

標本分布について

- 黒木玄
- 2022-04-11~2022-04-12

このノートではJulia言語 (<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 確率変数の独立性と標本分布](#)
 - [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](#)
 - [1.3 独立同分布の定義](#)
 - [1.4 標本分布の定義](#)
 - [1.5 試行回数 \$n\$ のBernoulli試行の分布はBernoulli分布の標本分布](#)
 - [1.6 二項分布による推定の確率的揺らぎの記述](#)
- ▼ [1.7 問題: 大阪都構想に関する住民投票の結果について](#)
 - [1.7.1 Julia言語による計算の例](#)
 - [1.7.2 注意: Clopper-Pearsonの信頼区間とそれを与えるP値](#)
 - [1.7.3 WolframAlphaによる計算の例:](#)
 - [1.8 互いに異なる任意の2つが独立であっても全体が独立であるとは限らない](#)
 - [1.9 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 無相関性](#)
 - [2.1 共分散と相関係数の定義および無相関の定義](#)
 - [2.2 問題: 独立ならば無相関である \(実質1行で解ける\)](#)
 - [2.3 問題: 無相関でも独立とは限らない](#)
 - [2.4 問題: 無相関な確率変数達の和の分散](#)
 - [2.5 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](#)
 - [2.6 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](#)

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

1 確率変数の独立性と標本分布

1.1 同時確率質量関数と同時確率密度関数

複数の確率変数 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、 $P(x_1, \dots, x_n) \geq 0$ であり、 $P(x_1, \dots, x_n)$ の総和が 1 にあり、

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

が成立することである。このとき、 X_1 単独の確率質量関数 $P(x_1)$ は

$$P(x_1) = \sum_{x_2} \cdots \sum_{x_n} P(x_1, \dots, x_n)$$

になる. なぜならば,

複数の確率変数 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, $p(x_1, \dots, x_n) \geq 0$ であり, $p(x_1, \dots, x_n)$ 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.

X_1

1.2 確率変数の独立性の定義

複数の確率変数 X_1, \dots, X_n が同時に与えられており, 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,

$$E[f(X_1) \cdots f(X_n)] = E[f(X_1)] \cdots E[f(X_n)]$$

と X_i ごと関数 $f(X_i)$ の積の期待値が各々の $f(X_i)$ 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.)

上の条件は, 例えば, X_i 達のそれぞれが確率密度関数 $p_i(x_i)$ を持つとき, (X_1, \dots, X_n) の同時確率密度関数 $p(x_1, \dots, x_n)$ が $p(x_1) \cdots p(x_n)$ と積の形になること

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

と同値だと考えてよい. すなわち, (X_1, \dots, X_n) の関数の期待値が次のように書かれることと同値だと思ってよい:

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

ここで $\int \cdots \int$ は適切な範囲での定積分を表す.

X_i 達のそれぞれが確率質量関数 $P_i(x_i)$ を持つ場合には, X_1, \dots, X_n の独立性は

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

が成立することと同値である.

確率密度関数と確率質量関数が混ざっている場合も同様である.

大雑把に誤解を恐れずに言うと, 確率が積になるときに独立だという.

実際の計算では, 積の期待値が期待値の積になるという形式で確率変数の独立性を使う.

1.3 独立同分布の定義

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 達が従う分布が等しいことであると定める.

1.4 標本分布の定義

独立同分布な n 個の確率変数達 X_1, \dots, X_n の同時確率分布を各 X_k の分布の **サイズ n の標本分布** (サンプル分布, sample distribution) と呼ぶ.

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 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

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

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。この専門用語の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。

1.6 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、未知の当たりの確率は k/n である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 の推定の仕方のモデル化になっている。

$\hat{p} = K/n$ 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:

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

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 $\text{std}(\hat{p})$ の値を推定するためにはここで使った p に \hat{p} を代入して得られる公式

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

を使えばよいだろう。

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。

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

$n = 1000, p = 0.3$

Legend: $\hat{p} = K/n$ (blue bars), $p = 0.3$ (red line)

X-axis: \hat{p} (0.24 to 0.36)

Y-axis: Frequency (0 to 25)

現実の統計分析ではこのような未知の確率 $p = 0.3$ に当たるものが使えないので、このようなグラフを描くことはできない。真つ暗闇の中を手探りで進むような感じになる。

解答例:

(1) $P(K \leq 694844) \approx 5.652\text{e-}20$

(2) $P(K \leq 675829) \approx 4.844\text{e-}49$

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

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

解答終

1.7.1 Julia言語による計算の例

```
In [3]: 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函数の対数である.  
3 logP(n, p, k) = logabsbinomial(n, k)[1] + k*log(p) + (n-k)*log(1-p)
```

Out[3]: logP (generic function with 1 method)

```
In [4]: 1 # (1)  
2 @show exp(logsumexp(logP(694844 + 705585, 0.5, k) for k in 0:694844))  
3 @show cdf(Binomial(694844 + 705585, 0.5), 694844);  
  
exp(logsumexp((logP(694844 + 705585, 0.5, k) for k = 0:694844))) = 5.65211286550343e-20  
cdf(Binomial(694844 + 705585, 0.5), 694844) = 5.652112867175311e-20
```

```
In [5]: 1 # (2)  
2 @show exp(logsumexp(logP(675829 + 692996, 0.5, k) for k in 0:675829))  
3 @show cdf(Binomial(675829 + 692996, 0.5), 675829);  
  
exp(logsumexp((logP(675829 + 692996, 0.5, k) for k = 0:675829))) = 4.843721062457042e-49  
cdf(Binomial(675829 + 692996, 0.5), 675829) = 4.8437210657569034e-49
```

```
In [6]: 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 [7]: 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 [8]: 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 [9]: 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 [10]: 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 [11]: 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
```

```
In [12]: 1 @doc beta_inc_inv
```

Out[12]: beta_inc_inv(a, b, p, q=1-p)

Return a tuple $(x, 1-x)$ where x satisfies $I_x(a, b) = p$, i.e., x is the inverse of the regularized incomplete beta function $I_x(a, b)$.

See also: [beta_inc_inv](#) (@ref).

```
In [13]: 1 @doc beta_inc
```

Out[13]: beta_inc(a, b, x, y=1-x)

Return a tuple $(I_x(a, b), 1 - I_x(a, b))$ where $I_x(a, b)$ is the regularized incomplete beta function given by

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt,$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$.

External links: [DLMF \(https://dlmf.nist.gov/8.17.1\)](https://dlmf.nist.gov/8.17.1), [Wikipedia \(https://en.wikipedia.org/wiki/Beta_function#Incomplete_beta_function\)](https://en.wikipedia.org/wiki/Beta_function#Incomplete_beta_function)

See also: [beta_inc_inv](#) (@ref).

1.7.2 注意: 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)で計算した値は **片側検定のP値** の一種である。それらの値の2倍は **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不完全ベータ関数になっていることから、その逆関数を直接使ってもよい:

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

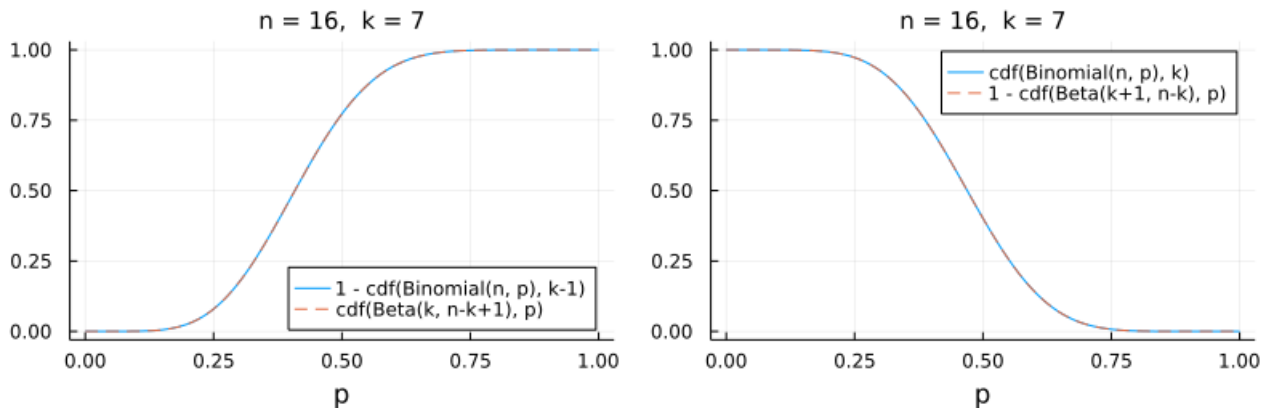
Wolfram言語では `InverseBetaRegularized` を使う。

Clopper-Pearsonの信頼区間を使うことのメリットは、二項分布での確率を和で計算して方程式を解くという面倒な計算を経由せずに、基本特殊関数の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Sternの信頼区間との比較で劣っていることは、Clopper-Pearsonの信頼区間の方が無駄に広くなってしまう場合が多いことである。 np が大きな場合にはどちらを使っても実践的な差は出ない。

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)
```

Out[14]:



1.7.3 WolframAlphaによる計算の例:

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

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

(3) p_L : `solve cdf(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: `InverseBetaRegularized(1/40, 694844, 705586)` (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%281%2F40%2C+694844%2C+705586%29>). もしくは、`InverseBetaRegularized(0.025, 694844, 705585 + 1)` (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%280.025%2C+694844%2C+705585+1%29>).

p_U : `solve cdf(BinomialDistribution(694844 + 705585, p), 694844) = 0.025` (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%280.025%2C+694844%2C+705585%29>). もしくは、`InverseBetaRegularized(1 - 0.025, 694844 + 1, 705585)` (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%281-0.025%2C+694844+1%2C+705585%29>).

(4) p_L : `solve cdf(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: `InverseBetaRegularized(1/40, 694844, 705586)` (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%281%2F40%2C+694844%2C+705586%29>). もしくは、`InverseBetaRegularized(0.025, 675829, 692996 + 1)` (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%280.025%2C+675829%2C+692996+1%29>).

p_U : `solve cdf(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、`InverseBetaRegularized(1 - 0.025, 675829 + 1, 692996)` (<https://www.wolframalpha.com/input?i=InverseBetaRegularized%281-0.025%2C+675829+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。

1.8 互いに異なる任意の2つが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2つは独立だが、 X, Y, Z の全体は独立ではない。

1.9 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Bを与えた。
- (2) $P(y) = 1/2$ の解釈: $Y = 1$ は薬に効果があったことを、 $Y = 0$ は効果が無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。
- (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。
- (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

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

は $x = 1$ の薬Aの場合も $x = 0$ の薬Bの場合も男女全体を見ると半々で効果があり、男女全体では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

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

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

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

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の表

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

より、 $x = 1$ の薬Aの場合には $y = 1$ の確率が正で効果があるが、 $x = 0$ の薬Bの場合には $y = 1$ の確率が0で効果がないことがわかる。 $z = 0$ の男性の場合に制限した確率質量関数の表

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

より、 $x = 1$ の薬Aの場合には $y = 1$ の確率が0で効果がないが、 $x = 0$ の薬Bの場合には $y = 1$ の確率が正で効果があることがわかる。

このように確率変数達が独立か否かは現実において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」と結論すると大変なことになってしまう場合がある。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 無相関性

2.1 共分散と相関係数の定義および無相関の定義

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。

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。

X, Y の共分散が $\text{cov}(X, Y) = 0$ のとき(この条件は相関係数が $\text{cor}(X, Y) = 0$ となることと同値、直観的には「直交している」と考える)、確率変数達 X, Y は **無相関** であるという。

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。

2.2 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

2.3 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: 確率質量函数 $P(x, y)$ 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 は同時確率質量函数 $P(x, y)$ 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

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

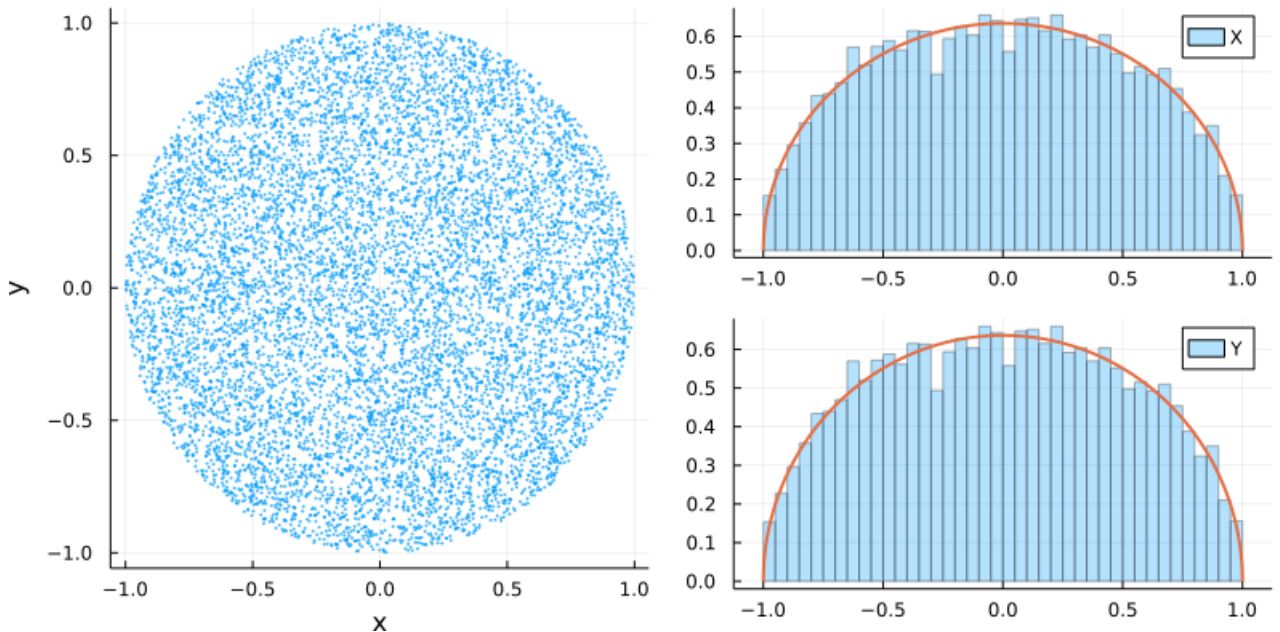
となり、同様にして $p(y) = (2/\pi)\sqrt{1-y^2}$ となるが、 $-1 < x, y < 1$ かつ $x^2 + y^2 > 1$ のとき、 $p(x, y) = 0$ となるが、 $p(x)p(y) \neq 0$ となるので、それらは等しくない。ゆえに X, Y は独立ではない。

解答終

```
In [15]: 1 n = 10^4
2 XY = [(r = sqrt(rand())); t = 2*pi*rand(); (r*cos(t), r*sin(t))] for _ in range(1:n)
3 X, Y = first.(XY), last.(XY)
4 @show cov(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

Out[15]:



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。

2.4 問題: 無相関な確率変数達の和の分散

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:

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

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、 $(v_i, v_j) = \|v_i\|^2 \delta_{ij}$ 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 でそれ以外の場合 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

2.5 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に求めてみよ。

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 が出るまでに出た 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!

2.6 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 $u_1, v_1, \dots, u_n, v_n$ 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 $(\ , \)$ について

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

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

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

の成立を意味する.

解答終

In []:

1