖度の定義として採用する. 3 を引いた方の $\bar{\kappa}_4(X) = \bar{\mu}_4(X) - 3$ は **過剰尖度** (かじょうせんど, **excess kurtosis**) と呼ばれることも多い. 正規分布の尖度を ($\bar{\kappa}_4$ の方の 0 ではなく) $\bar{\mu}_4$ の方の 3 とするときに, そこからどれだけ分布の尖り具合が増したかを $\bar{\kappa}(X)$ が表していることを「過剰」と表現している.

In [34]:

```
1 @vars t μ3 μ4 μ5 κ3 κ4 κ5
2 Mt = 1 + t^2/2 + μ3*t^3/6 + μ4*t^4/24 + μ5*t^5
3 expr = series(log(Mt), t)
```

Out[34]: $\frac{t^2}{2} + \frac{t^3 \mu_3}{6} + t^4 \left(\frac{\mu_4}{24} - \frac{1}{8} \right) + t^5 \left(-\frac{\mu_3}{12} + \mu_5 \right) + O(t^6)$

4.6 問題: 正規分布の歪度と尖度

X が正規分布に従う確率変数であるとき, その歪度 $\bar{\kappa}_3(X)$ と尖度 $\bar{\kappa}_4(X)$ が以下になることを示せ:

$$\bar{\kappa}_3(X) = \bar{\kappa}_4(X) = 0.$$

解答例: X が正規分布 $\text{Normal}(\mu, \sigma)$ に従う確率変数のとき, その標準化 $Z = (X - \mu)/\sigma$ は標準正規分布に従うので, $Z = (X - \mu)/\sigma$ のキウムラント母関数は $t^2/2$ になり, その 3 次以上の項がすべて 0 になる. ゆえに $\bar{\kappa}_m(X) = 0 (m = 3, 4, 5, \dots)$. 特に $\bar{\kappa}_3(X) = \bar{\kappa}_4(X) = 0$.

解答終

4.7 問題: 一様分布のキウムラント母関数と歪度と尖度

一様分布 $\text{Uniform}(0, 1)$ に従う確率変数 X について, そのキウムラント母関数 $K_X(t) = \log E[e^{tX}]$ と歪度 $\bar{\kappa}_3(X)$ と尖度 $\bar{\kappa}_4(X)$ が以下になることを示せ:

$$K_X(t) = \log \frac{e^t - 1}{t}, \quad \bar{\kappa}_3(X) = 0, \quad \bar{\kappa}_4(X) = -\frac{6}{5}.$$

一様分布は歪度が 0 でかつ尖度がかなり小さな連続分布の例になっている.

解答例: モーメント母関数とキウムラント母関数は以下のように計算される:

$$E[e^{tX}] = \int_0^1 e^{tx} dx = \left[\frac{e^{tx}}{t} \right]_0^1 = \frac{e^t - 1}{t},$$

$$K_X(t) = \log \frac{e^t - 1}{t} = \frac{1}{2}t + \frac{1}{12} \frac{t^2}{2} - \frac{1}{120} \frac{t^4}{4!} + O(t^5).$$

ゆえに, $\mu = 1/2, \sigma = \sqrt{1/12}, Z = (X - \mu)/\sigma$ とおくと,

$$K_Z(t) = K_X\left(\frac{t}{\sigma}\right) - \frac{\mu}{\sigma}t = \frac{t^2}{2} - \frac{6}{5} \frac{t^4}{4!} + O(t^5).$$

ゆえに $\bar{\kappa}_3(X) = 0, \bar{\kappa}_4 = -6/5$.

解答終

```
In [35]: 1 @vars t
2 K_X = log((exp(t) - 1)/t)
3 κ = [limit(diff(K_X, t, m) , t, 0) for m in 1:4]
```

```
Out[35]: [ 1/2
           1/12
           0
          -1/120]
```

```
In [36]: 1 μ, σ = κ[1], √κ[2]
2 κ = [limit(diff(K_X(t⇒t/σ) - μ/σ*t, t, m) , t, 0) for m in 1:4]
```

```
Out[36]: [ 0
           1
           0
          -6/5]
```

4.8 問題: 独立な確率変数達の和のモーメント母関数と特性函数とキュムラント母関数

独立な確率変数達 X_1, \dots, X_n の和のモーメント母関数と特性函数とキュムラント母関数が次のように表されることを示せ:

$$\begin{aligned} M_{X_1+\dots+X_n}(t) &= M_{X_1}(t) \cdots M_{X_n}(t), \\ \varphi_{X_1+\dots+X_n}(t) &= \varphi_{X_1}(t) \cdots \varphi_{X_n}(t), \\ K_{X_1+\dots+X_n}(t) &= K_{X_1}(t) + \cdots + K_{X_n}(t). \end{aligned}$$

注意: この結果は空気のごとく使われる.

解答例: 独立な確率変数達 X_1, \dots, X_n について

$$E[f_1(X_1) \cdots f_n(X_n)] = E[f_1(X_1)] \cdots E[f_n(X_n)]$$

が成立することより,

$$\begin{aligned} M_{X_1+\dots+X_n}(t) &= E[e^{t(X_1+\dots+X_n)}] = E[e^{tX_1} \cdots e^{tX_n}] = M_{X_1}(t) \cdots M_{X_n}(t), \\ \varphi_{X_1+\dots+X_n}(t) &= E[e^{it(X_1+\dots+X_n)}] = E[e^{itX_1} \cdots e^{itX_n}] = \varphi_{X_1}(t) \cdots \varphi_{X_n}(t). \end{aligned}$$

ゆえに

$$\begin{aligned} K_{X_1+\dots+X_n}(t) &= \log M_{X_1+\dots+X_n}(t) = \log(M_{X_1}(t) \cdots M_{X_n}(t)) \\ &= \log M_{X_1}(t) + \cdots + \log M_{X_n}(t) = K_{X_1}(t) + \cdots + K_{X_n}(t). \end{aligned}$$

解答終

4.9 問題: ガンマ分布のキュムラント母関数と歪度と尖度

ガンマ分布 $\text{Gamma}(\alpha, \theta)$ に従う確率変数 X について, そのキュムラント母関数 $K_X(t) = \log E[e^{tX}]$ と歪度 $\bar{\kappa}_3(X)$ と尖度 $\bar{\kappa}_4(X)$ が以下になることを示せ:

$$K_X(t) = -\alpha \log(1 - \theta t), \quad \bar{\kappa}_3(X) = \frac{2}{\sqrt{\alpha}}, \quad \bar{\kappa}_4(X) = \frac{6}{\alpha}.$$

解答例: $t < 1/\theta$ と仮定する. このとき,

$$\mu = \alpha\theta, \quad \sigma = \sqrt{\alpha}\theta, \quad Z = \frac{X - \mu}{\sigma}$$

とおくと,

$$\begin{aligned}
E[e^{tX}] &= \int_0^\infty e^{tx} \frac{e^{-x/\theta} x^{\alpha-1}}{\theta^\alpha \Gamma(\alpha)} dx = \frac{1}{\theta^\alpha \Gamma(\alpha)} \int_0^\infty e^{-((1-\theta t)/\theta)x} x^{\alpha-1} dx \\
&= \frac{1}{\theta^\alpha \Gamma(\alpha)} \frac{\theta^\alpha}{(1-\theta t)^\alpha} \Gamma(\alpha) = \frac{1}{(1-\theta t)^\alpha}, \\
K_X(t) &= -\alpha \log(1-\theta t) \\
&= \alpha \theta t + \alpha \theta^2 \frac{t^2}{2} + \alpha \theta^3 \frac{t^3}{3} + \alpha \theta^4 \frac{t^4}{4} + O(t^5) \\
&= \mu t + \sigma^2 \frac{t^2}{2} + \frac{2\sigma^3}{\sqrt{\alpha}} \frac{t^3}{3!} + \frac{6\sigma^4}{\alpha} \frac{t^4}{4!} + O(t^5), \\
K_Z(t) &= K_X\left(\frac{t}{\sigma}\right) - \frac{\mu}{\sigma} t = \frac{t^2}{2} + \frac{2}{\sqrt{\alpha}} \frac{t^3}{3!} + \frac{6}{\alpha} \frac{t^4}{4!} + O(t^5).
\end{aligned}$$

ゆえに,

$$\bar{\kappa}_3(X) = \frac{2}{\sqrt{\alpha}}, \quad \bar{\kappa}_3(X) = \frac{6}{\alpha}.$$

解答終

```
In [37]: 1 @vars t
2 @vars α θ positive=true
3 K_X = -α*log(1 - θ*t)
4 κ = [diff(K_X, t, m)(t⇒0).simplify().factor() for m in 1:4]
```

```
Out[37]: [ αθ
           αθ²
          2αθ³
          6αθ⁴ ]
```

```
In [38]: 1 μ, σ = κ[1], √κ[2]
2 [diff(K_X(t⇒t/σ) - μ/σ*t, t, m)(t⇒0).simplify().factor() for m in 1:4]
```

```
Out[38]: [ 0
           1
          2/√α
           6/α ]
```

4.10 問題: 二項分布のキウムラント母関数と歪度と尖度

二項分布 $\text{Binomial}(n, p)$ に従う確率変数 X について, そのキウムラント母関数 $K_X(t) = \log E[e^{tX}]$ と歪度 $\bar{\kappa}_3(X)$ と尖度 $\bar{\kappa}_4(X)$ が以下のようになることを示せ:

$$K_X(t) = n \log(1 + p(e^t - 1)), \quad \bar{\kappa}_3(X) = \frac{1 - 2p}{\sqrt{np(1-p)}}, \quad \bar{\kappa}_4(X) = \frac{1 - 6p(1-p)}{np(1-p)}.$$

注意: $n = 1, p = 1/2$ のとき $\bar{\kappa}_4(X) = -2$ となる. 実はその場合の分布 $\text{Bernoulli}(1/2)$ が確率分布全体の中で最小の尖度を持つ分布になっている.

解答例: X_1, \dots, X_n をそれぞれが $\text{Bernoulli}(p)$ に従う独立同分布な確率変数達であるとする. このとき, 二項分布に従う確率変数は $X = \sum_{i=1}^n X_i$ と作れる. ゆえに

$$K_X(t) = \sum_{i=1}^n K_{X_i}(t) = nK_{X_1}(t).$$

$X_1 \sim \text{Bernoulli}(p)$ について,

$$\begin{aligned}
E[e^{tX_1}] &= e^{t^1} p + e^{t^0} (1-p) = 1 + p(e^t - 1), \\
K_{X_1}(t) &= \log(1 + p(e^t - 1)) \\
K_X(t) &= n \log(1 + p(e^t - 1))
\end{aligned}$$

なので,

$$\mu = np, \quad \sigma = \sqrt{np(1-p)}$$

とおくと,

$$\begin{aligned} K_X(t) &= n \log(1 + p(e^t - 1)) \\ &= npt + np(1-p)\frac{t^2}{2} + np(1-p)(1-2p)\frac{t^3}{3!} + np(1-p)(6p^2 - 6p + 1)\frac{t^4}{4!} + O(t^5) \\ &= \mu t + \sigma^2 \frac{t^2}{2} + \sigma^2(1-2p)\frac{t^3}{3!} + \sigma^2(1-6p(1-p))\frac{t^4}{4!} + O(t^5), \\ K_{(X-\mu)/\sigma}(t) &= K_X\left(\frac{t}{\sigma}\right) - \frac{\mu}{\sigma}t = \frac{t^2}{2} + \frac{1-2p}{\sigma} \frac{t^3}{3!} + \frac{1-6p(1-p)}{\sigma^2} \frac{t^4}{4!} + O(t^5). \end{aligned}$$

ゆえに

$$\bar{\kappa}_3(X) = \frac{1-2p}{\sigma} = \frac{1-2p}{\sqrt{np(1-p)}}, \quad \bar{\kappa}_4(X) = \frac{1-6p(1-p)}{\sigma^2} = \frac{1-6p(1-p)}{np(1-p)}.$$

解答終

In [39]: `1 kurtosis(Bernoulli(0.5))`

Out[39]: -2.0

In [40]: `1 @vars n p t
2 K_X = n * log(1 + p*(exp(t) - 1))
3 κ = [diff(K_X, t, m)(t⇒0).simplify().factor() for m in 1:4]`

Out[40]:
$$\begin{bmatrix} np \\ -np(p-1) \\ np(p-1)(2p-1) \\ -np(p-1)(6p^2-6p+1) \end{bmatrix}$$

In [41]: `1 μ, σ = κ[1], √κ[2]
2 [diff(K_X(t⇒t/σ) - μ/σ*t, t, m)(t⇒0).simplify().factor() for m in 1:4]`

Out[41]:
$$\begin{bmatrix} 0 \\ 1 \\ -\frac{2p-1}{\sqrt{-np(p-1)}} \\ -\frac{6p^2-6p+1}{np(p-1)} \end{bmatrix}$$

5 独立同分布な確率変数達の不偏分散の分散

以下の節においては, X_1, \dots, X_n は独立同分布な確率変数達であるとし, それらの標本分散と不偏分散を以下のように書くことにする:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad S_X^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

さらに, X_i の期待値と分散をそれぞれ μ, σ^2 と書くことにする.

このとき, 以下が成立することはすでに示している:

$$E[\bar{X}] = \mu, \quad \text{var}(X) = E[(\bar{X} - \mu)^2] = \frac{\sigma^2}{n}, \quad E[S_X^2] = \sigma^2.$$

さらに以下では, X_i の **歪度** (わいど, skewness) と **尖度** (せんど, kurtosis) をそれぞれ $\bar{\kappa}_3 = \bar{\kappa}_3(X_i)$, $\bar{\kappa}_4 = \bar{\kappa}_4(X_i)$ と書くことにする:

$$\bar{\kappa}_3 = E \left[\left(\frac{X_i - \mu}{\sigma} \right)^3 \right], \quad \bar{\kappa}_4 = E \left[\left(\frac{X_i - \mu}{\sigma} \right)^4 \right] - 3.$$

標本平均と不偏分散の共分散と分散は X_i の歪度 $\bar{\kappa}_3$ と尖度 $\bar{\kappa}_4$ を使って次のように書ける:

$$\text{cov}(\bar{X}, S_X^2) = \frac{\bar{\kappa}_3}{n} \sigma^3, \quad \text{var}(S_X^2) = \frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}.$$

以下の節ではこの結果を証明する.

5.1 期待値 0, 分散 1 の場合への帰着

$Z = (X - \mu)/\sigma$ とおくと, Z の期待値と分散はそれぞれ 0 と 1 になる.

$Z_i = (X_i - \mu)/\sigma$ とおくと, Z_1, \dots, Z_n はそれぞれが Z と同じ分布に従う独立同分布な確率変数達になる. そして, $X_i = \sigma Z_i + \mu$ より,

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i, \quad S_Z^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2$$

とおくと,

$$\bar{X} = \sigma \bar{Z} + \mu, \quad S_X^2 = \sigma^2 S_Z^2$$

なので,

$$\text{cov}(\bar{X}, S_X^2) = \sigma^3 \text{cov}(\bar{Z}, S_Z^2), \quad \text{var}(S_X^2) = \sigma^4 \text{var}(S_Z^2).$$

これより, 標本分散と不偏分散の分散と共分散の計算は $\mu = 0, \sigma = 1$ の場合に帰着する.

5.2 標本平均と不偏分散の共分散の計算

標本平均と不偏分散の共分散は X_i の歪度 $\bar{\kappa}_3$ を使って次のように書ける:

$$\text{cov}(\bar{X}, S_X^2) = \frac{\bar{\kappa}_3}{n} \sigma^3.$$

証明: $\mu = 0, \sigma = 1$ と仮定する. このとき, $\mu = 0$ という仮定より,

$$\text{cov}(\bar{X}, S_X^2) = E[\bar{X}(S_X^2 - \sigma^2)] = E[\bar{X}S_X^2] - \underbrace{\sigma^2 E[\bar{X}]}_{=0} = E[\bar{X}S_X^2].$$

そして,

$$\begin{aligned} n\bar{X}^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 + \frac{2}{n} \sum_{1 \leq i < j \leq n} X_i X_j, \\ (n-1)S_X^2 &= \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i^2 - 2\bar{X}X_i + \bar{X}^2) = \sum_{i=1}^n X_i^2 - n\bar{X}^2 \\ &= \sum_{i=1}^n X_i^2 - \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{2}{n} \sum_{1 \leq i < j \leq n} X_i X_j = \frac{n-1}{n} \sum_{i=1}^n X_i^2 - \frac{2}{n} \sum_{1 \leq i < j \leq n} X_i X_j, \\ \therefore S_X^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} X_i X_j. \end{aligned}$$

ゆえに

$$\bar{X}S_X^2 = \frac{1}{n} \sum_{i=1}^n X_i \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} X_i X_j \right).$$

$E[X_i] = \mu = 0$ という仮定と X_i 達の独立性より, i, j, k がすべて等しいとき以外には $E[X_i X_j X_k] = 0$ となり, $\mu = 0, \sigma = 1$ という仮定より $E[X_i^3] = \bar{\kappa}_3$ となるので,

$$\text{cov}(\bar{X}, S_X^2) = E[\bar{X}S_X^2] = \frac{1}{n^2} \sum_{i=1}^n E[X_i^3] = \frac{\bar{\kappa}_3}{n}.$$

$\mu = 0, \sigma = 1$ と仮定しない一般の場合には, これが σ^3 倍されて,

$$\text{cov}(\bar{X}, S_X^2) = \frac{\bar{\kappa}_3}{n} \sigma^3.$$

証明終

5.3 不偏分散の分散の計算

不偏分散の分散は X_i の尖度 $\bar{\kappa}_4$ を使って次のように書ける:

$$\text{var}(S_X^2) = \left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1} \right) \sigma^4.$$

証明: $\mu = 0, \sigma = 1$ と仮定する. このとき, $E[S_X^2] = \sigma^2 = 1$ より,

$$\text{var}(S_X^2) = E[(S_X^2)^2] - E[S_X^2]^2 = E[(S_X^2)^2] - 1.$$

そして, 前節の計算より

$$S_X^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} X_i X_j$$

なので,

$$\begin{aligned} (S_X^2)^2 &= \frac{1}{n^2} \left(\sum_{i=1}^n X_i^4 + 2 \sum_{1 \leq i < j \leq n} X_i^2 X_j^2 \right) \\ &\quad - \frac{2}{n^2(n-1)} \sum_{i=1}^n \sum_{1 \leq j < k \leq n} X_i^2 X_j X_k \\ &\quad + \frac{4}{n^2(n-1)^2} \sum_{1 \leq i < j \leq n} \sum_{1 \leq k < l \leq n} X_i X_j X_k X_l. \end{aligned}$$

$\mu = 0, \sigma = 1$ でかつ X_1, \dots, X_n は独立であるという仮定と, $E[X_i] = 0, E[X_i^2] = 1, E[X_i^4] = \bar{\kappa}_4 + 3$ であることを使うと,

$$\begin{aligned} E \left[\sum_{i=1}^n X_i^4 \right] &= n(\bar{\kappa}_4 + 3), \\ E \left[2 \sum_{1 \leq i < j \leq n} X_i^2 X_j^2 \right] &= 2 \frac{n(n-1)}{2} = n(n-1), \\ E \left[\sum_{1 \leq j < k \leq n} X_i^2 X_j X_k \right] &= 0, \\ E \left[\sum_{1 \leq i < j \leq n} \sum_{1 \leq k < l \leq n} X_i X_j X_k X_l \right] &= \frac{n(n-1)}{2}. \end{aligned}$$

最後の等式は $i = k, j = l$ の場合のみが和に寄与することを使った. ゆえに

$$\begin{aligned} \text{var}(S_X^2) &= \frac{n(\bar{\kappa}_4 + 3) + n(n-1)}{n^2} + \frac{4}{n^2(n-1)^2} \frac{n(n-1)}{2} - 1 \\ &= \frac{\bar{\kappa}_4 + n + 2}{n} + \frac{2}{n(n-1)} - 1 \\ &= \frac{\bar{\kappa}_4 + n + 2}{n} + \frac{2}{n-1} - \frac{2}{n} - \frac{n}{n} \\ &= \frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}. \end{aligned}$$

$\mu = 0, \sigma = 1$ と仮定しない一般の場合には, これが σ^4 倍されて,

$$\text{var}(S_X^2) = \left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1} \right) \sigma^4.$$

証明終

5.4 歪度と尖度に関する不等式

以上の結果を用いて, 歪度 $\bar{\kappa}_3$ と尖度 $\bar{\kappa}_4$ のあいだには

$$\bar{\kappa}_4 \geq \bar{\kappa}_3^2 - 2 \geq -2$$

が成立することを示そう. 標本平均 \bar{X} と不偏分散 S_X^2 の分散共分散行列

$$\begin{bmatrix} \text{var}(\bar{X}) & \text{cov}(\bar{X}, S_X^2) \\ \text{cov}(S_X^2, \bar{X}) & \text{var}(S_X^2) \end{bmatrix} = \begin{bmatrix} \sigma^2/n & \sigma^3 \bar{\kappa}_3/n \\ \sigma^3 \bar{\kappa}_3/n & \sigma^4 (\bar{\kappa}_4/n + 2/(n-1)) \end{bmatrix}$$

は半正定値になる. 特にその行列式は 0 以上になる:

$$\frac{\sigma^6}{n^2} \left(\bar{\kappa}_4 + \frac{2}{1 - 1/n} - \bar{\kappa}_3^2 \right) \geq 0.$$

ゆえに,

$$\bar{\kappa}_4 \geq \bar{\kappa}_3^2 - \frac{2}{1 - 1/n}.$$

$n \rightarrow \infty$ とすることによって,

$$\bar{\kappa}_4 \geq \bar{\kappa}_3^2 - 2.$$

を得る.

6 正規分布の標本分布の場合

In []:

1	
---	--