数理統計学01 Bernoulli試行と関連確率分布

二項分布, 幾何分布, 負の二項分布, など

(確率)分布にしたがう確率変数 X とその函数として得られる確率変数

確率変数 = ランダムに値が決まる量のモデル化

現実　ランダムに値が決まると考えられる量

- サイコロやルーレットの出目
- ギャンブルで得られる賞金の額
- 無作為抽出で得られた標本(サンプル, データ)

↓

数学的モデル化　ある(確率)分布にしたがう確率変数

データのモデル化として確率変数達の組 $(X_1, \dots, X_n)$ を使う.

数学的には「確率変数はその函数の期待値が定義されているもの」と思ってよい.

X

$E[f(X)]$ と書く　←　$f(X)$ も確率変数

例 (一様乱数生成のくりかえし) コンピュータでの rand() のくりかえし,

rand() は確率変数でモデル化される.

rand() のくりかえしの結果は一様分布の標本(サンプル)の例になっている.

一様乱数の生成の繰り返し

確率変数 の理解に向けて, コンピュータでの `rand()` 函数について説明する.

0以上1未満の区間上の一様分布の(擬似)乱数(以下ではこれを **一様乱数** と短く呼ぶことにする)を10個生成すると以下のようになる.

```
[rand() for _ in 1:10]
```

```
10-element Vector{Float64}:
 0.5105030210947423
 0.13229885279256504
 0.05873542170272006
 0.742812017444488
 0.8417141593154808
 0.44238748406950934
 0.5002487759410339
 0.7049642022880609
 0.6687158029467306
 0.7672724745322447
```

0〜1の一様乱数生成のくりかえし

コンピュータでの `rand()` は **値がランダムに決まる量という意味での確率変数** のわかりやすい例になっていると思える.

$\boxed{例}$ (Bernoulli試行, ベルヌーイ試行, ベルヌイ試行)

pを成功確率とよぶ

Bernoulli試行とは,

確率	p	$1-p$
値	1	0

で1と0をランダムに生成すること

0.7086835170539888 → 0 → 外れ
0.3129878077458778 → 0 → 外れ
0.4432299124211534 → 0 → 外れ
0.6380100590067396 → 0 → 外れ
0.7457553759007464 → 0 → 外れ
0.0042479988608554 → 1 → 当たり
0.4282692167397828 → 0 → 外れ
0.2591080281694401 → 1 → 当たり
0.6056891396222264 → 0 → 外れ
0.4718831513039459 → 0 → 外れ
0.2359750577106341 → 1 → 当たり
0.3658417371594386 → 0 → 外れ
0.4169159761488047 → 0 → 外れ
0.2234702295437706 → 1 → 当たり
0.2267949185649013 → 1 → 当たり
0.1054052960574101 → 1 → 当たり
0.5401955636162634 → 0 → 外れ
0.3620455865151407 → 0 → 外れ
0.9332297268604005 → 0 → 外れ
0.1009962782059489 → 1 → 当たり

0~1のあいだの一様乱数Tについて

$$\begin{cases} T < 0.3 \Rightarrow 1\text{を生成} \\ T \geq 0.3 \Rightarrow 0\text{を生成} \end{cases}$$

$p = 0.3$

1 → 当たり, 0 → 外れ

当たりの例

- くじが当たる.
- 商品を買ってもらえる.
- 病気になる.
- 製品が不良である.

離散(確率)分布

確率	p_1	p_2	p_3	$\cdots$
値	a_1	a_2	a_3	$\cdots$

probability (mass) function, pmf

確率(質量)関数 $P(x)$

$$P(a_i) = p_i \geq 0, \quad \sum_i P(a_i) = 1$$

(以下ではカンタンのため $a_i \in \mathbb{R}$ であると考える.)

有限または可算な離散集合 $\{a_1, a_2, \dots\}$ 上の確率(質量)関数 $P(x)$ が与えられているとき, $P(x)$ によって <u>離散分布</u> (discrete distribution) が定義されたという.

$\boxed{例}$ (ベルヌイ Bernoulli 分布)

確率	p	$1-p$
値	1	0

, $0 \leq p \leq 1$, $P(1) = p$, $P(0) = 1-p$.

このpを<u>成功確率</u>とよぶ

すなわち, $P(x) = p^x (1-p)^{1-x}$ $(x = 1, 0)$. この分布を Bernoulli(p) と書く.

(n 個 Bernoulli 分布を合わせると試行回数 n の Bernoulli 試行が得られる.)

$\boxed{\text{連続(確率)分布}}$ (probability) density function, pdf

$\mathbb{R}$ 上の(確率)密度函数： $p(x) \geq 0$, $\int_{-\infty}^{\infty} p(x)\,dx = 1$

実際には $\int_0^\infty$ や $\int_a^b$ になったりする。

$\mathbb{R}^n$ 上の確率密度函数も同様に定める。 → 多変量の連続分布

$\mathbb{R}$ 上の確率密度函数 $p(x)$ が与えられたとき、

$p(x)$ によって <u>連続分布</u> (continuous distribution) が定義されたという。

$\boxed{\text{例}}$ 0〜1の一様分布 (rand()のモデル化, Uniform(0,1) と書く.)

$$p(x) = \begin{cases} 1 & (0 < x < 1) \\ 0 & (\text{それ以外}) \end{cases}$$

$0 \leq x < 1$ でも $0 < x \leq 1$ でも $0 \leq x \leq 1$ でもよい。

コンピュータでの rand() は $0 \leq x < 1$

□

$\boxed{\text{例}}$ 平均 μ, 分散 σ^2 の <u>正規分布</u> (Normal(μ, σ) と書く)

σ は標準偏差

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

$$p(z \mid 0, 1) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$

平均0, 分散 1^2 の正規分布を <u>標準正規分布</u> とよび, Normal() と書く.

□

連続分布(密度函数 $p(x)$)での確率

連続分布において，確率を

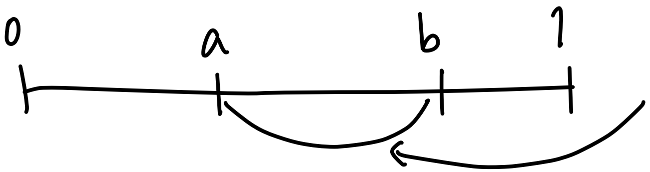
$$\left(c\text{ 以上で}(c\text{ より大きくて}) \ d\text{ 以下の}(d\text{ より小さい})\text{確率}\right) = \int_c^d p(x)\,dx$$

と定める．

$$\left(\text{ぴったり } c \text{ に等しい確率}\right) = \int_c^c p(x)\,dx = 0$$

例 $0 \leq p \leq 1$ のとき，$\left(\text{Uniform}(0,1)\text{ での値が } p \text{ 以下(未満)になる確率}\right) = \int_0^p 1\,dt = p.$

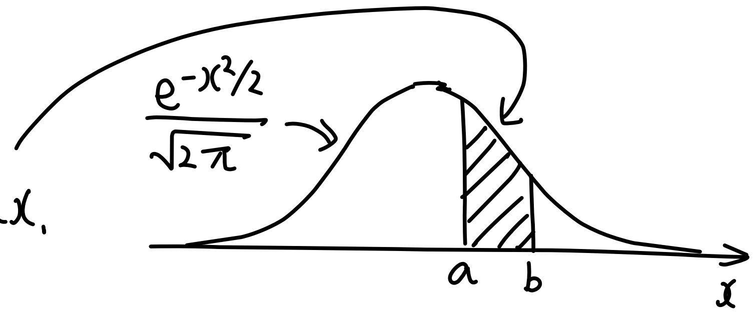
$0 \leq a \leq b \leq 1$ のとき，$\left(\text{Uniform}(0,1)\text{ で値が } a \text{ 以上 } b \text{ 以下になる確率}\right) = \int_a^b 1\,dt = b-a.$



この長さが 0〜1 での一様分布における確率．

例 $a \leq b$ のとき 標準正規分布において

$$\left(\text{値が } a \text{ 以上 } b \text{ 以下になる確率}\right) = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx.$$



$$\int_{-1}^{1} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx \approx 68.27\%, \quad \int_{-2}^{2} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx \approx 95.45\%, \quad \int_{-1.96}^{1.96} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx \approx 95.00\%$$

自分でコンピュータを使って確認せよ！

離散分布における確率変数

仮にこの分布の名前は D とする。

離散分布

確率	$P(a_1)$	$P(a_2)$	$\cdots$
値	a_1	a_2	$\cdots$

において、

random variable

$\downarrow$

各 a_i に実数 $f(a_i)$ を対応させる関数 $f(x)$ を離散分布上の確率変数と呼ぶ。

合わせる習慣

特に、a_i を a_i 自身に対応させる確率変数 X を（X は a_i を a_i に対応させる関数）

大文字

分布 D にしたがう確率変数 とよび、$X \sim D$ と書く。（$X \sim P(x)$ と書くこともある。）

X は確率質量関数 $P(x)$ を持つという。

そして、各 a_i に $f(a_i)$ を対応させる確率変数を $f(X)$ と表す。

直観的には $\begin{cases} X \text{ の値は確率 } P(a_i) \text{ で } a_i \text{ になる。} \\ f(X) \text{ の値は確率 } P(a_i) \text{ で } f(a_i) \text{ になる。} \end{cases}$

$f(X)$ の期待値 (expectation value) の定義 $E[f(X)] = \sum_i f(a_i)P(a_i)$.

$E[X] = \sum_i a_i P(a_i)$ を分布 D の期待値 もしくは 分布 D の平均 とよぶ。

注意 標本 $x_1, x_2, \ldots, x_n$ の平均 $\frac{1}{n}\sum_{i=1}^{n} x_i$ と分布 D の平均を区別せよ！

例 X は成功確率 $p=0.3$ の Bernoulli 分布にしたがう確率変数だとする:

$$X \sim \mathrm{Bernoulli}(0.3), \qquad P(1)=0.3, \quad P(0)=0.7$$

この Bernoulli 分布の期待値は

$$E[X] = 1 \times 0.3 + 0 \times 0.7 = 0.3.$$

確率 0.3 で当たりが出るくじで,

当たりが出たら 1000円もらえ, 外れが出たら 300円とられるという状況は,

上の Bernoulli 分布における次のように定義された確率変数 $f(X)$ でモデル化される:

大文字

$$f(x) = \begin{cases} 1000\text{円} & (x=1) \\ -300\text{円} & (x=0) \end{cases}$$

← ふつうの変数 x は小文字

このとき, $f(X)$ の期待値は

$$E[f(X)] = 1000\text{円} \times 0.3 + (-300\text{円}) \times 0.7 = 90\text{円}.$$

このように期待値の計算で使える<u>函数</u>たち $(X, f(X), \dots)$ を<u>確率変数</u>とよぶ.

連続分布における確率変数

確率密度関数 $p(x)$ で定めれた確率分布 D における確率変数とは x の関数 $f(x)$ のことである.

X は確率密度関数 $p(x)$ を持つという

特に, x を x 自身に対応させる確率変数 X を分布 D にしたがう確率変数と呼び, $X \sim D$ と書く. ($X \sim p(x)$ と書くこともある.)

following D

x の関数 $f(x)$ が定める確率変数を $f(X)$ と書く.

$\boxed{f(X)\text{の期待値の定義}}$ $E[f(X)] = \int f(x)\,p(x)\,dx$. ← 適切な定積分

$E[X] = \int x\,p(x)\,dx$ (定積分) を分布 D の期待値もしくは分布 D の平均とよぶ.

$\boxed{\text{注意}}$ 標本 $x_1, x_2, \ldots, x_n$ の平均 $\frac{1}{n}\sum_{i=1}^{n} x_i$ と分布 D の平均を区別せよ!

$\boxed{例}$ T は 0~1 の一様分布にしたがう確率変数であるとする:

$$T \sim \mathrm{Uniform}(0,1) \sim p(t) = \begin{cases} 1 & (0<t<1) \\ 0 & (\text{それ以外}) \end{cases}$$

この分布の期待値は

$$E[T] = \int_{-\infty}^{\infty} t\, p(t)\, dt = \int_0^1 t\, dt = \left[\frac{t^2}{2}\right]_0^1 = \frac{1}{2}.$$

← 区間 0~1 の中点

$$f(t) = \begin{cases} 1000\text{円} & (t < 0.3) \\ -300\text{円} & (t \geq 0.3) \end{cases}$$

で定まる確率変数 f(T) について,

$$E[f(T)] = \int_{-\infty}^{\infty} f(t)\, p(t)\, dt = \int_0^{0.3} 1000\text{円}\, dt + \int_{0.3}^{1} (-300\text{円})\, dt$$

$$= 1000\text{円} \times 0.3 + (-300\text{円}) \times 0.7 = 90\text{円}.$$

他の例:

$$E[T^m] = \int_0^1 t^m\, dt = \left[\frac{t^{m+1}}{m+1}\right]_0^1 = \frac{1}{m+1}.$$

← Uniform(0,1) のモーメント

$$E\left[\left(T-\tfrac{1}{2}\right)^2\right] = \int_0^1 \left(t-\tfrac{1}{2}\right)^2 dt = \left[\frac{1}{3}\left(t-\frac{1}{2}\right)^3\right]_0^1 = \frac{1/8 - (-1/8)}{3} = \frac{1}{12}.$$

Uniform(0,1) の分散

$\boxed{\text{確率変数に関する確率}}$ 確率変数 X について，次のように書く：

条件を書いた $P(\underline{c \leqq f(X) \leqq d}) = (c \leq f(X) \leq d$ となる確率$)$

$\leq$ と $\leqq$ は同じ意味

X が $\{a_1, a_2, \ldots\}$ 上の確率質量函数 $P(x)$ で定まる離散分布にしたがうとき，

$$P(c \leqq f(X) \leqq d) = \sum_{c \leqq f(a_i) \leq d} P(a_i) = \left(\begin{array}{l} c \leqq f(a_i) \leqq d \text{ をみたす } a_i \text{ の全体} \\ \text{に関する } P(a_i) \text{ の和} \end{array} \right)$$

X が確率密度函数 $p(x)$ を持つとき，

$$P(c \leqq f(X) \leqq d) = \int_{\{x \in \mathbb{R} \mid c \leqq f(x) \leqq d\}} p(x)\,dx.$$

$$1_{c \leqq f(X) \leqq d}(x) = \begin{cases} 1 & (c \leqq f(x) \leqq d) \\ 0 & (\text{それ以外}) \end{cases} \text{ のとき，}$$

$$P(c \leqq f(X) \leqq d) = E\left[1_{c \leqq f(X) \leqq d}(X)\right] = \begin{cases} \displaystyle\sum_i 1_{c \leqq f(X) \leqq d}(a_i)\, P_i \\ \displaystyle\int_{\mathbb{R}} 1_{c \leqq f(X) \leqq d}(x)\, p(x)\,dx \end{cases}$$

カイ

$\chi_{c \leqq f(X) \leqq d}(x)$ と書くこともある．

確率は期待値で書ける．

$E[\ ]$の基本性質

(1) 線形性： $E[\alpha f(X)+\beta g(X)] = \alpha E[f(X)] + \beta E[g(X)]$.

(2) 単調性： $f(x) \leqq g(x)$ （すべてのxについて） $\Rightarrow$ $E[f(X)] \leqq E[g(X)]$.

(3) 規格化条件： 定数αについて $E[\alpha]=\alpha$. （特に $E[1]=1$）.

計算のコツ　できるだけ基本性質だけを使ってすます！

連続分布での証明

定積分

(1) $E[\alpha f(X)+\beta g(X)] = \int (\alpha f(x)+\beta g(x))\, p(x)\, dx$

$$= \alpha \int f(x)p(x)\,dx + \beta \int g(x)p(x)\,dx = \alpha E[f(X)] + \beta E[g(X)].$$

(2) $p(x) \geqq 0$ と $f(x) \leqq g(x)$ より, $f(x)p(x) \leqq g(x)p(x)$ となるので,

$$E[f(X)] = \int f(x)p(x)\,dx \leqq \int g(x)p(x)\,dx = E[g(X)].$$

(3) $\int p(x)\,dx = 1$ より

$$E[\alpha] = \int \alpha\, p(x)\,dx = \alpha \int p(x)\,dx = \alpha.$$

$\boxed{\text{q.e.d.}}$

確率変数 $Y=f(X)$ の分散

平均からの距離の2乗の平均を分散と呼ぶ

$Y=f(X)$ の期待値を $\mu_Y = E[Y] = E[f(X)]$ と書くとき,

$$\sigma_Y^2 = \mathrm{Var}(Y) = \mathrm{Var}(f(X)) = E[(Y-\mu)^2] = E[(f(X)-\mu_Y)^2]$$

を $Y=f(X)$ の分散 (variance) とよぶ.

(注意 標本 $x_1, x_2, \dots, x_n$ の平均を $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ と書くときの

標本の分散 $\underbrace{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$ と確率変数の分散はちがうものである.

慣習的に不偏分散 $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})$ を使うことが多い.)

分散の平方根を標準偏差 (standard deviation) とよび, 次のように書く;

$$\sigma_Y = \mathrm{std}(Y) = \sqrt{\mathrm{Var}(Y)}.$$

分散の計算には, $\mu_Y = E[Y]$ のときの $\mathrm{Var}(Y) = E[(Y-\mu_Y)^2]$ だけではなく

$$\mathrm{Var}(Y) = E[Y^2] - E[Y]^2$$

もよく使われる.

よく使われる

$E[\ ]$ の線形性と規格化条件しか使っていない. →

証明 $\mathrm{Var}(Y) = E[(Y-\mu_Y)^2] = E[Y^2 - 2\mu_Y Y + \mu_Y^2]$

$= E[Y^2] - 2\mu_Y E[Y] + \mu_Y^2 = E[Y^2] - 2\mu_Y^2 + \mu_Y^2$

$= E[Y^2] - \mu_Y^2 = E[Y^2] - E[Y]^2.$

自習問題

答えは github においてある資料にある.

問題: 期待値と分散の特徴付け

X は有限離散分布に従う確率変数であるとし, 実数 t について $f(t) = E[(X-t)^2]$ とおく. このとき, $f(t)$ を最小化する t の値は X の期待値 $E[X]$ になり, 最小値 $f(E[X])$ は X の分散 $\mathrm{var}(X)$ になることを示せ.

この条件は使わない

注意: a を確率変数 X の中央値とするとき, $E[|X-a|]$ は中央値からの距離の期待値になる. これも確率変数 X の分布の広がりの大きさの指標の1つとして使える. 分散の定義について「どうして2乗するのか?」という質問は非常によくある質問だが, 分布の広がりの大きさの指標は無数にあるので, **分散に特権的な優位性があるわけではない**. 期待値の代わりに中央値を考え, 2乗せずに単なる絶対値を考えるのでもよい. その他にも中央値と四分位数の組み合わせもよく使われている. 分散や四分位数について教えることが多いのは単によく使われるからである. しかし, 期待値と分散の組み合わせを頻繁に目にする数学的理由もあって, **中心極限定理** によって正規分布で近似されると考えられる分布が応用上よく現れるという事情がある.

$\boxed{\text{確率変数のロケーションスケール変換}}$

E[ ]の基本性質の応用

$X \mapsto aX+b$ という変換のこと

確率変数 X と定数 a, b について,

(1) $E[aX+b] = aE[X]+b$ (これは E[ ]の基本性質より易しい.)

(2) $\mathrm{var}(aX+b) = a^2\,\mathrm{var}(X)$.

<u>証明</u> $\mu = E[X]$, $\mu' = E[aX+b] = a\mu+b$

$$\mathrm{Var}(aX+b) = E[(aX+b-\mu')^2] = E[(aX+b-(a\mu+b))^2]$$
$$= E[(a(X-\mu))^2] = E[a^2(X-\mu)^2] = a^2 E[(X-\mu)^2] = a^2\,\mathrm{var}(X).$$

(3) $a \geqq 0$ のとき, $\mathrm{std}(aX+b) = a\,\mathrm{std}(X)$.

<u>証明</u> $\mathrm{std}(aX+b) = \sqrt{\mathrm{var}(aX+b)} = \sqrt{a^2\,\mathrm{var}(X)} = a\sqrt{\mathrm{var}(X)} = a\,\mathrm{std}(X)$. □

← よく使われる.

特に, $\mu = E[X]$, $\sigma = \mathrm{std}(X)$ のとき, $\boxed{Z = \dfrac{X-\mu}{\sigma}}$ とおくと, $E[Z]=0$, $\mathrm{var}(X)=1$.

<u>証明</u> $E[Z] = E\left[\dfrac{X-\mu}{\sigma}\right] = \dfrac{1}{\sigma}(E[X]-\mu) = \dfrac{1}{\sigma^2}(E[X]-E[X]) = 0$.

$$\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) \underset{\substack{\uparrow \\ \text{上の(2)}}}{=} \frac{\mathrm{var}(X)}{\sigma^2} = \frac{\mathrm{var}(X)}{\mathrm{std}(X)^2} = 1.$$

$\boxed{\text{q.e.d.}}$

$\boxed{\text{例}}$ $X \sim \mathrm{Bernoulli}(p)$ のとき,

$E[X] = 1 \cdot p + 0 \cdot (1-p) = p$,

$E[X^2] = 1^2 \cdot p + 0^2 \cdot (1-p) = p$,

$\mathrm{Var}(X) = E[X^2] - E[X]^2 = p - p^2 = p(1-p)$. □

$\boxed{\text{例}}$ $T \sim \mathrm{Uniform}(0,1)$ のとき,

$E[T] = \int_0^1 t\, dt = \frac{1}{2}$

$E[T^2] = \int_0^1 t^2\, dt = \frac{1}{3}$

$\mathrm{Var}(T) = E[T^2] - E[T]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$. □

$\boxed{\text{例}}$ $X \sim \mathrm{Normal}(0,1)$ のとき,

$E[X] = \int_{-\infty}^{\infty} x \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 0$ (被積分函数が奇函数だから).

$\mathrm{Var}(X) = E[X^2] = \int_{-\infty}^{\infty} x^2 \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = 1$ となることを示そう.

$\left(\int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = \sqrt{2\pi} \text{ を示せばよい.} \right)$

Gauss積分の公式 $\int_{-\infty}^{\infty} e^{-y^2} dy = \sqrt{\pi}$ において,

$\alpha > 0$ について $y = \sqrt{\alpha}\, x$ とおくと, $\sqrt{\alpha} \int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\pi}$.

ゆえに, $\int_{-\infty}^{\infty} e^{-\alpha x^2} dx = \sqrt{\pi}\, \alpha^{-1/2}$.

両辺を α で微分して -1 倍すると, $\int_{-\infty}^{\infty} x^2 e^{-\alpha x^2} dx = \frac{1}{2} \sqrt{\pi}\, \alpha^{-3/2}$.

$\alpha = \frac{1}{2}$ とおくと, $\int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = \frac{1}{2} \sqrt{\pi}\ 2^{3/2} = \sqrt{2\pi}$.

□

試行回数 n, 成功確率 p の Bernoulli 試行の確率分布

与えられた1と0からなる長さ n の数列 $(x_1, \ldots, x_n)$ $(x_i = 1, 0)$ に対して,

毎回確率 p で1が, 確率 $1-p$ で0がランダムに得られる Bernoulli 試行を

n 回行ったときに得られる1と0からなる長さ n の数列が $(x_1, \ldots, x_n)$ になる確率は

Bernoulli 分布での確率

$$P(x_i) = p^{x_i}(1-p)^{1-x_i} = \begin{cases} 1 & (x_i = 1) \\ 0 & (x_i = 0) \end{cases}$$

の積になる:

n 回の Bernoulli 試行の確率質量函数 pmf

$$P(x_1, \ldots, x_n) = \prod_{i=1}^{n} \left(p^{x_i}(1-p)^{1-x_i} \right) = p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)}$$

ここで, $x_1 + \cdots + x_n$ は1と0からなる数列 $(x_1, \ldots, x_n)$ の中の1の個数になる.

$P(x_1, \ldots, x_n)$ $(x_i = 1, 0)$ によって定まる多変量離散分布を

$$\mathrm{Bernoulli}(p)^n = \underbrace{\mathrm{Bernoulli}(p) \times \cdots \times \mathrm{Bernoulli}(p)}_{n\text{個}}$$

と表す.

試行回数 n, 成功確率 p の二項分布 (binomial distribution)

試行回数 n, 成功確率 p の Bernoulli 試行で得られる

1と0からなる長さ n の数列の中に1が k 個含まれる確率は,

$x_1+\cdots+x_n=k$ となるような

$$P(x_1,\dots,x_n) = p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)} = p^k(1-p)^{n-k}$$

の和になり, その個数は k 個の1の位置を n 個の x_i たちから選ぶ組み合わせの数

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)\cdots(n-k+1)}{k!}$$

← ${}_nC_k$ ではなく $\binom{n}{k}$ と書く.

に等しいので

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

になる. $P(k)$ が定める離散分布を二項分布と呼び, $\mathrm{Binomial}(n,p)$ と表す.

$$\sum_{k=0}^{n} P(k) = \sum_{k=0}^{n}\binom{n}{k}p^k(1-p)^{n-k} = (p+(1-p))^n = 1$$

二項定理 $(x+y)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} x^k y^{n-k}$.

$\boxed{\text{自習問題}}$ $K \sim \mathrm{Binomial}(n,p)$ のときの $E[K]=np$, $\mathrm{Var}(K)=np(1-p)$ を示せ. □

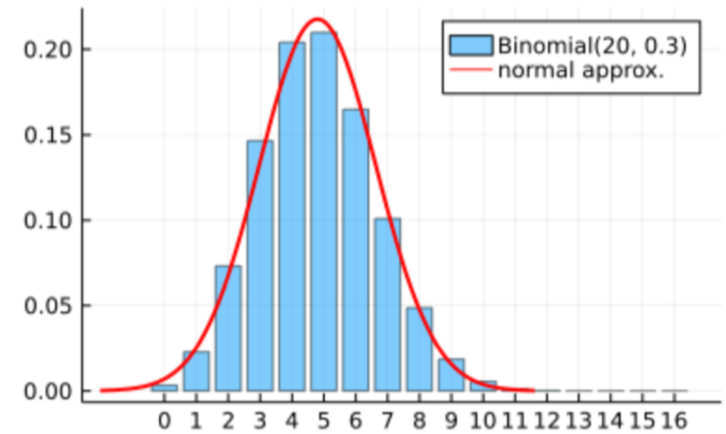
答えは github においてある資料にある. インターネットで検索しても答えが見付かる.

二項分布の確率質量函数のグラフ

二項分布 $\mathrm{Binomial}(n,p)$ はその期待値 np と反対側の期待値 $n(1-p)$ の両方が大きなとき, $\mathrm{Binomial}(n,p)$ と等しい平均 $\mu = np$ と分散 $\sigma^2 = np(1-p)$ を持つ正規分布で近似されること(二項分布の中心極限定理)がよく知られている. (実際には np がそう大きくなくても悪くない近似になる. これは階乗のStirlingの公式による近似がそう悪くならないからだと考えられる.)

二項分布の中心極限定理は階乗に関するStirlingの(近似)公式を使えば示せるが, そこそこ面倒な計算が必要になる. そのような計算を避けたい人も避けなかった人も, コンピュータで二項分布と正規分布のグラフを重ねてプロットすれば二項分布が正規分布で近似されることをすぐに納得できると思う. 正規分布については後で詳しく説明する.

```
bin = Binomial(16, 0.3)
μ, σ = mean(bin), std(bin)
x = support(bin)
bar(x, x -> pdf(bin, x); label="Binomial(20, 0.3)", alpha=0.5)
plot!(; xtick=x)
plot!(Normal(μ, σ); label="normal approx.", lw=2, c=:red)
```

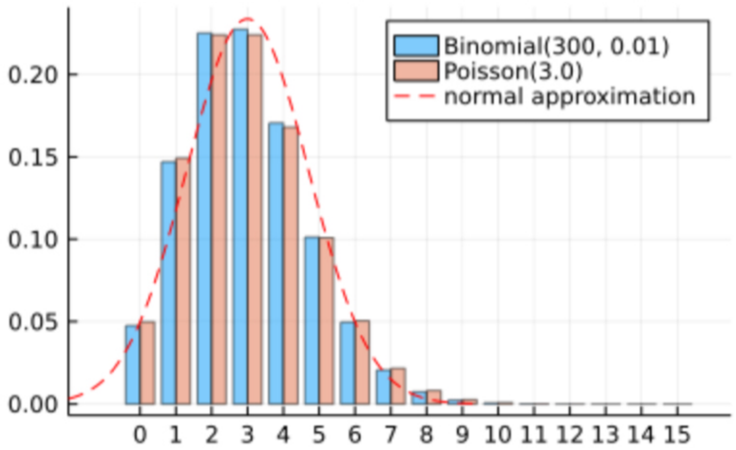


正規分布近似 (中心極限定理)

次のセルを見ればわかるように, p が小さな場合の二項分布は期待値 np のPoisson分布で近似される.

より正確に言うと, 固定された $\lambda > 0$ について, $p = \lambda/n$ とおくと, 二項分布 $\mathrm{Binomial}(n,p)$ は $n \to \infty$ でPoisson分布 $\mathrm{Poisson}(\lambda)$ に近付く. Poisson分布(ポアソン分布)の定義を知らない人はこれをPoisson分布の定義だと思ってもよい. この点については後で説明する予定である.

```
λ, n = 3, 100
p = λ/n
bin = Binomial(n, p)
μ, σ = mean(bin), std(bin)
x = 0:5λ
groupedbar(x, [pdf.(bin, x) pdf.(Poisson(n*p), x)];
    label=["Binomial(300, 0.01)" "Poisson($(n*p))"], alpha=0.5)
plot!(; xtick=x, xlim=(-2, 16.5))
plot!(Normal(μ, σ); label="normal approximation", lw=1, c=:red, l
```



Poisson分布近似

自習問題

問題: 当たりが30%の確率で出るくじを100回引いた場合

当たりが毎回 $p = 30\% = 0.3$ の確率で出るくじを $n = 100$ 回引いたときに当たりが出た回数 K は二項分布 $\mathrm{Binomial}(n, p)$ に従う確率変数とみなされる.

(1) 当たりの回数 K の期待値と分散と標準偏差を求めよ.

(2) 当たりの回数 K が25回以下になる確率と20回以下になる確率を求めよ.

(3) 当たりの回数 K が35回以上になる確率と40回以上になる確率を求めよ.

有効桁3桁で求めよ.

答えは githubにおいてある資料にある.

成功確率 p での幾何分布

$1 < p \leq 1$ と仮定し, $m=0,1,2,\ldots$ とする.

成功確率 p の Bernoulli 試行をつづけたとき,
ちょうど $n=m+1$ 回目で 1 がはじめて出る確率は,
m 回 0 がつづいて, $m+1$ 回目に 1 が出る確率に等しいので,

$$P(m) = (1-p)^m \times p = p(1-p)^m \quad (m=0,1,2,\ldots)$$

になる. この $P(m)$ が定める離散分布を幾何分布とよび, Geometric(p) と表す.

m は 0 が出た回数
n は試行回数
1 は 1 が出た回数

$\displaystyle\sum_{m=0}^{\infty} P(m) = 1$ となることの証明: $\displaystyle\sum_{m=0}^{\infty} p(1-p)^m = p\,\frac{1}{1-(1-p)} = p \times \frac{1}{p} = 1.$

等比級数の和

$\boxed{\text{自習問題}}$ $M \sim \text{Geometric}(p)$, $N = M+1 =$ (1 が出るまでの試行回数) のとき,

$$E[M] = \frac{1-p}{p}, \quad E[N] = E[M]+1 = \frac{1}{p}, \quad \mathrm{Var}(M) = \mathrm{Var}(N) = \frac{1-p}{p^2}$$

となることを示せ. □

答えは github においてある資料にある.

自習問題　幾何分布の応用 → ガチャ

問題: 当たりが出る確率が1%のガチャを当たりが出るまで回す場合

当たりが1%の確率で出るガチャを引き続けて当たりが出るまで回す回数を N と書く.

(1) N の期待値と分散と標準偏差を求めよ.

(2) 当たりが初めて出るまでの回数が150回以上になる確率と200回以上になる確率と300回以上になる確率を求めよ.

確率や標準偏差の値は有効桁3桁以上が合っていれば正解とする.

答えはgithubにおいてある資料にある.

成功回数 k, 成功確率 p の負の二項分布

$k>0$, $m=0,1,2,\dots$ とする.

成功確率 p の Bernoulli 試行をつづけたとき,　← 当たりが k 回出るまでガチャをまわした回数

ちょうど $n=k+m$ 回目で 1 が k 個出る確率は,

最初の $k+m-1$ 回中に 0 が m 回出て, $k+m$ 回目に 1 が出る確率に等しいので

$$P(m)=\binom{k+m-1}{m}p^{k-1}(1-p)^m\times p=\binom{k+m-1}{m}p^k(1-p)^m \quad (m=0,1,2,\dots)$$

になる. この $P(m)$ が定める m の分布を負の二項分布とよび, 次のように表す:

$$\text{Negative Binomial}(k,p)=\text{NegBin}(k,p).$$

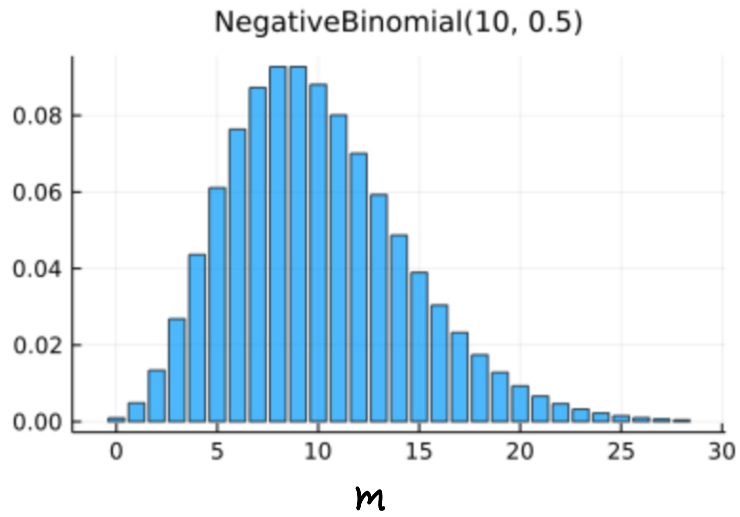
m は k 回 1 が出るまでに出た 0 の個数を意味している.

$P(m)$ を $n=k+m$ で表わすと, $\binom{k+m-1}{m}=\binom{k+m-1}{k-1}=\binom{n-1}{k-1}$ なので,

$$P(n-k)=\binom{n-1}{k-1}p^k(1-p)^{n-k} \quad (n=k,k+1,k+2,\dots).$$

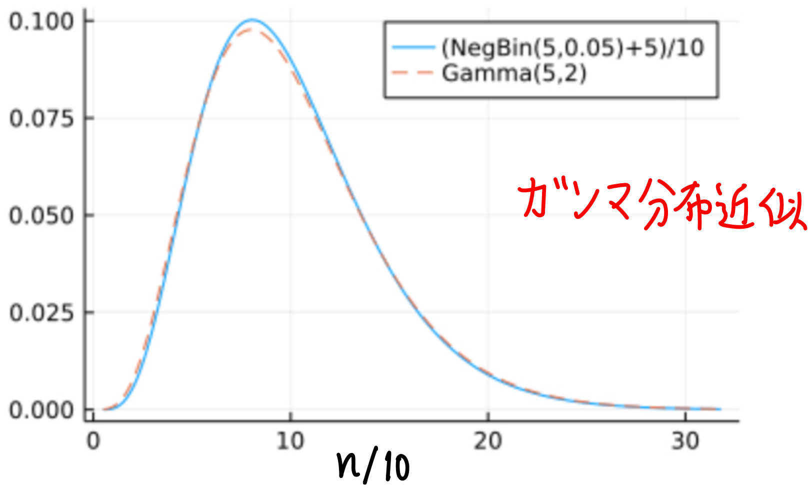
これは二項分布の確率の式に似ているが少しちがう.

負の二項分布の確率質量函数のグラフ



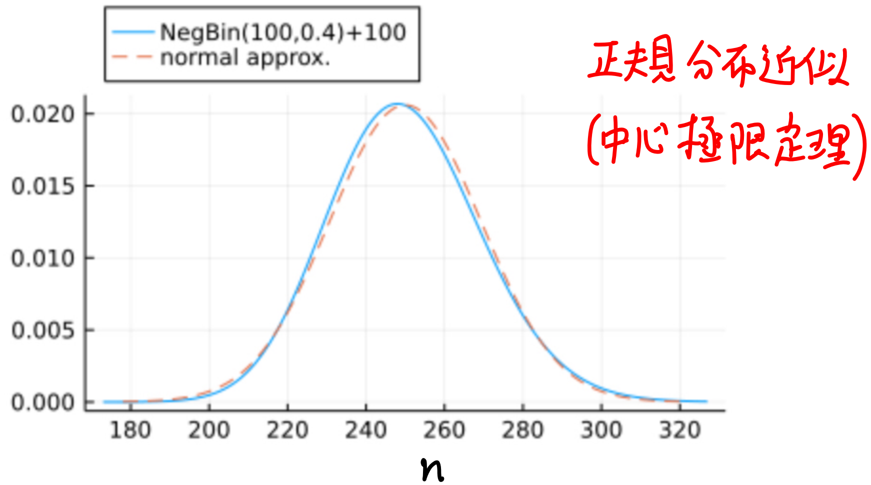
M は負の二項分布 $\mathrm{NegativeBinomial}(k,p)$ に従う確率変数であるとし, $N=M+k$ とおく. L は正の整数であるとし, $\theta=1/(Lp)$, $T=N/L$ とおくと, L を大きくしたとき, T の分布はガンマ分布 $\mathrm{Gamma}(k,\theta)$ で近似される.

```
k, θ = 5, 2
L = 10
p = 1/(L*θ)

negbin = NegativeBinomial(k, p)
μ, σ = mean(negbin), std(negbin)
m = 0:round(Int, μ + 5σ)
n = m .+ k
t = n ./ L
plot(t, pdf.(negbin, m)*L; label="(NegBin($k,$p)+$k)/$L")
plot!(t, pdf.(Gamma(k, θ), t); label="Gamma($k,$θ)", ls=:dash)
```



k が大きなとき, 負の二項分布 $\mathrm{NegativeBinomial}(k,p)$ は正規分布で近似される.

```
k, p = 100, 0.4

negbin = NegativeBinomial(k, p)
μ, σ = mean(negbin), std(negbin)
m = round(Int, μ - 4σ):round(Int, μ + 4σ)
n = m .+ k
plot(n, pdf.(negbin, m); label="NegBin($k,$p)+$k")
plot!(Normal(μ+k, σ); label="normal approx.", ls=:dash)
plot!(; legend=:outertop)
```



$\boxed{\text{どうして負の二項分布とよぶか？}}$ 以下，$|x|<1$とする．

二項係数：$\displaystyle \binom{a}{m} = \frac{a(a-1)(a-2)\cdots(a-m+1)}{m!}$

（aは0以上の整数でなくてもよい．$a=0,1,2,\dots$のときは $\displaystyle \binom{a}{m} = \frac{a!}{m!(a-m)!}$）

$f(x)=(1+x)^a$ とおくと，

$f^{(m)}(x) = a(a-1)\cdots(a-m+1)(1+x)^{a-m}$ なので，$\displaystyle \frac{1}{m!}f^{(m)}(0) = \binom{a}{m}$，

$f(x)$のTaylor展開を<u>二項展開</u>と呼ぶ：$\displaystyle (1+x)^a = \sum_{m=0}^{\infty}\binom{a}{m}x^m \quad (|x|<1)$

これの(a,x)を$(-a,-x)$におきかえると

$$(1-x)^{-a} = \sum_{m=0}^{\infty} \underbrace{(-1)^m\binom{-a}{m}}_{} x^m \quad (|x|<1).$$

−1倍の ← negative

仮に 負の二項係数 と呼ぼう．

（このノート内での特殊な用語）

負の二項展開（仮の名前）

さらに，

$$(-1)^m\binom{-a}{m} = (-1)^m\frac{(-a)(-a-1)\cdots(-a-m+1)}{m!} = \frac{a(a+1)\cdots(a+m-1)}{m!} = \binom{a+m-1}{m}$$

これの $a=k$ の場合が負の二項分布に出て来ている．

自習問題 $M \sim \mathrm{NegativeBinomial}(k, p)$, $N = k + M$ のとき,

$$E[M] = \frac{k(1-p)}{p}, \quad E[N] = k + E[M] = \frac{k}{p}, \quad \mathrm{Var}(M) = \mathrm{Var}(N) = \frac{k(1-p)}{p^2}$$

となることを示せ. 答えは github においてある資料にある.

自習問題

問題: 当たりが出る確率が5%のガチャを当たりが5回出るまで回す場合

当たりが $p = 5\% = 0.05$ で出るガチャを当たりがちょうど $k = 5$ 回出るまでに回した回数を N と書くことにする. N は負の二項分布 $\mathrm{NegativeBinomial}(k, p)$ に従う確率変数 M によって $N = M + k$ と書ける.

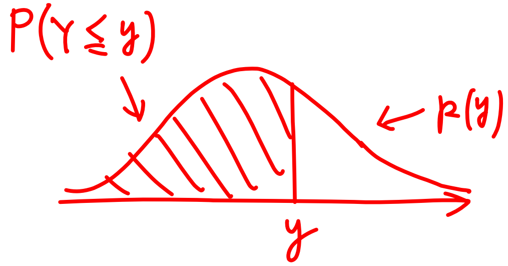
(1) N の期待値と分散と標準偏差を求めよ.

(2) N が150回以上, 200回以上になる確率を求めよ.

確率や標準偏差の値は有効桁3桁以上が合っていれば正解とする.

答えは github においてある資料にある.

累積分布函数 (cumulative distribution function, cdf) と 分位点函数 (quantile function)



確率変数 $Y=f(X)$ がしたがう分布の累積分布函数 $F_Y(y)$ を

$$F_Y(y) = P(Y \leq y) = (Y \leq y \text{ となる確率})$$

と定める. 累積分布函数 $F_Y(y)$ は y について広義単調増加函数になり,

$\lim_{y\to-\infty} F_Y(y) = 0$, $\lim_{y\to\infty} F_Y(y) = 1$, $\lim_{y\downarrow b} F_Y(y) = F_Y(b)$ (右側連続) になる.

$\boxed{例}$ $X \sim \mathrm{Bernoulli}(0.3)$ のとき,

$$F_X(x) = \begin{cases} 0 & (x<0) \\ 0.7 & (0\leq x<1) \\ 1 & (1\leq x) \end{cases}$$

$\boxed{例}$ $T \sim \mathrm{Uniform}(0,1)$ のとき,

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

累積分布函数の逆函数 (存在しない場合は適当に代替物を定める) を分位点函数 (クオンタイル函数) と呼び, $Q(p) = F_Y^{-1}(p)$ と書くことにする.

$Q_Y(0.5) = F_Y^{-1}(0.5)$ を Y がしたがう分布の中央値 (median) と呼び,

$Q_Y(0.25)$ と $Q_Y(0.75)$ を Y がしたがう分布の四分位点 (四分位数, quartile) と呼ぶ.

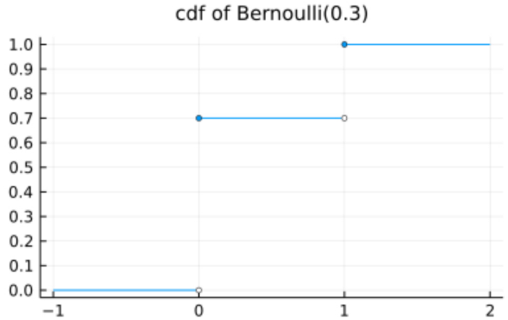
標本 (サンプル) $x_1, \dots, x_n$ の中央値や四分位数とはちがうものであることに注意せよ.

Bernoulli分布の累積分布函数

Bernoulli分布 $\mathrm{Bernoulli}(p)$ の累積分布函数は次になる:

$$
F(x) = \begin{cases} 0 & (x < 0) \\ 1-p & (0 \le x < 1) \\ 1 & (1 \le x) \end{cases}
$$

```
ber = Bernoulli(0.3)
x = vcat([-1:0.02:-0.02; NaN], [0:0.02:0.98; NaN], 1:0.02:2)
plot(x, x -> cdf(ber, x); label="")
title!("cdf of Bernoulli(0.3)")
plot!(; ytick = 0:0.1:1)
scatter!(0:1, x -> cdf(ber, x - 0.02) ; label="", c=:white, ms=3)
scatter!(0:1, x -> cdf(ber, x) ; label="", c=1, ms=3)
```



一様分布の累積分布函数

$a < b$ のとき, 一様分布 $\mathrm{Uniform}(a, b)$ の確率密度函数は

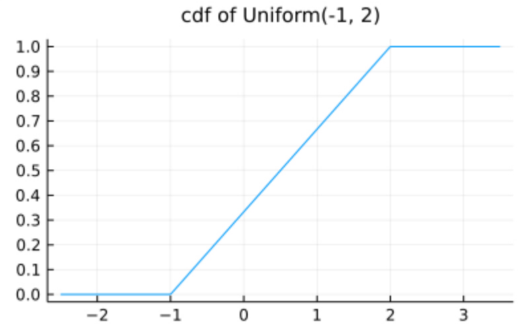
$$
p(x) = \begin{cases} 1/(b-a) & (a < x < b) \\ 0 & (\text{otherwise}) \end{cases}
$$

である. ゆえに, 一様分布 $\mathrm{Uniform}(a, b)$ の累積分布函数は

$$
F(x) = \int_{-\infty}^{x} p(x')\,dx' = \begin{cases} 0 & (x \le a) \\ (x-a)/(b-a) & (a \le x \le b) \\ 1 & (b \le x) \end{cases}
$$

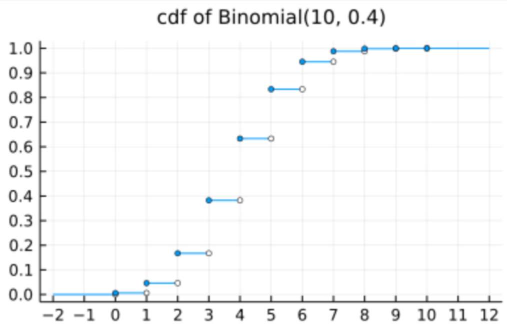
になる.

```
plot(x -> cdf(Uniform(-1, 2), x), -2.5, 3.5; label="")
title!("cdf of Uniform(-1, 2)")
yticks!(0:0.1:1)
```



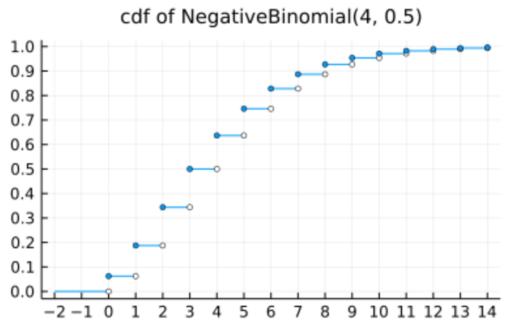
二項分布の累積分布函数のグラフ

```
bin = Binomial(10, 0.4)
x = vcat([-2:0.02:-0.02; NaN], ([k:0.02:k+0.98; NaN] for k in 0:9
plot(x, x -> cdf(bin, x); label="")
title!("cdf of Binomial(10, 0.4)")
plot!(; xtick = -2:12, ytick = 0:0.1:1)
scatter!(0:10, x -> cdf(bin, x - 0.02) ; label="", c=:white, ms=3
scatter!(0:10, x -> cdf(bin, x) ; label="", c=1, ms=3)
```



負の二項分布の累積分布函数のグラフ

```
negbin = NegativeBinomial(4, 0.5)
x = vcat([-2:0.02:-0.02; NaN], ([k:0.02:k+0.98; NaN] for k in 0:1
plot(x, x -> cdf(negbin, x); label="")
title!("cdf of NegativeBinomial(4, 0.5)")
plot!(; xtick = -2:14, ytick = 0:0.1:1)
scatter!(0:14, x -> cdf(negbin, x - 0.02) ; label="", c=:white, ms
scatter!(0:14, x -> cdf(negbin, x) ; label="", c=1, ms=3)
```



標準正規分布の累積分布函数と分位点函数

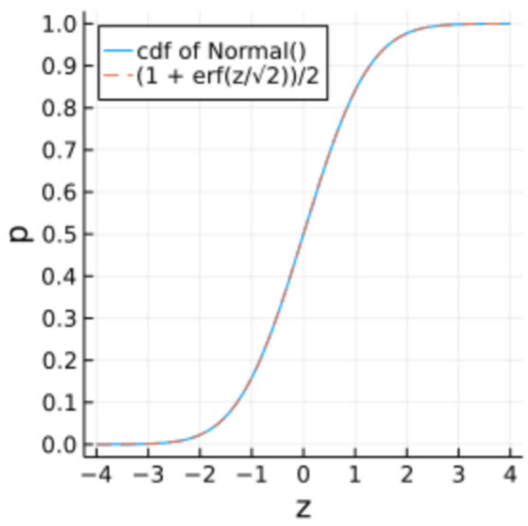
誤差函数が

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2)\,du$$

と定義される. この函数はコンピュータでの基本特殊函数ライブラリに含まれており, 効率的に計算可能である. この函数を使うと, 正規分布の累積分布函数は次のように書ける: $t = \sqrt{2}\,u$ とおくと,

$$\begin{aligned} F(z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2}\,dt = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^z e^{-t^2/2}\,dt \\ &= \frac{1}{2} + \frac{1}{\sqrt{\pi}} \int_0^{z/\sqrt{2}} e^{-u^2}\,du = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) = \frac{1+\operatorname{erf}(z/\sqrt{2})}{2}. \end{aligned}$$

```
plot(z -> cdf(Normal(), z), -4, 4;
    label="cdf of Normal()", legend=:topleft)
plot!(z -> (1 + erf(z/√2))/2, -4, 4;
    label="(1 + erf(z/√2))/2", ls=:dash)
plot!(; xtick=-5:5, ytick = 0:0.1:1,
    xlabel="z", ylabel="p")
plot!(; size=(300, 300))
```



$y = \operatorname{erf}(x)$ の逆函数 $x = \operatorname{erfinv}(y)$ もコンピュータでの基本特殊函数ライブラリに含まれており, 効率的に計算可能である. 標準正規分布の累積分布函数 $p = F(z)$ の逆函数(分位点函数, quantile function)は

$$z = Q_{\mathrm{Normal()}}(p) = \sqrt{2}\,\operatorname{erfinv}(2p-1)$$

と書ける. 標準正規分布の分位点函数は統計学で非常によく使われる. 例えば

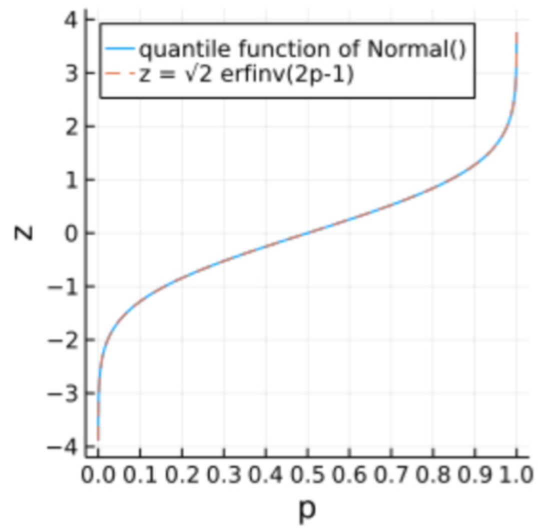
$$Q_{\mathrm{Normal()}}(0.975) \approx 1.96$$

は非常に有名な数値である. これは標準正規分布において値が 1.96 以上になる確率が 2.5% になることを意味している.

```
quantile(Normal(), 0.975)
```

```
1.9599639845400576
```

```
plot(p -> quantile(Normal(), p), 0, 1;
    label="quantile function of Normal()", legend=:topleft)
plot!(p -> √2*erfinv(2p-1), 0, 1;
    label="z = √2 erfinv(2p-1)", ls=:dash)
plot!(; xtick = 0:0.1:1, ytick=-5:5,
    xlabel="p", ylabel="z", ylim=(-4.2, 4.2))
plot!(; size=(300, 300))
```



自習問題

問題: 標準正規分布の四分位数

標準正規分布の累積分布函数と分位点函数はそれぞれ

$$
p = F(z) = \int_{-\infty}^{z} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt = \frac{1+\operatorname{erf}(z/\sqrt{2})}{2},
$$
$$
z = Q(p) = F^{-1}(p) = \sqrt{2}\,\operatorname{erfinv}(2p-1)
$$

と書ける. 標準正規分布の四分位数 $Q(0.25)$, $Q(0.75)$ をコンピュータを使って小数点以下第2桁まで求めよ.

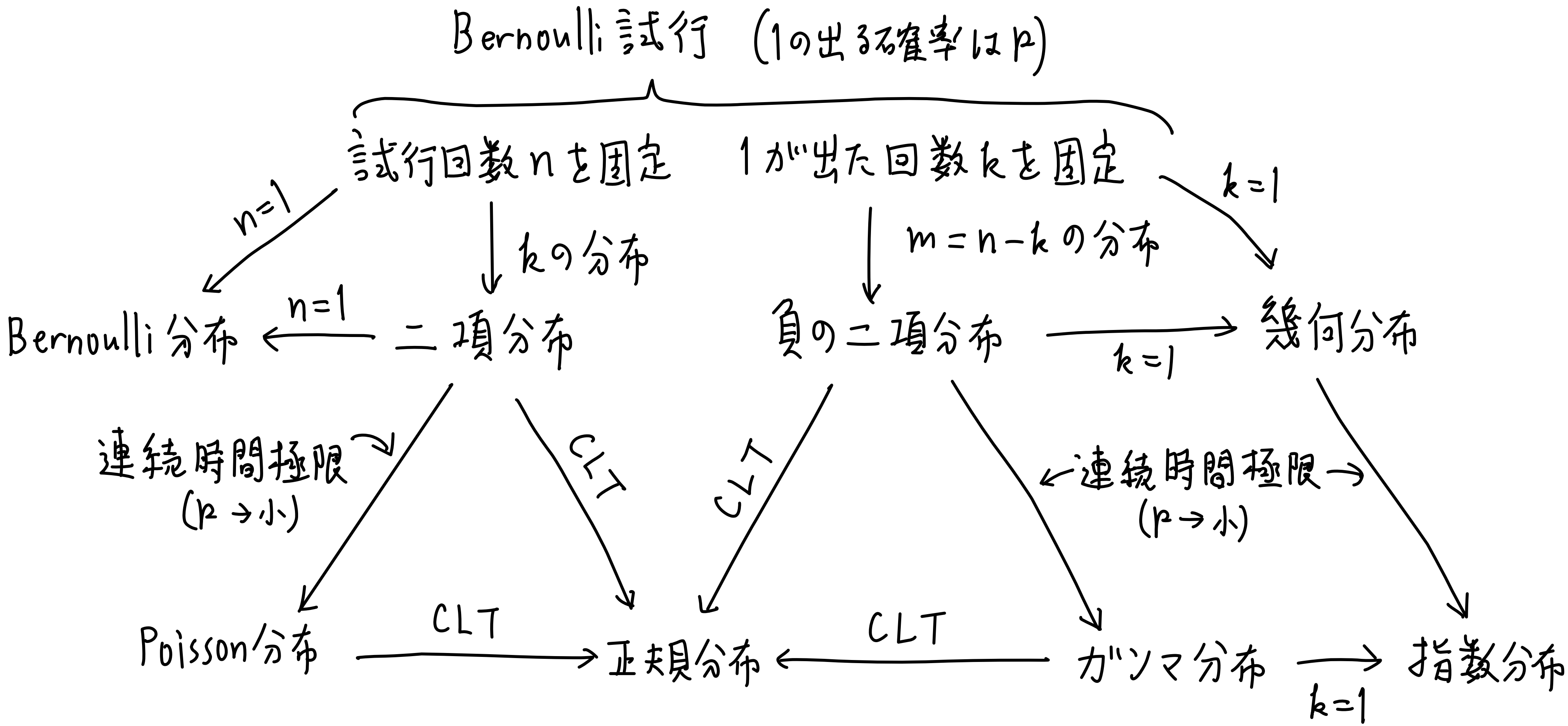
答えはgithubにおいてある資料にある.

まずはWolfram Alphaなどでerfinv(x)を使って求めることをためしてみよ.

$$
\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x \exp(-u^2)\,du
$$

は誤差函数(error function)とよばれている.
(erfinvはその逆函数)

Bernoulli試行の関連確率分布の図



CLT = 中心極限定理 (central limit theorem)

Gauss積分、ガンマ函数、ベータ函数

ガンマ分布とベータ分布も定義しておこう.

Gauss積分は正規分布の「分母」

Gauss積分 $\displaystyle\int_{-\infty}^{\infty} e^{-y^2}\,dy = \sqrt{\pi} \xrightarrow{\ y = \frac{x-\mu}{\sqrt{2\sigma^2}}\ } \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = \sqrt{2\pi\sigma^2}$

↓ 一般化, $\alpha > 0$

$y = x^{\frac{1}{2}}$, $dy = \frac{1}{2}x^{-\frac{1}{2}}\,dx$

$$\int_{-\infty}^{\infty} e^{-y^2}|y|^{2\alpha-1}\,dy = 2\int_0^{\infty} e^{-y^2} y^{2\alpha-1}\,dy = \int_0^{\infty} e^{-x} x^{\alpha-1}\,dx.$$

ガンマ函数 $\displaystyle\Gamma(\alpha) = \int_0^{\infty} e^{-x} x^{\alpha-1}\,dx = \int_{-\infty}^{\infty} e^{-y^2}|y|^{2\alpha-1}\,dy.$

ガンマ函数は Gauss積分の一般化

ベータ函数 $\displaystyle B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$

ガンマ函数とベータ函数は正規分布に関する統計学に自然に出て来る.

これらは統計学的には正規分布モデルの話になっているとみなされる. (それ以外の見方もできる.)

$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$ の証明

$I = \int_{-\infty}^{\infty} e^{-x^2} dx$ とおくと,

$$I^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)} dx\, dy$$

$= $ (右の山の体積)

$= \int_0^1$ (高さ z での断面の面積) dz

$$= \int_0^1 \pi \left(\sqrt{-\log z}\right)^2 dz$$

$$= -\pi \int_0^1 \log z\, dz$$

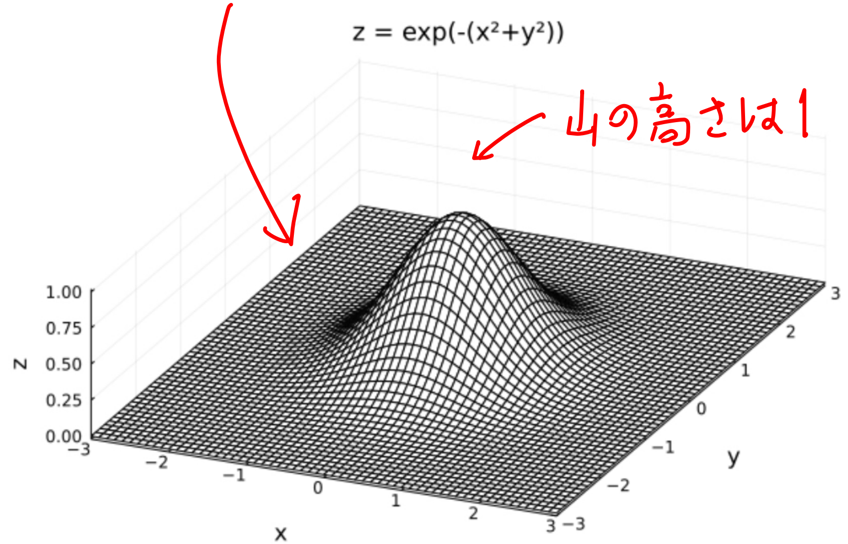
$$= -\pi \left[z \log z - z\right]_0^1$$

$$= -\pi \left(\underbrace{1 \log 1}_{=0} - 1 - \left(\underbrace{0 \log 0}_{=0} - 0\right)\right) = \pi,$$

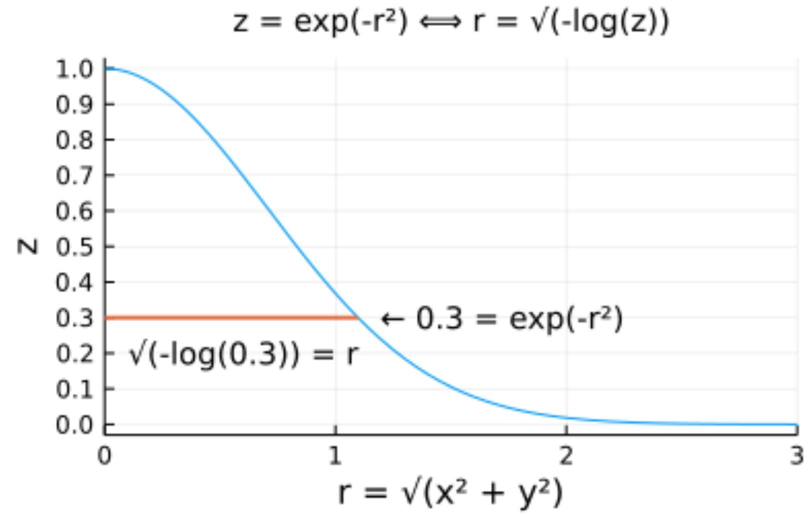
$\therefore\ I = \sqrt{\pi}$.

この山の体積が I^2 になる.

山の高さは1



この山は z 軸 $x = y = 0$ を中心にして回転対称である. 次のグラフは横軸を半径 $r = \sqrt{x^2+y^2}$ とした場合のこの山の高さのグラフである.



半径は $\sqrt{-\log z}$ ⇐

$\boxed{\text{Gauss積分の一般化}}$ $\alpha > 0$ と仮定する。 $y = x^{\frac{1}{2}},\ dy = \frac{1}{2} x^{-\frac{1}{2}} dx$

$$\int_{-\infty}^{\infty} e^{-y^2} |y|^{2\alpha-1} dy = 2\int_0^{\infty} e^{-y^2} y^{2\alpha-1} dy = \underbrace{\int_0^{\infty} e^{-x} x^{\alpha-1} dx}_{= \text{ガンマ関数 } \Gamma(\alpha)}.$$

$\boxed{\text{計算例}}$ $k = 0, 1, 2, \ldots$ について，$\alpha = k + \frac{1}{2} = \frac{2k+1}{2}$ のとき，

$$\int_{-\infty}^{\infty} e^{-y^2} y^{2k} dy = \frac{1 \cdot 3 \cdots (2k-1)}{2^k} \sqrt{\pi}.$$

$\boxed{\text{証明}}$ $\int_{-\infty}^{\infty} e^{-y^2} dy = \sqrt{\pi}$ で $y = \sqrt{a}\, t$ とおくと，$\int_{-\infty}^{\infty} e^{-at^2} \sqrt{a}\, dt = \sqrt{\pi}$.

$\int_{-\infty}^{\infty} e^{-at^2} dt = \sqrt{\pi}\, a^{-\frac{1}{2}}$. これの両辺に $\left(-\frac{\partial}{\partial a}\right)^k$ を作用させると，

$$\int_{-\infty}^{\infty} e^{-at^2} t^{2k} dt = \frac{1}{2} \frac{3}{2} \cdots \frac{2k-1}{2} \sqrt{\pi}\, a^{-\frac{2k+1}{2}}.$$

$a = 1$ とおくと，

$$\int_{-\infty}^{\infty} e^{-t^2} t^{2k} dt = \frac{1 \cdot 3 \cdots (2k-1)}{2^k} \sqrt{\pi}. \qquad \square$$

口頭での説明は略す.

ガンマ函数

$\alpha > 0$ と仮定する.

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1}\,dx = \int_{-\infty}^\infty e^{-y^2} |y|^{2\alpha-1}\,dy.$$

他にも色々役に立つ.

これは Gauss 積分の一般化になっているので, 正規分布で~~も~~役に立つ.

例 $\Gamma\left(\frac{1}{2}\right) = \int_{-\infty}^\infty e^{-y^2}\,dy = \sqrt{\pi}.$ **例** $\Gamma(1) = \int_0^\infty e^{-x}\,dx = \left[-e^{-x}\right]_0^\infty = 1.$

$\boxed{\Gamma(\alpha+1) = \alpha\Gamma(\alpha)}$ 函数等式

$$\Gamma(\alpha+1) = \int_0^\infty e^{-x} x^\alpha\,dx = \int_0^\infty (-e^{-x})' x^\alpha\,dx$$

部分積分

$$= \left[-e^{-x} x^\alpha\right]_0^\infty - \int_0^\infty (-e^{-x})(x^\alpha)'\,dx$$

$\alpha > 0$ より $0^\alpha = 0$

$$= \int_0^\infty e^{-x} \alpha x^{\alpha-1}\,dx = \alpha\Gamma(\alpha).$$

$k = 0, 1, 2, \dots$ とする.

例 $\Gamma(k+1) = k\Gamma(k) = k(k-1)\Gamma(k-1) = \cdots = k(k-1)\cdots 2\cdot 1\,\Gamma(1) = k!$

$\Gamma(n+1) = n!$

ガンマ函数は Gauss 積分と階乗の同時一般化!

例 $\Gamma\left(\frac{2k+1}{2}\right) = \frac{2k-1}{2}\Gamma\left(\frac{2k-1}{2}\right) = \frac{2k-1}{2}\frac{2k-3}{2}\Gamma\left(\frac{2k-3}{2}\right) = \cdots$

$$= \frac{2k-1}{2}\frac{2k-3}{2}\cdots\frac{3}{2}\frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1\cdot 3\cdots(2k-1)}{2^k}\sqrt{\pi}$$

前ページですでに示している.

自習問題 $\Gamma\left(\frac{2k+1}{2}\right) = \frac{(2k)!}{2^{2k}k!}\sqrt{\pi}$ $(k = 0, 1, 2, \dots)$ を示せ.

Github にある資料に答えが書いてある.

通常ガンマ函数は次の形で使われる. $\alpha > 0,\ \theta > 0$ と仮定する.

$$\boxed{\int_0^\infty e^{-x/\theta} x^{\alpha-1}\,dx = \theta^\alpha \Gamma(\alpha).}$$

証明 $x = \theta t$ とおくと, $(\text{左辺}) = \int_0^\infty e^{-t} \theta^{\alpha-1} t^{\alpha-1} \theta\,dt = \theta^\alpha \Gamma(\alpha).$ □

$\left\{\begin{array}{l}\alpha : \text{形状パラメータ} \\ \theta : \text{スケールパラメータ}\end{array}\right.$

ガンマ分布の定義 Gamma(α, θ) と表す.

次の確率密度函数で定義される $x > 0$ における分布をガンマ分布と呼ぶ.

$$p(x \mid \alpha, \theta) = \frac{1}{\theta^\alpha \Gamma(\alpha)} e^{-x/\theta} x^{\alpha-1} \quad (x > 0).$$

$\alpha = 1$ のとき, 指数分布と呼び, Exponential(θ) と表す.

$X \sim \text{Gamma}(\alpha, \theta)$ のとき, $E[f(X)] = \dfrac{1}{\theta^\alpha \Gamma(\alpha)} \displaystyle\int_0^\infty f(x)\, e^{-x/\theta} x^{\alpha-1}\,dx.$

自習問題 分布 Gamma(α, θ) の期待値と分散がそれぞれ $\alpha\theta,\ \alpha\theta^2$ となることを示せ.

ヒント $E[X^k]$ を求めよ. $\mathrm{Var}(X) = E[X^2] - E[X]^2$ を使え. □

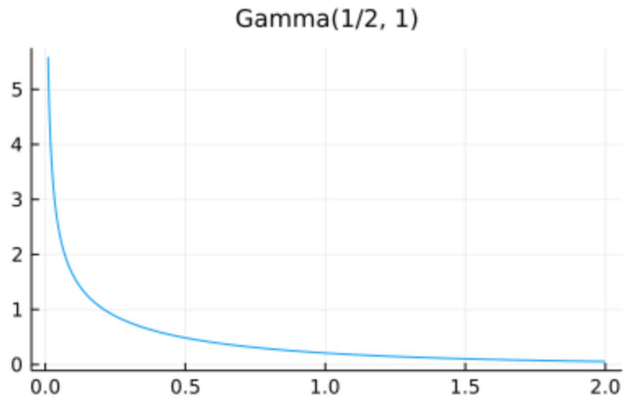
答えは Github においてある資料にある.

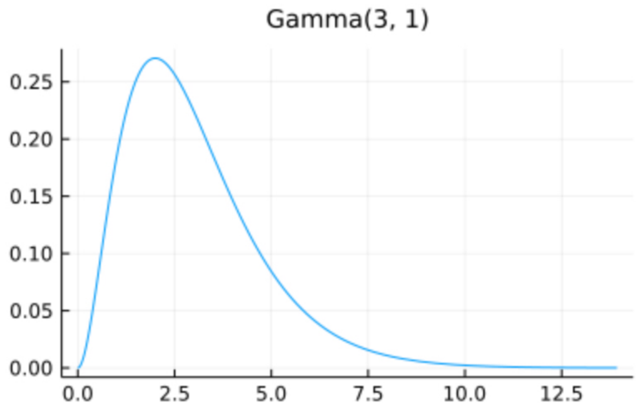
ガンマ函数について非常に良い練習になる.

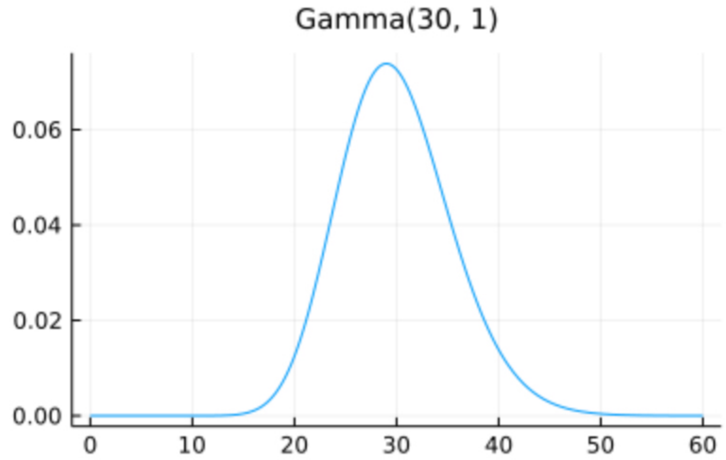
ガンマ分布
Gamma(α, θ)
のグラフ

スケールパラメータ θ を変えても、グラフの形状は変わらない。

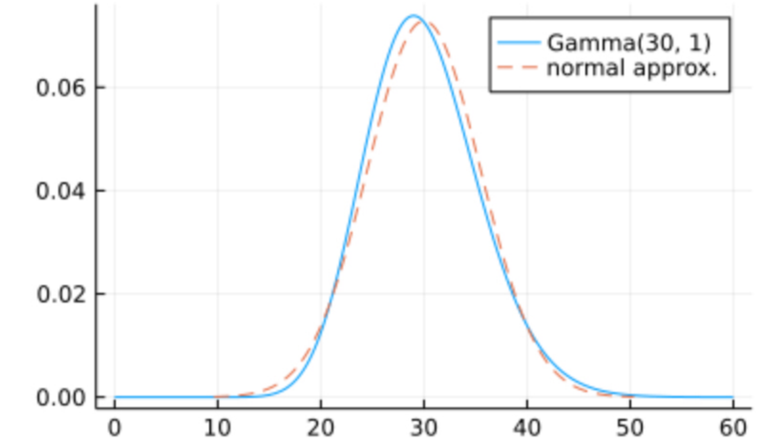
だから、グラフの形状を知りたければ形状パラメータ α を動かしてグラフを描けばよい。



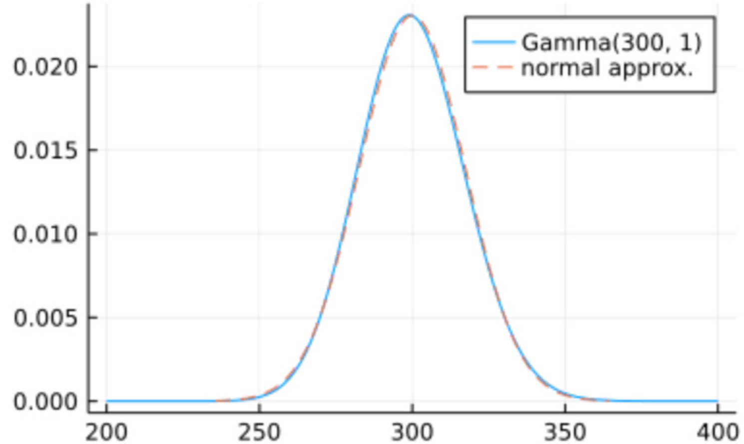




形状パラメータ α が大きなガンマ分布は正規分布で近似される.



形状パラメータを大きくすると、ガンマ分布は正規分布に近づく



自習問題 $n-1$次元単位球面S^{n-1}とn次元単位球体B^nが次のように定義される:

$$S^{n-1} = \{(x_1,\dots,x_n)\in\mathbb{R}^n \mid x_1^2+\cdots+x_n^2=1\},\quad B^n = \{(x_1,\dots,x_n)\in\mathbb{R}^n \mid x_1^2+\cdots+x_n^2 \leqq 1\}$$

このとき, 次が成立することを示せ:

$$A_{n-1} = (S^{n-1}\text{の面積}) = \frac{2\pi^{n/2}}{\Gamma(n/2)},\quad V_n = (B^n\text{の体積}) = \frac{A_{n-1}}{n} = \frac{\pi^{n/2}}{\Gamma(n/2+1)}.\qquad \square$$

答えはGithubにおいてある資料にある. 人生全体の中で1週間くらい考える価値のある問題.

例 ($n=1,2,3,4,5$の場合)

$$A_0 = (S^0=\{\pm1\}\text{の点の個数}) = \frac{2\pi^{1/2}}{\Gamma(1/2)} = 2,\qquad V_1 = (\text{区間}[-1,1]\text{の長さ}) = \frac{2}{1} = 2.$$

$$A_1 = (S^1\text{の長さ}) = \frac{2\pi}{\Gamma(1)} = 2\pi,\qquad V_2 = (B^2\text{の面積}) = \frac{2\pi}{2} = \pi.$$

$$A_2 = (S^2\text{の面積}) = \frac{2\pi^{3/2}}{\Gamma(3/2)} = \frac{2\pi^{3/2}}{(1/2)\pi^{1/2}} = 4\pi,\qquad V_3 = (B^3\text{の体積}) = \frac{4\pi}{3}$$

$$A_3 = \frac{2\pi^2}{\Gamma(2)} = 2\pi^2,\qquad V_4 = \frac{2\pi^2}{4} = \frac{\pi^2}{2}$$

$$A_4 = \frac{2\pi^{5/2}}{\Gamma(5/2)} = \frac{2\pi^{5/2}}{(3/2)(1/2)\pi^{1/2}} = \frac{8\pi^2}{3},\qquad V_5 = \frac{8\pi^2}{3\cdot5} = \frac{8\pi^2}{15}.\qquad \square$$

$\boxed{\text{Stirlingの公式}}$ $n! = n^n e^{-n} \sqrt{2\pi n}\,(1+o(1))$ ($o(1)$ は $n\to\infty$ で 0 に収束する量)

空気のごとく使われる。 小文字

$\boxed{\text{証明}}$ $n\to\infty$ で $\dfrac{n!}{n^n e^{-n}\sqrt{n}} \to \sqrt{2\pi}$ となることを示せばよい。

ポイント!

$n! = \Gamma(n+1) = \int_0^\infty e^{-x} x^n\,dx$ で $\boxed{x = n+\sqrt{n}\,y = n\left(1+\dfrac{y}{\sqrt{n}}\right)}$ とおくと、

$$n! = \int_{-\sqrt{n}}^\infty e^{-(n+\sqrt{n}\,y)}\left(n\left(1+\frac{y}{\sqrt{n}}\right)\right)^n \sqrt{n}\,dy = \underline{n^n e^{-n}\sqrt{n}} \int_{-\sqrt{n}}^\infty e^{-\sqrt{n}\,y}\left(1+\frac{y}{\sqrt{n}}\right)^n dy.$$

大文字

$$\log\left(e^{-\sqrt{n}\,y}\left(1+\frac{y}{\sqrt{n}}\right)^n\right) = n\log\left(1+\frac{y}{\sqrt{n}}\right) - \sqrt{n}\,y = n\left(\frac{y}{\sqrt{n}} - \frac{y^2}{2n} + O(n^{-3/2})\right) - \sqrt{n}\,y$$

$\log(1+a) = a - \dfrac{a^2}{2} + \dfrac{a^3}{3} + \cdots$

$$= -\frac{y^2}{2} + O(n^{-1/2}) \longrightarrow -\frac{y^2}{2} \quad (n\to\infty)$$

Laplace近似という方法を使っている。積分表示の御利益!

ゆえに、$n\to\infty$ のとき、$e^{-\sqrt{n}}\left(1+\dfrac{y}{\sqrt{n}}\right)^n \to e^{-y^2/2}$ なので、

$$\frac{n!}{n^n e^{-n}\sqrt{n}} = \int_{-\sqrt{n}}^\infty e^{-\sqrt{n}\,y}\left(1+\frac{y}{\sqrt{n}}\right)^n dy \longrightarrow \int_{-\infty}^\infty e^{-y^2/2}\,dy = \sqrt{2\pi}.$$

$\boxed{\text{q.e.d.}}$

極限の交換が気になる人は自分でどうにかすること。Lebesgueの収束定理を使えば容易。

$\boxed{\text{自習問題}}$ $n=1,2,\ldots,10$ について，$a=n!$ と $b=n^n e^{-n}\sqrt{2\pi n}$ と相対誤差 $\dfrac{b}{a}-1$ を計算せよ．b については小数点以下第3桁まで，相対誤差については小数点以下第5桁まで求めよ．

Github においてある資料に答えがある．

$\boxed{\text{例}}$ $a=3!=6$，$b=3^3e^{-3}\sqrt{2\pi\cdot 3}\approx 5.836$，$\dfrac{b}{a}-1\approx -0.02730 < 3\%$

$n=3$ ですでに相対誤差が 3% を切っている！

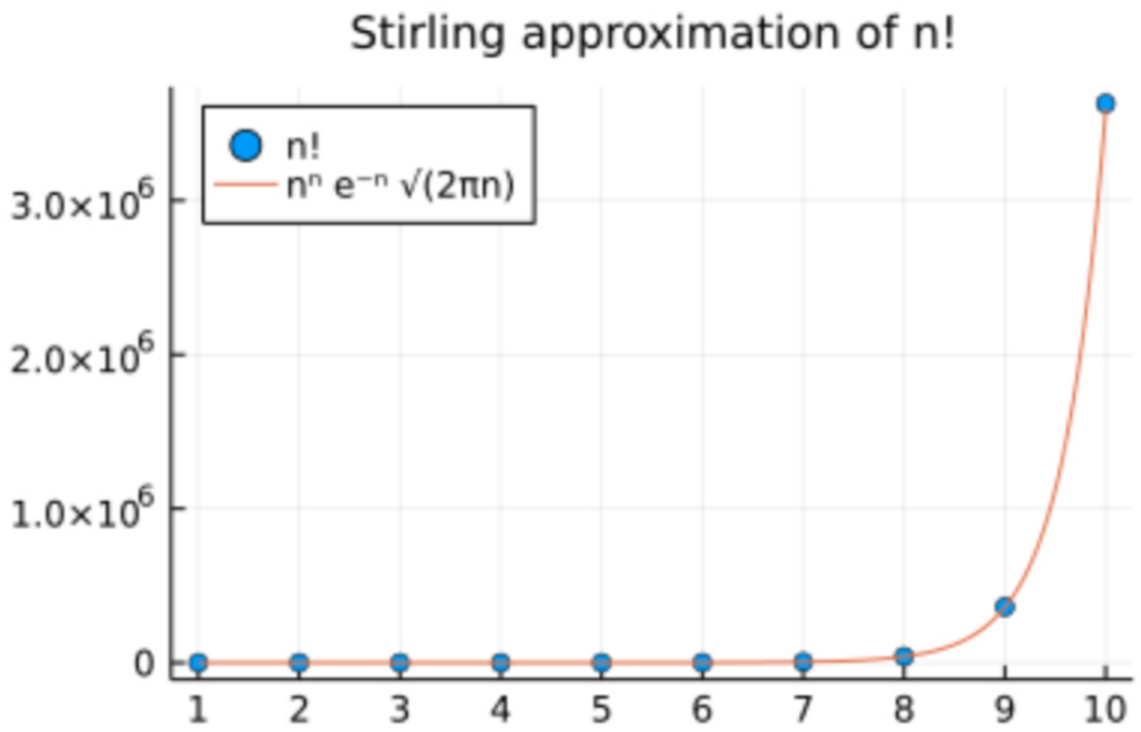
Stirling の公式は $n\to\infty$ として証明された $n!$ の近似公式だが，大きくない n であってもかなり精度が高い．

重要

$\boxed{\text{おまけ}}$ $n! = n^n e^{-n}\sqrt{2\pi n}\left(1+\dfrac{1}{12n}+O\left(\dfrac{1}{n^2}\right)\right)$ が成立している．

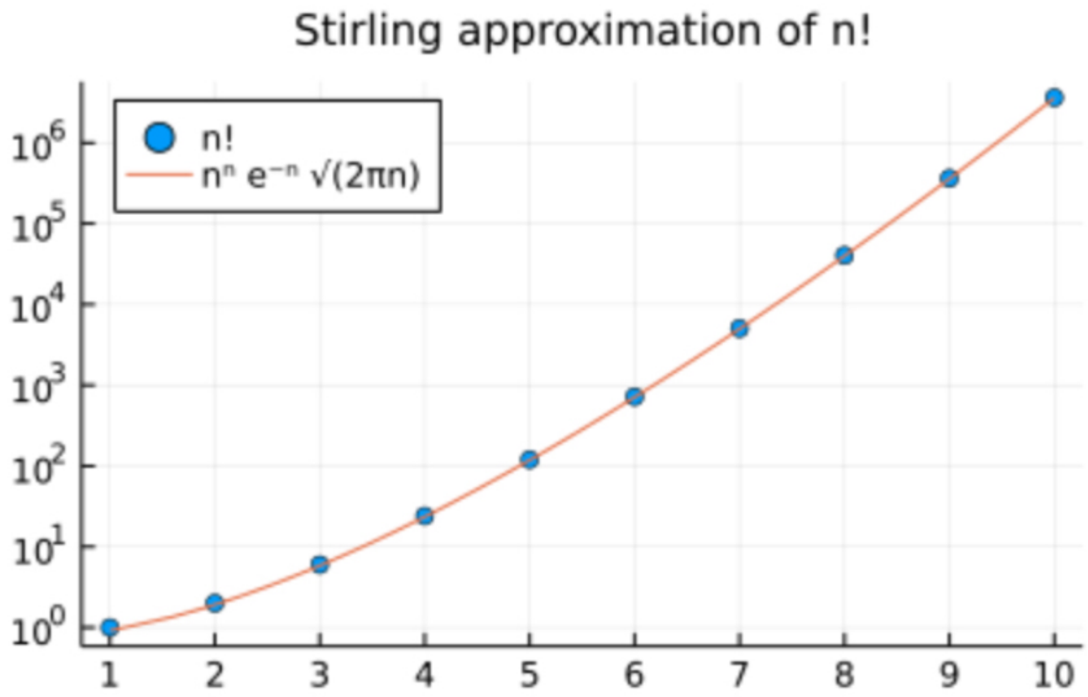
$1!=1$，$1^1e^{-1}\sqrt{2\pi\cdot 1}\left(1+\dfrac{1}{12\cdot 1}\right)=\dfrac{13\sqrt{2\pi}}{12e}\approx 0.99898\cdots$ ← ほぼ 1！

$\dfrac{1}{12n}$ で補正した近似公式であれば $n=1$ で相当な精度になっている！



スターリングの公式による
n! の近似のグラフ

n=10 以下でも
よく一致している。



対数スケールで比較

Stirling の公式を使った
近似は精度が高いので
使い易い。

ディガンマ，トリガンマ，ポリガンマ函数

応用上，ガンマ函数の導函数が必要になる場合がある．

$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1}\,dx$ より，$\Gamma^{(k)}(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} (\log x)^k\,dx$．

$\Gamma^{(k)}(\alpha)$ そのものではなく，

$$\psi(\alpha) = \frac{\partial}{\partial\alpha}(\log\Gamma(\alpha)) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}, \quad \psi'(\alpha), \quad \psi^{(k)}(\alpha) = \left(\frac{\partial}{\partial\alpha}\right)^{k+1} \log\Gamma(\alpha)$$

はそれぞれディガンマ，トリガンマ，ポリガンマ函数と呼ばれる基本特殊函数であり，コンピュータ上のライブラリで容易に計算できる．

このことを知っていると，ガンマ函数の導函数が必要な計算（たとえばガンマ分布モデルの最尤法）のコンピュータでの実装で困らない．

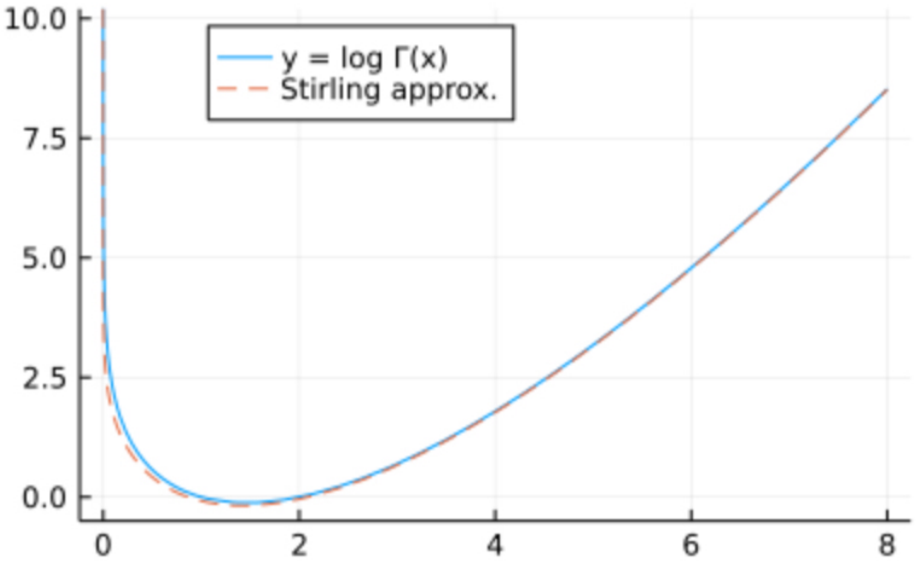
Gaussの乗法公式

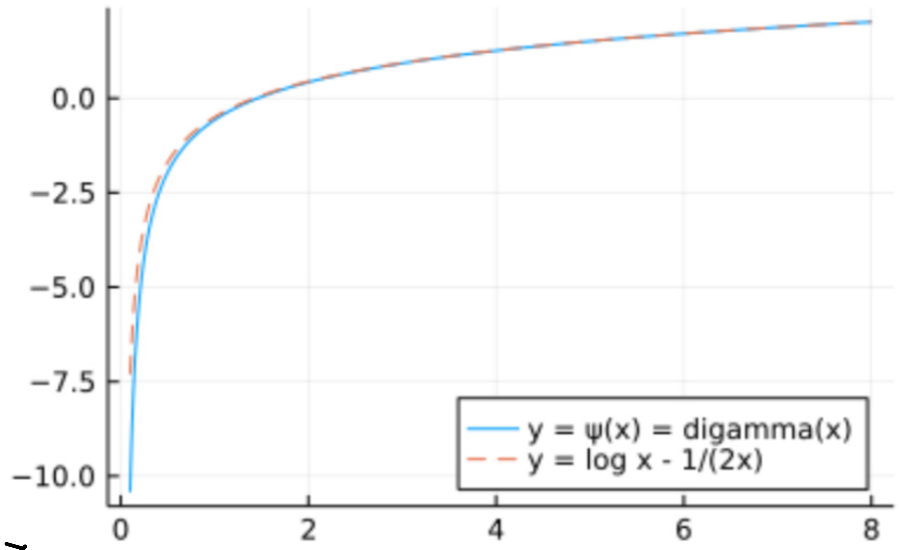
$$\Gamma(\alpha) = \lim_{n\to\infty} \frac{n^\alpha n!}{\alpha(\alpha+1)\cdots(\alpha+n)}$$ より，$$\psi(\alpha) = (\log\Gamma(\alpha))' = \lim_{n\to\infty}\left(\log n - \sum_{k=0}^{n} \frac{1}{\alpha+k}\right),$$

$$\psi'(\alpha) = \sum_{k=0}^{\infty} \frac{1}{(\alpha+k)^2}, \quad -\psi(1) = \gamma \approx 0.5772 \ (\text{Euler の } \gamma)$$ などが得られ，

$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ から，$\psi(\alpha+1) = \psi(\alpha) + \frac{1}{\alpha}$，$\psi'(\alpha+1) = \psi'(\alpha) - \frac{1}{\alpha^2}$ などが得られる．

対数ガンマ函数 $\log\Gamma(x)$, ディガンマ函数 $\psi(x) = (\log\Gamma(x))'$, トリガンマ函数 $\psi'(x)$ のグラフ

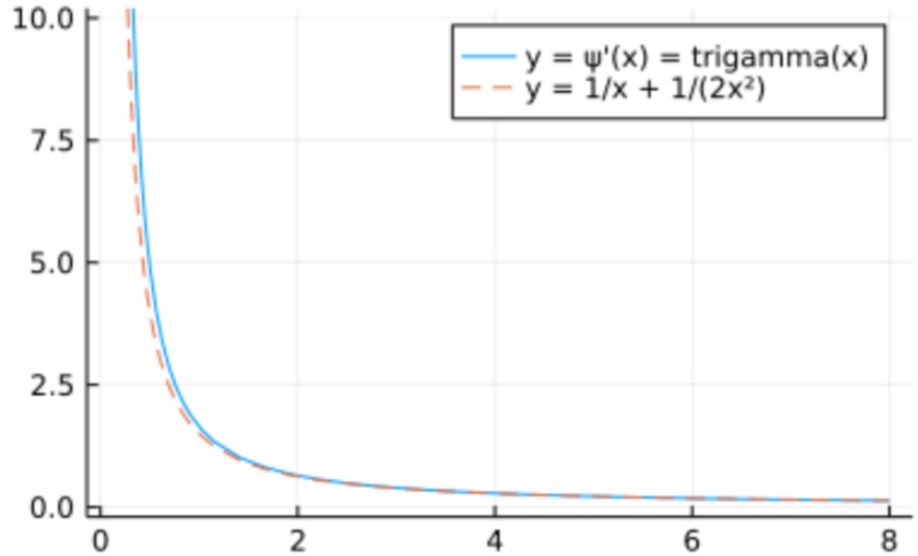




対数ガンマ

ディガンマ

トリガンマ



$$\log\Gamma(x) \approx x\log x - x - \frac{1}{2}\log x + \log\sqrt{2\pi}$$

$$\psi(x) \approx \log x - \frac{1}{2x}$$

$$\psi'(x) \approx \frac{1}{x} + \frac{1}{2x^2}$$

$\boxed{\text{ベータ函数}}$ $B(\alpha,\beta)=\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \quad (\alpha,\beta>0)$.

($t=1-u$ とおくと $B(\alpha,\beta)=B(\beta,\alpha)$ だとわかる)

$\boxed{\text{ベータ函数のガンマ函数表示}}$ $B(\alpha,\beta)=\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ ← (後で示す.)

$0<t<1$
$\Updownarrow$
$0<u<\infty$
$\Updownarrow$
$-\infty<x<\infty$

$\boxed{\text{他の表示}}$ $B(\alpha,\beta)=\int_0^1 \dfrac{u^{\alpha-1}du}{(1+u)^{\alpha+\beta}}=2\int_0^{\pi/2}(\cos\theta)^{2\alpha-1}(\sin\theta)^{2\beta-1}d\theta$

$t=\dfrac{u}{1+u}=1-\dfrac{1}{1+u}$ ↗ $\qquad t=\cos^2\theta$ ↗

$\Updownarrow$

$u=\dfrac{t}{1-t}$ ← t が確率のとき, $\dfrac{t}{1-t}$ をオッズ (odds) と呼ぶ.

$x=\log u=\log\dfrac{t}{1-t}$ を logit変換と呼び,

その逆変換 $t=\dfrac{e^x}{1-e^x}=\dfrac{1}{1-e^{-x}}$ を logistic変換と呼ぶ.

$$\begin{cases}\mathrm{logit}(t)=\log\dfrac{t}{1-t}\\[2mm] \mathrm{logistic}(x)=\dfrac{1}{1+e^{-x}}\end{cases}$$

$\boxed{\text{ベータ分布の定義}}$ 次の確率密度函数が定める $0<t<1$ における分布をベータ分布と呼ぶ:

$$p(t|\alpha,\beta)=\frac{1}{B(\alpha,\beta)}t^{\alpha-1}(1-t)^{\beta-1}\quad(0<t<1).$$

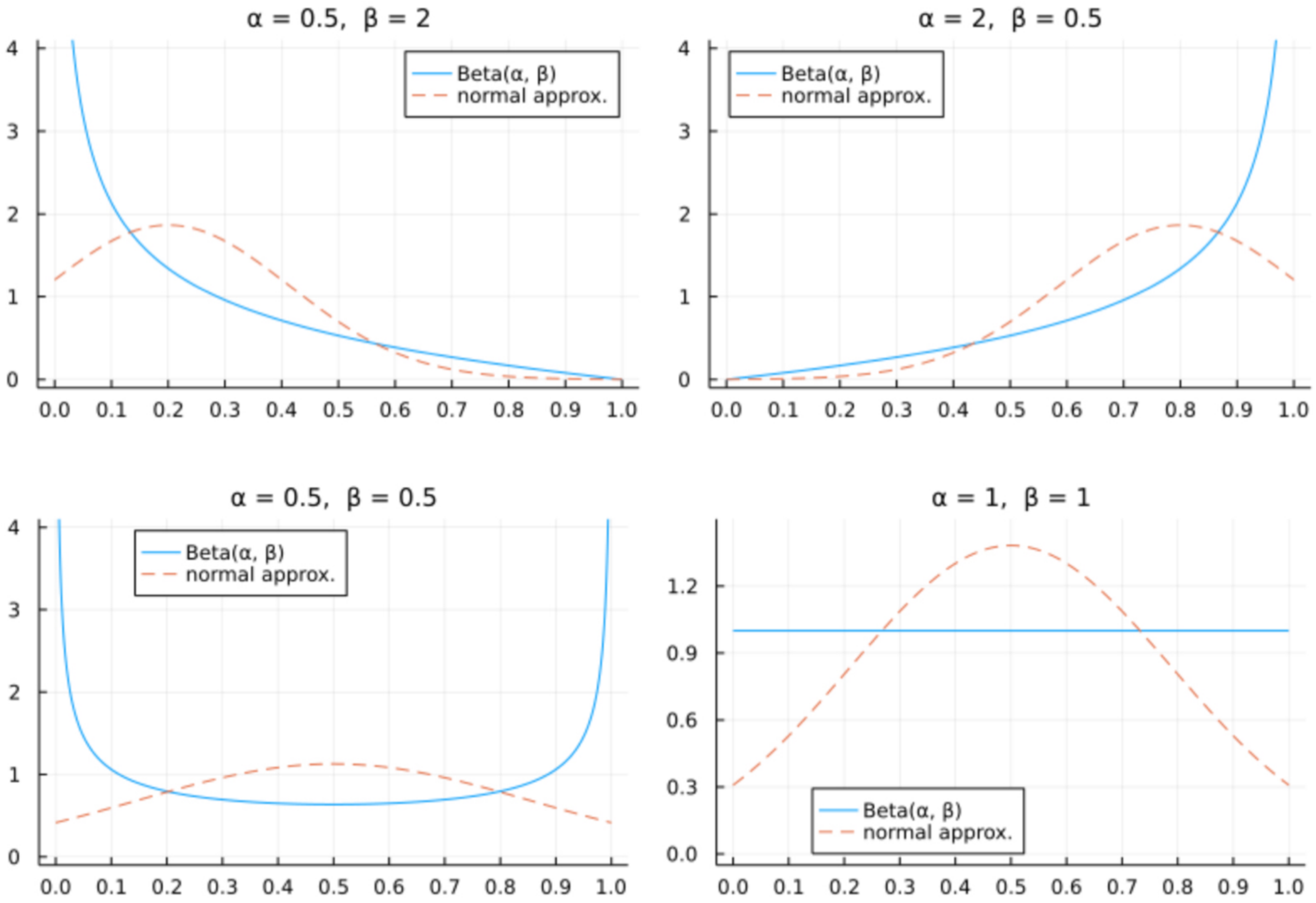
↑ $\mathrm{Beta}(\alpha,\beta)$ と書く

$\boxed{\text{自習問題}}$ ベータ函数のガンマ函数表示を用いて,

分布 $\mathrm{Beta}(\alpha,\beta)$ の期待値と分散がそれぞれ $\dfrac{\alpha}{\alpha+\beta}$, $\dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ になることを示せ.

Github においてある資料に答えがある.

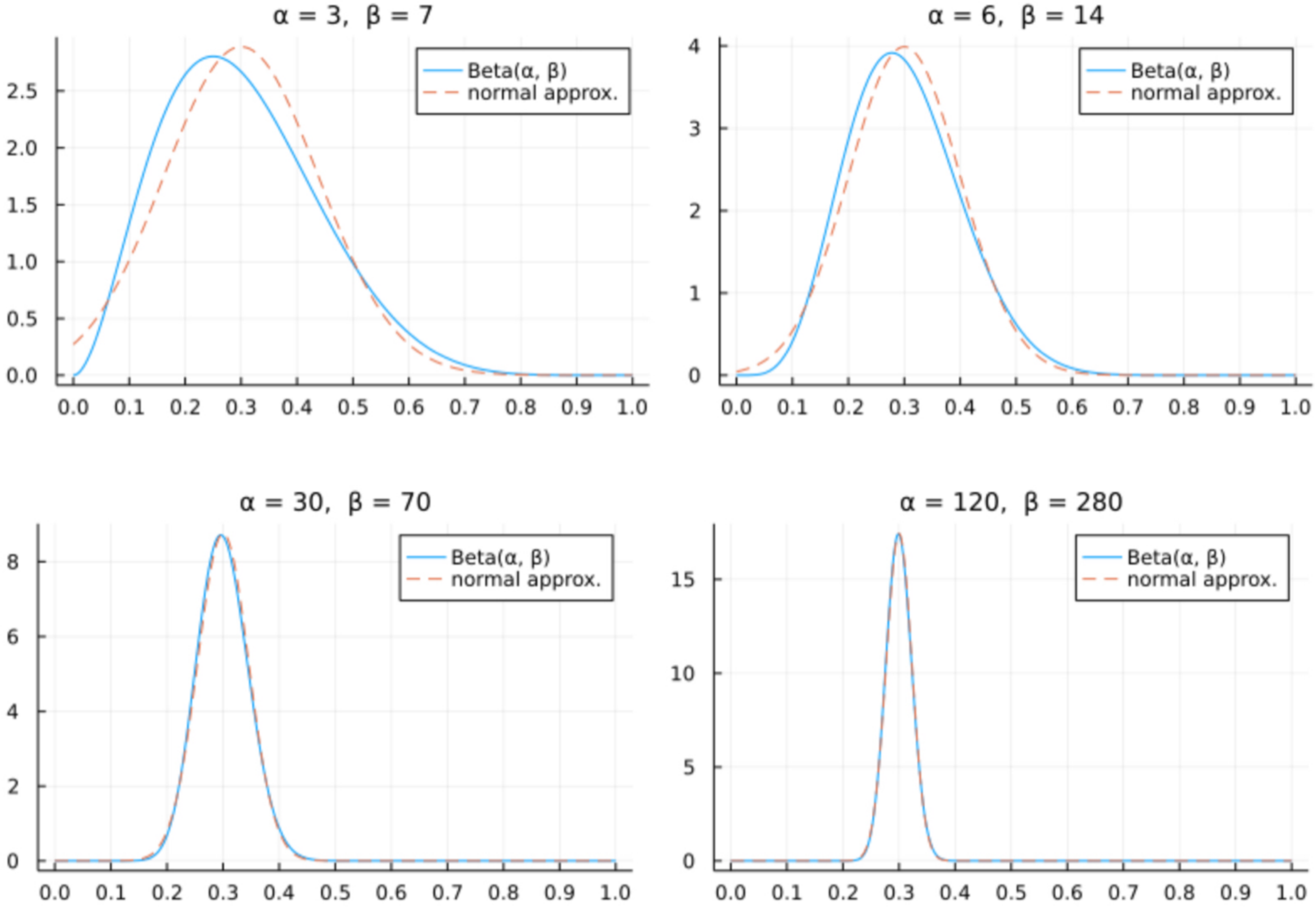
ベータ分布のグラフ



Uniform(0,1) と同じ

ベータ分布のグラフ(つづき)

パラメータを大きくすると正規分布で近似される.



ベータ函数の極限でガンマ函数が得られること

$$\lim_{n\to\infty} n^\alpha B(\alpha, n+b) = \Gamma(\alpha)$$

正規分布から外れた世界 $\xrightarrow{\text{極限}}$ 正規分布の世界

証明

$$t = \frac{x}{n}, \quad t^{\alpha-1}dt = \frac{x^{\alpha-1}}{n^\alpha}dx$$

$$n^\alpha B(\alpha, n+b) = n^\alpha \int_0^1 t^{\alpha-1}(1-t)^{n+b-1}dt = \int_0^n x^{\alpha-1}\left(1-\frac{x}{n}\right)^{n+b-1}dx,$$

$$\lim_{n\to\infty}\left(1-\frac{x}{n}\right)^{n+b-1} = e^{-x} \quad \left(\because \lim_{n\to\infty}\left(1-\frac{x}{n}\right)^n = e^{-x}, \ \lim_{n\to\infty}\left(1-\frac{x}{n}\right)^{b-1} = 1\right)$$

なので,
$$\lim_{n\to\infty} n^\alpha B(\alpha, n+b) = \int_0^\infty x^{\alpha-1}e^{-x}dx = \Gamma(\alpha). \qquad \square$$

例 $\nu\to\infty$ のとき,
$$\sqrt{\nu}\,B\left(\frac{1}{2}, \frac{\nu}{2}\right) = \sqrt{2}\left(\frac{\nu}{2}\right)^{\frac{1}{2}} B\left(\frac{1}{2}, \frac{\nu}{2}\right) \to \sqrt{2}\,\overbrace{\Gamma\left(\frac{1}{2}\right)}^{\sqrt{\pi}} = \sqrt{2\pi},$$

これは次のようにしても示せる:

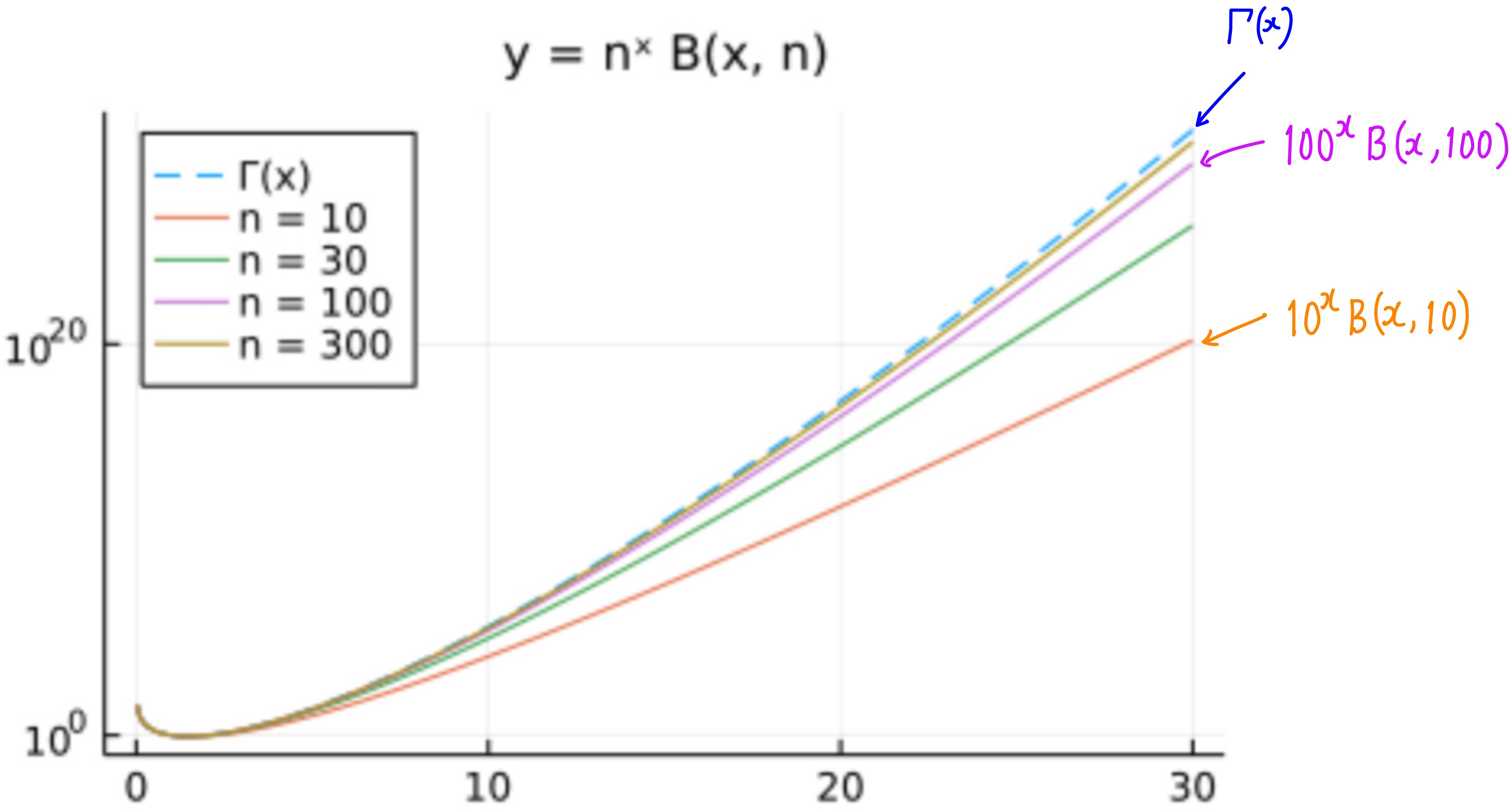
$$u = \frac{t^2}{\nu}, \quad \frac{du}{u} = 2\frac{dt}{t}, \quad u^{1/2} = \frac{t}{\sqrt{\nu}}$$

$$\sqrt{\nu}\,B\left(\frac{1}{2}, \frac{\nu}{2}\right) = \sqrt{\nu}\int_0^\infty \frac{u^{\frac{1}{2}-1}du}{(1+u)^{(\nu+1)/2}} = 2\int_0^\infty \left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}dt = \int_{-\infty}^\infty \left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}dt$$

後で使う

$$= \int_{-\infty}^\infty \left(1+\frac{t^2/2}{\nu/2}\right)^{-\nu/2-1/2}dt \xrightarrow{\nu\to\infty} \int_{-\infty}^\infty e^{-t^2/2}dt = \sqrt{2\pi}. \qquad \square$$

$y = n^x B(x, n)$ と $y = \Gamma(x)$ を対数スケールで同時プロット



$\boxed{\text{自習問題(易)}}$ $B(\alpha,\beta)=2\int_0^{\pi/2}(\cos\theta)^{2\alpha-1}(\sin\theta)^{2\beta-1}d\theta$ を示せ ($t=\cos^2\theta$ とおく).

答えはGithubにおいてある資料にある.

特に $B(\frac{1}{2},\beta)=2\int_0^{\pi/2}(\sin\theta)^{2\beta-1}d\theta$, $B(\alpha,\frac{1}{2})=2\int_0^{\pi/2}(\cos\theta)^{2\alpha-1}d\theta$.

$\sin\theta$ や $\cos\theta$ のべきの 0 から $\frac{\pi}{2}$ までの積分はベータ函数で計算できる.

$\boxed{\text{ベータ函数の別の基本表示}}$ (確率 t ではなくオッズ $u=\dfrac{t}{1-t}$ による表示)

$$\boxed{B(\alpha,\beta)=\int_0^\infty\frac{u^{\alpha-1}\,du}{(1+u)^{\alpha+\beta}}.}$$

重要!

$$\Updownarrow$$

$$t=\frac{u}{1+u}=1-\frac{1}{1+u}$$

$\boxed{\text{証明}}$ $B(\alpha,\beta)=\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt$ で $t=\dfrac{u}{1+u}=1-\dfrac{1}{1+u}$ とおくと,

$dt=\dfrac{du}{(1+u)^2}$ と $u=\dfrac{t}{1-t}$ より,

t	0	1
u	0	∞

より,

$$B(\alpha,\beta)=\int_0^\infty\left(\frac{u}{1+u}\right)^{\alpha-1}\left(\frac{1}{1+u}\right)^{\beta-1}\frac{du}{(1+u)^2}=\int_0^\infty\frac{u^{\alpha-1}\,du}{(1+u)^{\alpha+\beta}}.$$

$\square$

この計算は t分布 や F分布と深く関係している.

自習問題(易) $\displaystyle\int_{-\infty}^{\infty}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}dt = \sqrt{\nu}\,B\left(\frac{1}{2},\frac{\nu}{2}\right) \quad (\nu>0).$ 答えはGithubにおいてある資料にある.

ヒント (左辺) $\displaystyle = 2\int_0^{\infty}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}dt$ で $t=\sqrt{\nu u}$ とおいて, ベータ函数の別の基本表示を使う.

TDist(ν)と書く

t分布の定義 次の確率密度函数によって定義される分布を t分布 と呼ぶ:

$$p(t|\nu) = \frac{1}{\sqrt{\nu}\,B\left(\frac{1}{2},\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} \quad (t\in\mathbb{R}).$$

← νを自由度と呼ぶ.

ニュー

μ ミュー　ν ニュー

u ユー　v ヴイ

$\nu=1$のt分布をCauchy分布とよび Cauchy() と書く.

$$p(t|1) = \frac{1}{\pi}\frac{1}{1+t^2} \quad (t\in\mathbb{R})$$

t分布TDist(ν)は$\nu\to\infty$で標準正規分布に収束する

$$\lim_{\nu\to\infty} p(t|\nu) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}.$$

$\log\left(1+\frac{a}{n}\right)^{-n}$
$= -n\log\left(1+\frac{a}{n}\right)$
$= -n\left(\frac{a}{n}-\frac{a^2}{2n^2}+\cdots\right)$
$= -a+\frac{a^2}{2n}+\cdots$
$\longrightarrow -a$

証明 すでに $\displaystyle\lim_{\nu\to\infty}\sqrt{\nu}\,B\left(\frac{1}{2},\frac{\nu}{2}\right)=\sqrt{2\pi}$ を示した. その計算ですでに

$\nu\to\infty$のとき, $\displaystyle\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} = \left(1+\frac{t^2/2}{\nu/2}\right)^{-\frac{\nu}{2}-\frac{1}{2}} \longrightarrow e^{-\frac{t^2}{2}}$ となることを使っている.

$\left(1+\frac{a}{n}\right)^n \longrightarrow e^{-a} \quad (n\to\infty)$

□

t分布 TDist(ν)のグラフ

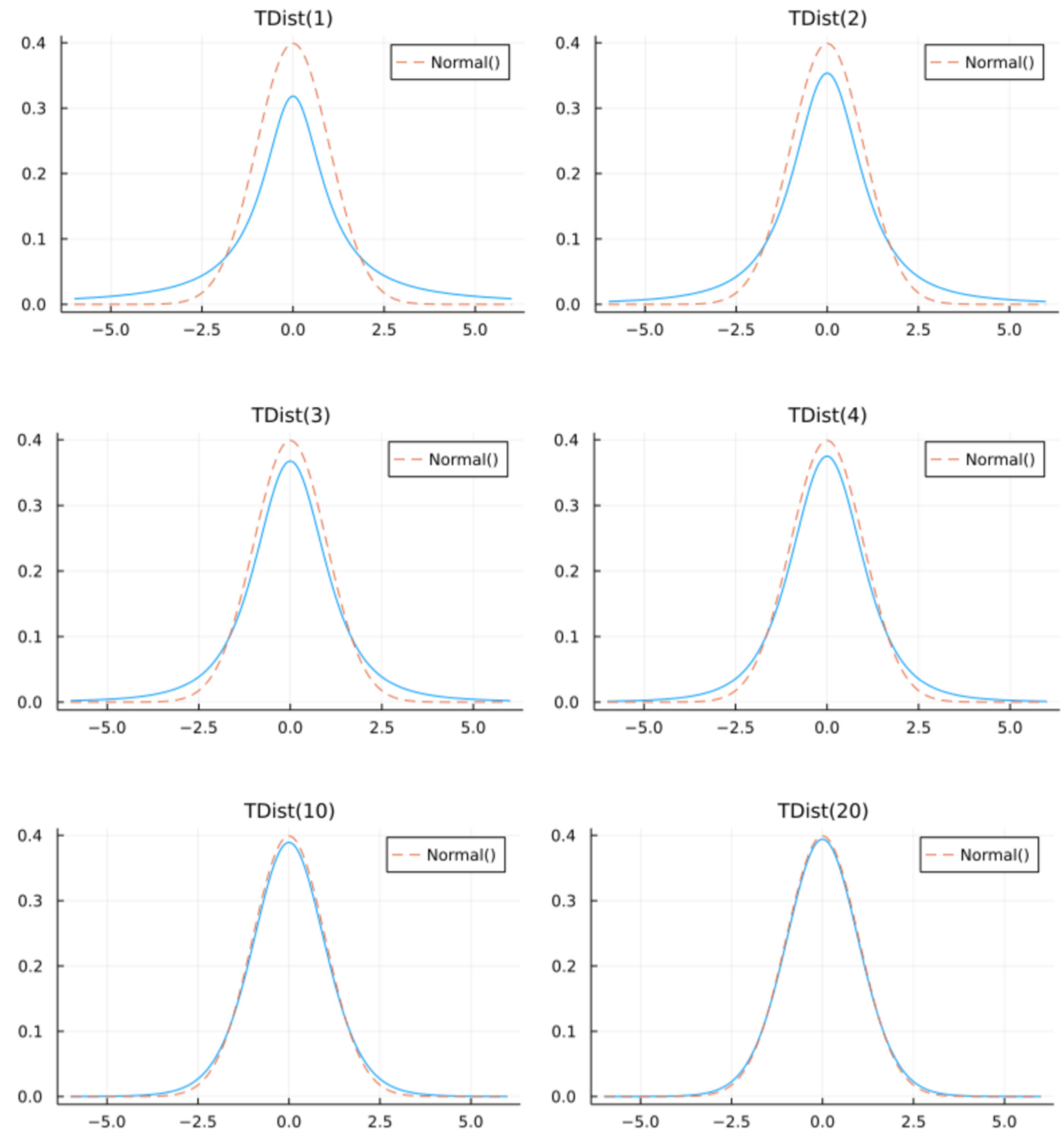
点線は標準正規分布

$$\frac{e^{-t^2/2}}{\sqrt{2\pi}}$$

実線はt分布

$$\frac{1}{\sqrt{\nu}\,B\left(\frac{1}{2},\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

自由度νを大きくすると,
t分布は標準正規分布
に近づく.



ベータ函数のガンマ函数表示

大学新入生向けの微積分の講義ですでにやっていると思うが確率分布と相性のよい方法でやりなおそう。

$\Gamma(\alpha)\Gamma(\beta) = \Gamma(\alpha+\beta)B(\alpha,\beta)$ を示したい。

積分 $J[f]$ に関する公式 函数 $f(z,t)$ に対して，$J[f]$ を次のように定める：

$$J[f] = \int_0^\infty \int_0^\infty f\left(x+y, \frac{x}{x+y}\right) e^{-x} x^{\alpha-1} \cdot e^{-y} y^{\beta-1}\,dx\,dy.$$

このとき，次の公式が成立している：

$$J[f] = \int_0^\infty \left(\int_0^1 f(z,t)\, e^{-z} z^{\alpha+\beta-1} \cdot t^{\alpha-1}(1-t)^{\beta-1}\,dt \right) dz. \qquad \square$$

後で証明する。

上の公式を $f=1$（$f(z,t)$ は z と t によらずに 1）の場合に適用すると，

$$J[1] = \int_0^\infty e^{-x} x^{\alpha-1}\,dx \cdot \int_0^\infty e^{-y} y^{\beta-1}\,dy = \Gamma(\alpha)\Gamma(\beta),$$

$$J[1] = \int_0^\infty e^{-z} z^{\alpha+\beta-1}\,dz \cdot \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt = \Gamma(\alpha+\beta)B(\alpha,\beta).$$

$$\therefore\ \Gamma(\alpha)\Gamma(\beta) = \Gamma(\alpha+\beta)B(\alpha,\beta).$$

J[f]に関する公式の証明

定義

$$
J[f] = \int_0^\infty \int_0^\infty f\left(x+y, \frac{x}{x+y}\right) e^{-x} x^{\alpha-1} \cdot e^{-y} x^{\beta-1}\, dx\, dy
$$

$$
= \int_0^\infty \left(\int_0^\infty f\left(x+y, \frac{x}{x+y}\right) e^{-(x+y)} x^{\alpha-1} y^{\beta-1}\, dy \right) dx
\qquad
\begin{cases} y = z - x, & dy = dz \\ y > 0 \iff z > x \end{cases}
$$

$$
= \int_0^\infty \left(\int_x^\infty f\left(z, \frac{x}{z}\right) e^{-z} x^{\alpha-1} (z-x)^{\beta-1}\, dz \right) dx
$$

積分領域は $0 < x < z < \infty$

$$
= \int_0^\infty \left(\int_0^z f\left(z, \frac{x}{z}\right) e^{-z} x^{\alpha-1} (z-x)^{\beta-1}\, dx \right) dz
\qquad
\begin{cases} x = zt, & dx = z\, dt \\ 0 < x < z \iff 0 < t < 1 \end{cases}
$$

$$
= \int_0^\infty \left(\int_0^1 f(z, t)\, e^{-z} (zt)^{\alpha-1} (z(1-t))^{\beta-1} z\, dt \right) dz
$$

$$
= \int_0^\infty \left(\int_0^1 f(z, t)\, e^{-z} z^{\alpha+\beta-1} \cdot t^{\alpha-1} (1-t)^{\beta-1}\, dt \right) dz
$$

q.e.d.

この結果はガンマ分布からベータ分布を作る方法を与える。

確率変数達の同時確率密度函数と独立性

同時確率質量函数も同様に定める．

後で再度説明する予定

確率変数達 $X_1,\dots,X_n$ が同時確率密度函数 $p(x_1,\dots,x_n)$ を持つとは

$$E[f(X_1,\dots,X_n)] = \int\cdots\int f(x_1,\dots,x_n)\,p(x_1,\dots,x_n)\,dx_1\cdots dx_n \quad (\text{定積分})$$

が成立することだと定める．このとき

$$p_i(x_i) = \int\cdots\int p(x_1,\dots,x_n)\,dx_1\cdots\widehat{dx_i}\cdots dx_n \quad (\widehat{dx_i}\text{ は }dx_i\text{ を除くという意味})$$

とおくと，$p_i(x_i)$ は X_i 単独の確率密度函数になる：

$$E[f_i(X_i)] = \int\cdots\int f_i(x_i)\,p(x_1,\dots,x_n)\,dx_1\cdots dx_n$$

$$= \int f_i(x_i)\left(\int\cdots\int p(x_1,\dots,x_n)\,dx_1\cdots\widehat{dx_i}\cdots dx_n\right)dx_i = \int f_i(x_i)\,p_i(x_i)\,dx_i.$$

確率変数達 $X_1,\dots,X_n$ が独立であるとは，

$$p(x_1,\dots,x_n) = p_1(x_1)\cdots p_n(x_n)$$ が成立することだと定める．

同時確率質量函数の場合も同様

このとき，次が成立する：

$$E[f_1(X_1)\cdots f_n(X_n)] = \int\cdots\int f_1(x_1)\cdots f_n(x_n)\,p_1(x_1)\cdots p_n(x_n)\,dx_1\cdots dx_n$$

$$= \int f_1(x_1)\,p_1(x_1)\,dx_1\cdots\int f_n(x_n)\,p_n(x_n)\,dx_n$$

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

確率変数の独立性は，多くの場合，この形で使う．

ベータ分布のガンマ分布表示

これを使うとガンマ分布の乱数からベータ分布の乱数を作れる.

$$\left.\begin{array}{l} X \sim \mathrm{Gamma}(\alpha, 1),\ Y \sim \mathrm{Gamma}(\beta, 1) \\ \text{確率変数達 } X, Y \text{ は独立} \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} X+Y \sim \mathrm{Gamma}(\alpha+\beta, 1),\ \dfrac{X}{X+Y} \sim \mathrm{Beta}(\alpha, \beta) \\ X+Y,\ \dfrac{X}{X+Y} \text{ は独立} \end{array}\right.$$

証明 $\Rightarrow$ の左辺より,

$$E\left[f\left(X+Y, \frac{X}{X+Y}\right)\right] = \int_0^\infty \int_0^\infty f\left(x+y, \frac{x}{x+y}\right) \frac{e^{-x} x^{\alpha-1}}{\Gamma(\alpha)} \frac{e^{-y} y^{\beta-1}}{\Gamma(\beta)}\,dx\,dy = \frac{J[f]}{\Gamma(\alpha)\Gamma(\beta)}.$$

$J[f]$ に関する公式と $\Gamma(\alpha)\Gamma(\beta) = \Gamma(\alpha+\beta)B(\alpha, \beta)$ より,

$$\frac{J[f]}{\Gamma(\alpha)\Gamma(\beta)} = \frac{J[f]}{\Gamma(\alpha+\beta)B(\alpha,\beta)} = \int_0^\infty \left(\int_0^1 f(z, t) \frac{e^{-z} z^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)} \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha,\beta)}\,dt\right) dz.$$

これは $\Rightarrow$ の右辺の成立を意味する. □

注意 上の逆向き $\Leftarrow$ も同様に成立する. □

Githubにおいてある資料では実際にこれを実行している.

注意 仮に, ガンマ函数の乱数を

$x = \mathrm{rand}(\mathrm{Gamma}(\alpha, 1))$, $y = \mathrm{rand}(\mathrm{Gamma}(\beta, 1))$ と生成できるならば

$z = x + y$ は $\mathrm{Gamma}(\alpha+\beta, 1)$ にしたがう乱数になり,

$t = \dfrac{x}{z} = \dfrac{x}{x+y}$ は $\mathrm{Beta}(\alpha, \beta)$ にしたがう乱数になる. □

積分 $K[g]$ に関する公式　函数 $g(z,u)$ について

$$K[g] = \int_0^\infty\int_0^\infty g\left(x+y, \frac{x}{y}\right) e^{-x}x^{\alpha-1}\cdot e^{-y}y^{\beta-1}\,dx\,dy$$

と定めると,

$$K[g] = \int_0^\infty\int_0^\infty g(z,u)\, e^{-z}z^{\alpha+\beta-1}\cdot\frac{u^{\alpha-1}}{(1+u)^{\alpha+\beta}}\,du\,dz$$

$u = \dfrac{t}{1-t}$ (オッズ)

$\Updownarrow$

$t = \dfrac{u}{1+u}$ (確率)

証明　$f(z,t) = g\left(z, \dfrac{t}{1-t}\right)$ と定めると, $g(z,u) = f\left(z, \dfrac{u}{1+u}\right)$.

$g\left(x+y, \dfrac{x}{y}\right) = f\left(x+y, \dfrac{x/y}{1+x/y}\right) = f\left(x+y, \dfrac{x}{x+y}\right)$ なので, $K[g] = J[f]$.

そして, $J[f]$ の公式で $t = \dfrac{u}{1+u} = 1 - \dfrac{1}{1+u}$ とおくと, $0<t<1 \Leftrightarrow 0<u \Leftrightarrow \infty$ で

$$f(z,t)\, t^{\alpha-1}(1-t)^{\beta-1}\,dt = f\left(z, \frac{u}{1+u}\right)\left(\frac{u}{1+u}\right)^{\alpha-1}\left(\frac{1}{1+u}\right)^{\beta-1}\frac{du}{(1+u)^2} = g(z,t)\frac{u^{\alpha-1}}{(1+u)^{\alpha+\beta}}\,du$$

なので, $K[g] = J[f] = \displaystyle\int_0^\infty\int_0^\infty g(z,u)\, e^{-z}z^{\alpha+\beta-1}\cdot\frac{u^{\alpha-1}}{(1+u)^{\alpha+\beta}}\,du\,dz.$　q.e.d.

自習問題　以上の証明をノートにもっと詳しい形式で書いて内容を確認せよ.

私の経験では証明はすべて自分の手で書かないと理解できない. (大量に時間がとられる!)

ガンマ分布にしたがう独立な確率変数の商がしたがう分布

$X \sim \mathrm{Gamma}(\alpha,1)$, $Y \sim \mathrm{Gamma}(\beta,1)$ で X, Y が独立なとき,

$U = \dfrac{X}{Y}$ は次の確率密度函数を持つ:

$$p(u|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} \frac{u^{\alpha-1}}{(1+u)^{\alpha+\beta}} \quad (u>0).$$

すなわち, $E[g(U)] = \int_0^\infty g(u)\, p(u|\alpha,\beta)\, du$ が成立する.

証明 $T = \dfrac{X}{X+Y} \sim \mathrm{Beta}(\alpha,\beta)$ となるので

$$E[f(T)] = \frac{1}{B(\alpha,\beta)} \int_0^1 f(t)\, t^{\alpha-1}(1-t)^{\beta-1}\, dt.$$

そして, $\dfrac{T}{1-T} = \dfrac{X}{Y} = U$ なので,

$$\begin{aligned}
E[g(U)] &= \frac{1}{B(\alpha,\beta)} \int_0^1 g\left(\frac{t}{1-t}\right) t^{\alpha-1}(1-t)^{\beta-1}\, dt \\
&= \frac{1}{B(\alpha,\beta)} \int_0^\infty g(u) \left(\frac{u}{1+u}\right)^{\alpha-1} \left(\frac{1}{1+u}\right)^{\beta-1} \frac{du}{(1+u)^2} \\
&= \frac{1}{B(\alpha,\beta)} \int_0^\infty g(u) \frac{u^{\alpha-1}}{(1+u)^{\alpha+\beta}}\, du \\
&= \int_0^\infty g(u)\, p(u|\alpha,\beta)\, du.
\end{aligned}$$

q.e.d.

自習問題

ノート上に詳しい証明を自分の手で書き, 内容を確認せよ.

以上のような計算は後で平均に関する t 検定や信頼区間を理解するために役に立つ.

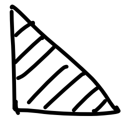
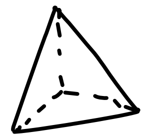
注 U がしたがう分布はスケール変換のちがいを除いて F分布 に等しい.
F検定で使われる.

注意 $Z = X+Y$ と $T = \dfrac{X}{Z}$ だけではなく, $Z = X+Y$ と $U = \dfrac{X}{Y}$ も独立になる. □

$\boxed{\text{Dirichlet 積分}}$ (ベータ函数の多変数化)

$$\Delta_{n-1} = \{(t_1, \dots, t_{n-1}) \in \mathbb{R}^{n-1} \mid t_1, \dots, t_{n-1} > 0,\ t_1 + \dots + t_{n-1} < 1\}$$ とおく。

四面体の内側

$\Delta_1 = \{t_1 \in \mathbb{R} \mid 0 < t_1 < 1\}$, $\Delta_2 = \{(t_1, t_2) \in \mathbb{R}^2 \mid t_1, t_2 > 0,\ t_1 + t_2 < 1\} =$ , $\Delta_3 =$ 

$=$ 

$$B(\alpha_1, \dots, \alpha_n) = \int \cdots \int_{\Delta_{n-1}} t_1^{\alpha_1 - 1} \cdots t_{n-1}^{\alpha_{n-1}-1} (1 - (t_1 + \dots + t_{n-1}))^{\alpha_n - 1} dt_1 \cdots dt_{n-1}$$

← 多変数ベータ函数。

$(\alpha_1, \dots, \alpha_{n-1}, \alpha_n > 0)$

$$= \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{n-1}) \Gamma(\alpha_n)}{\Gamma(\alpha_1 + \dots + \alpha_{n+1} + \alpha_n)}$$

← 証明は Github においてある資料にある。

$\boxed{\text{自習問題}}$ この結果をみとめて以下を示せ。

答えは Github においてある資料にある。

$$X = \left\{(x_1, \dots, x_n) \in \mathbb{R}^n_{>0} \;\middle|\; \left(\frac{x_1}{a_1}\right)^{p_1} + \dots + \left(\frac{x_n}{a_n}\right)^{p_n} < 1\right\}$$ とおくとき,

$$\int \cdots \int_X x_1^{\alpha_1 - 1} \cdots x_n^{\alpha_n - 1} dx_1 \cdots dx_n = \frac{a_1^{\alpha_1} \cdots a_n^{\alpha_n}}{p_1 \cdots p_n} \frac{\Gamma(\alpha_1/p_1) \cdots \Gamma(\alpha_n/p_n)}{\Gamma(\alpha_1/p_1 + \dots + \alpha_n/p_n + 1)}.$$

$\boxed{\text{Dirichlet 分布}}$ Dirichlet$(\alpha_1, \dots, \alpha_n)$

$$p(t_1, \dots, t_{n-1} \mid \alpha_1, \dots, \alpha_n) = \frac{t_1^{\alpha_1 - 1} \cdots t_{n-1}^{\alpha_{n-1}-1} (1 - (t_1 + \dots + t_{n-1}))^{\alpha_n - 1}}{B(\alpha_1, \dots, \alpha_n)}$$

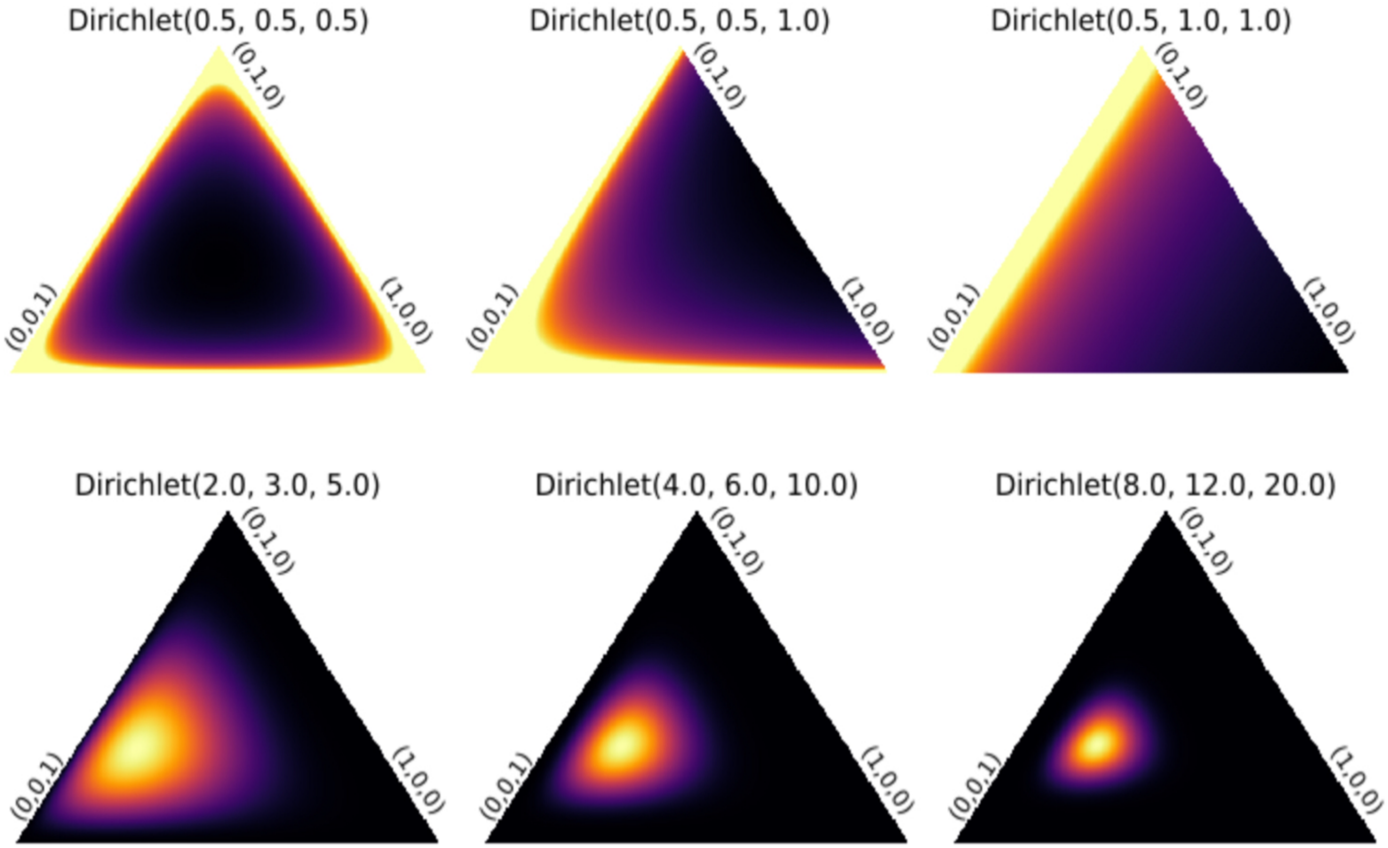
← 密度函数

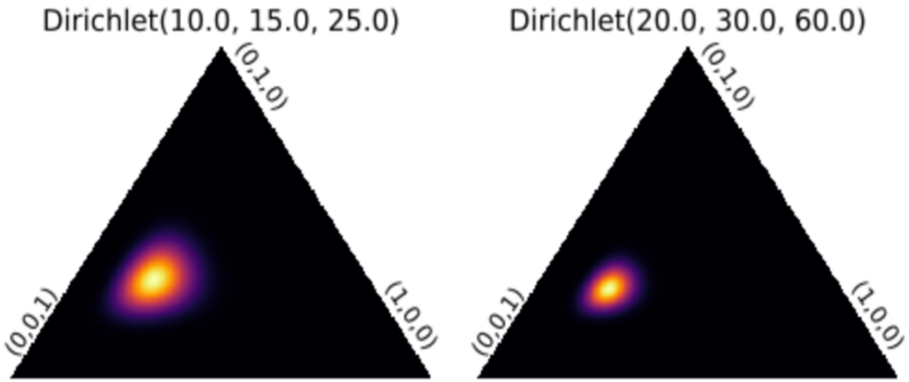
$\left(\begin{array}{l}(t_1, \dots, t_{n-1}) \\ \in \Delta_{n-1}\end{array}\right)$.

Dirichlet分布のヒートマップ

明るいところほど確率密度が高い。

ヒートマップの方が3次元のグラフより見易い場合がある。



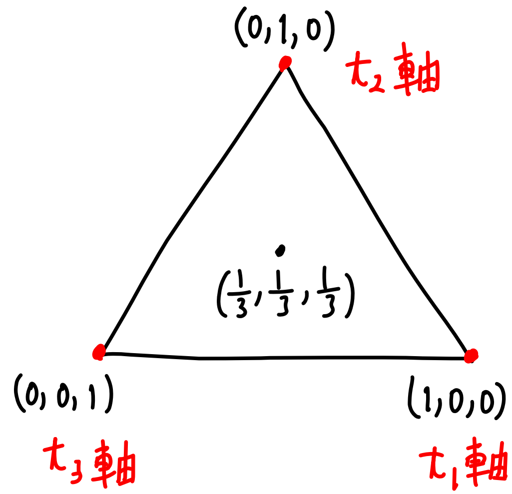


Dirichlet分布はパラメータ→大で多変量正規分布に近づく。

$t_1, t_2 > 0$ かつ

$$t_3 = 1-(t_1+t_2) > 0$$

(t_1, t_2, t_3)



① 確率変数 X が確率密度函数 $p(x)$ を持つ $\longleftarrow$ $X \sim p(x)$

$$\Longleftrightarrow \quad E[f(X)] = \int f(x)\,p(x)\,dx \quad (\text{定積分})$$

$X_1, \ldots, X_n \sim p(x_1, \ldots, x_n)$

② 確率変数達 $X_1, \ldots, X_n$ が(同時)確率密度函数 $p(x_1, \ldots, x_n)$ を持つ

$$\Longleftrightarrow \quad E[f(X_1, \ldots, X_n)] = \int \cdots \int f(x_1, \ldots, x_n)\,p(x_1, \ldots, x_n)\,dx_1 \cdots dx_n \quad (\text{定積分})$$

そのときの X_i 単独の確率密度函数 $p_i(x_i)$ は次になる:

$$p_i(x_i) = \int \cdots \int p(x_1, \ldots, x_n)\,dx_1 \cdots \widehat{dx_i} \cdots dx_n \quad (\widehat{dx_i} \text{ は } dx_i \text{ を除くという意味}).$$

(同時)確率質量函数についても同様

さらに,

③ $X_1, \ldots, X_n$ は独立である $\longleftarrow$ $X_i \sim p_i(x_i)$ and $X_1, \ldots, X_n$ are independent.

$$\Longleftrightarrow \quad p(x_1, \ldots, x_n) = p_1(x_1) \cdots p_n(x_n)$$

$$\Longleftrightarrow \quad E[f_1(X_1) \cdots f_n(X_n)] = E[f_1(X_1)] \cdots E[f_n(X_n)]$$ $\longleftarrow$ 確率変数の独立性はこの形で使う.

④ 確率変数の列 $Y_1, Y_2, \ldots$ の分布が確率変数 Y_∞ の分布に収束する (分布収束)

$$\Longleftrightarrow \quad \lim_{n \to \infty} E[f(Y_n)] = E[f(Y_\infty)].$$

分布の収束 $\Longleftrightarrow$ 確率変数の函数の期待値の収束

$\boxed{例}$ 試行回数 n, 成功確率 p の Bernoulli 試行の分布にしたがう確率変数 $X_1, \dots, X_n$ について,

$$E[f(X_1,\dots,X_n)] = \sum_{x_1=1,0} \cdots \sum_{x_n=1,0} f(x_1,\dots,x_n) \underbrace{\prod_{i=1}^{n} \left(p^{x_i}(1-p)^{1-x_i}\right)}_{\text{積の形}}$$

なので, $X_i \sim \mathrm{Bernoulli}(p)$ かつ $X_1, \dots, X_n$ は独立である.

$\boxed{例}$ $X \sim \mathrm{Gamma}(\alpha, 1)$, $Y \sim \mathrm{Gamma}(\beta, 1)$ で X と Y は独立

$$\Longleftrightarrow \quad E[f(X,Y)] = \int_0^\infty \int_0^\infty f(x,y) \underbrace{\frac{e^{-x} x^{\alpha-1}}{\Gamma(\alpha)} \frac{e^{-y} y^{\beta-1}}{\Gamma(\beta)}}_{\text{積の形}} \, dx\, dy$$

このとき,

$$E\left[g\left(X+Y, \frac{X}{X+Y}\right)\right] = \int_0^\infty \int_0^\infty g\left(x+y, \frac{x}{x+y}\right) \frac{e^{-x} x^{\alpha-1}}{\Gamma(\alpha)} \frac{e^{-y} y^{\beta-1}}{\Gamma(\beta)} \, dx\, dy$$

ガンマ函数とベータ函数の関係

$$= \int_0^\infty \left(\int_0^1 g(z,t) \underbrace{\frac{e^{-z} z^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)} \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha,\beta)}}_{\text{積の形}} \, dt \right) dz.$$

ゆえに, $X+Y \sim \mathrm{Gamma}(\alpha+\beta, 1)$, $\dfrac{X}{X+Y} \sim \mathrm{Beta}(\alpha, \beta)$ かつ $X+Y$ と $\dfrac{X}{X+Y}$ は独立.

$\boxed{例}$ $X_i \sim \mathrm{Normal}(\mu, \sigma)$ かつ $X_1, \dots, X_n$ は独立

$$\Longleftrightarrow \quad E[f(X_1,\dots,X_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1,\dots,x_n) \prod_{i=1}^{n} \left(\frac{e^{-(x_i-\mu)^2/(2\sigma^2)}}{\sqrt{2\pi\sigma^2}} \right) dx_1 \cdots dx_n.$$

例 (二項分布のPoisson分布への収束) 後で説明する予定

$$K_n \sim \mathrm{Binomial}\left(n, \frac{\lambda}{n}\right),\ K_\infty \sim \mathrm{Poisson}(\lambda) \Rightarrow \lim_{n\to\infty} E[f(K_n)] = E[f(K_\infty)]$$

例 (負の二項分布のガンマ分布への収束) pの小さな負の二項分布はガンマ分布で近似される.

$$M_L \sim \mathrm{NegativeBinomial}\left(\alpha, \frac{1}{L\theta}\right),\ T_L = \frac{M_L + k}{L},\ T_\infty \sim \mathrm{Gamma}(\alpha, \theta) \Rightarrow \lim_{L\to\infty} E[f(T_L)] = E[T_\infty].$$

例 (二項分布の中心極限定理) 後で解説する予定

$$K_n \sim \mathrm{Binomial}(n, p),\ Z_n = \frac{K_n - np}{\sqrt{n}},\ Z_\infty \sim \mathrm{Normal}(0, p(1-p)) \Rightarrow \lim_{n\to\infty} E[f(Z_n)] = E[f(Z_\infty)].$$

例 (ガンマ分布の中心極限定理) ガンマ分布はαが大きいとき正規分布で近似される.

$$X_\alpha \sim \mathrm{Gamma}(\alpha, \theta),\ Z_\alpha = \frac{X_\alpha - \alpha\theta}{\sqrt{\alpha\theta^2}},\ Z_\infty \sim \mathrm{Normal}(0,1) \Rightarrow \lim_{\alpha\to\infty} E[f(Z_\alpha)] = E[f(Z_\infty)].$$

例 (ベータ分布の正規分布近似) $0<p<1$とする. ベータ分布はパラメータが大きいとき正規分布で近似される.

$$T_n \sim \mathrm{Beta}(np, n(1-p)),\ Z_n = \frac{\sqrt{n}\,(T_n - p)}{p(1-p)},\ Z_\infty \sim \mathrm{Normal}(0,1) \Rightarrow \lim_{n\to\infty} E[f(Z_n)] = E[f(Z_\infty)].$$

例 (一般の中心極限定理) X_iたちは同じ分布にしたがい独立であるとし, $E[X_i] = \mu$が定義され, $E[(X_i - \mu)^2] = \sigma^2 < \infty$, $E[|X_i - \mu|^3] < \infty$であるとする. このとき,

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma},\ Z_\infty \sim \mathrm{Normal}(0,1) \Rightarrow \lim_{n\to\infty} E[f(Z_n)] = E[f(Z)].$$

後で解説する.

確率分布達の解釈

$\boxed{1}$ 正規分布の使い方について少し説明した後に Poisson分布を定義する.

$\boxed{2}$ Poisson分布とガンマ分布の以下の解釈について説明

離散時間の場合 (Bernoulli試行) $\xrightarrow{\text{連続時間極限}}$ 連続時間の場合

単位時間にL回試行

単位時間のあいだに平均 λ 回 $\underbrace{\text{イベント発生}}_{\text{1が出る}}$

単位時間のあいだに平均 λ 回イベントが起こる

二項分布 $K \sim \mathrm{Binomial}(L, \lambda/L)$ $\xrightarrow{L \to \infty}$ Poisson分布 $K \sim \mathrm{Poisson}(\lambda)$

L回の試行中のイベント発生回数Kの分布

単位時間のあいだに起こるイベントの回数Kの分布

イベント発生の時間的間隔の期待値は $\theta = \frac{1}{\lambda}$

イベント発生の時間的間隔の期待値は $\theta = \frac{1}{\lambda}$

負の二項分布

$M \sim \mathrm{NegativeBinomial}(\alpha, 1/(L\theta))$

$N := \alpha + M$

$= \left(\text{α回イベントが起こるまでの試行回数}\right)$

$\frac{N}{L} \xrightarrow{L \to \infty}$ (ちょうどα回イベントが起こるまでの時間) $= T$,

ガンマ分布

$T \sim \mathrm{Gamma}(\alpha, \theta)$

$\boxed{3}$ ベータ分布の一様乱数発生のくりかえしによる解釈

1 正規分布の使い方について少し説明した後にPoisson分布を定義する.

正規分布の使い方

正規分布に限らない一般論

常に $\sigma>0$ とする

確率変数 X の期待値と分散がそれぞれ μ, σ^2 のとき,

$\frac{X-\mu}{\sigma}$ の期待値と分散はそれぞれ 0 と 1 になる. ← $\frac{X-\mu}{\sigma}$ を X の標準化と呼ぶ.

証明 $E[X]=\mu$, $\mathrm{Var}(X)=E[(X-\mu)^2]=\sigma^2$ のとき, $E[\ ]$ の線形性と規格化条件より,

$$E\left[\frac{X-\mu}{\sigma}\right]=\frac{E[X]-\mu}{\sigma}=0.\quad \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right)=E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right]=\frac{E[(X-\mu)^2]}{\sigma^2}=\frac{\sigma^2}{\sigma^2}=1.$$ q.e.d.

一般の正規分布と標準正規分布の関係

$a, b\in\mathbb{R}, a>0$ とする.

Githubの資料に書いてある.

X が正規分布にしたがう確率変数ならば $aX+b$ も正規分布にしたがう(証明略).

$X\sim \mathrm{Normal}(\mu,\sigma)$ のとき, $Z=\frac{X-\mu}{\sigma}$ とおくと $Z\sim\mathrm{Normal}(0,1)$. ← 標準正規分布.

このことを使うと, 一般の正規分布に関する計算を標準正規分布に帰着できる.

具体的な数値の計算(コンピュータでの実装)でもこのことを使う.

正規分布に限らない一般論 X, Y が独立な確率変数ならば

$$E[X+Y] = E[X]+E[Y], \quad \mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y).$$

無相関に条件をゆるめられる.

証明 $E[X+Y]=E[X]+E[Y]$ は $E[\ ]$ の線形性の特別な場合にすぎない. $\mu_X = E[X]$, $\mu_Y = E[Y]$ とおく.

$$\mathrm{Var}(X+Y) = E\left[(X+Y-(\mu_X+\mu_Y))^2\right] = E\left[((X-\mu_X)+(Y-\mu_Y))^2\right] = E\left[(X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2\right]$$

$$= \underbrace{E[(X-\mu_X)^2]}_{=\mathrm{Var}(X)} + 2\underbrace{E[(X-\mu_X)(Y-\mu_Y)]}_{\substack{= \text{X と Y の独立性} \\ \underbrace{E[X-\mu_X]}_{=0}\underbrace{E[Y-\mu_Y]}_{=0} = 0.}} + \underbrace{E[(Y-\mu_Y)^2]}_{=\mathrm{Var}(Y)}$$

X と Y が無相関なら独立でなくても 0

$$= \mathrm{Var}(X) + \mathrm{Var}(Y)$$

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$

三平方の定理

正規分布に関する自習問題

Github資料に解答例がある.

確率変数達 X, Y は独立で(同時確率密度函数がそれぞれの確率密度函数の積になっており),

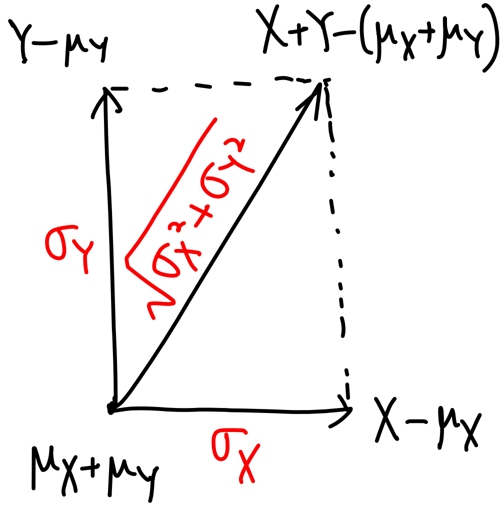
$$X \sim \mathrm{Normal}(\mu_X, \sigma_X), \quad Y \sim \mathrm{Normal}(\mu_Y, \sigma_Y)$$

となっているならば,

$$X+Y = \mathrm{Normal}\left(\mu_X+\mu_Y, \sqrt{\sigma_X^2+\sigma_Y^2}\right)$$

三平方の定理

となることを示せ.



最重要自習問題

基本中の基本!

必ず解いて欲しい問題: 正規分布における確率が95%または99%になる区間

$X \sim \mathrm{Normal}(\mu, \sigma)$ であると仮定する. このとき, X が区間 $[\mu - c\sigma, \mu + c\sigma]$ に含まれる確率

$$P(\mu - c\sigma \leq X \leq \mu + c\sigma)$$

答えはGithubにある.

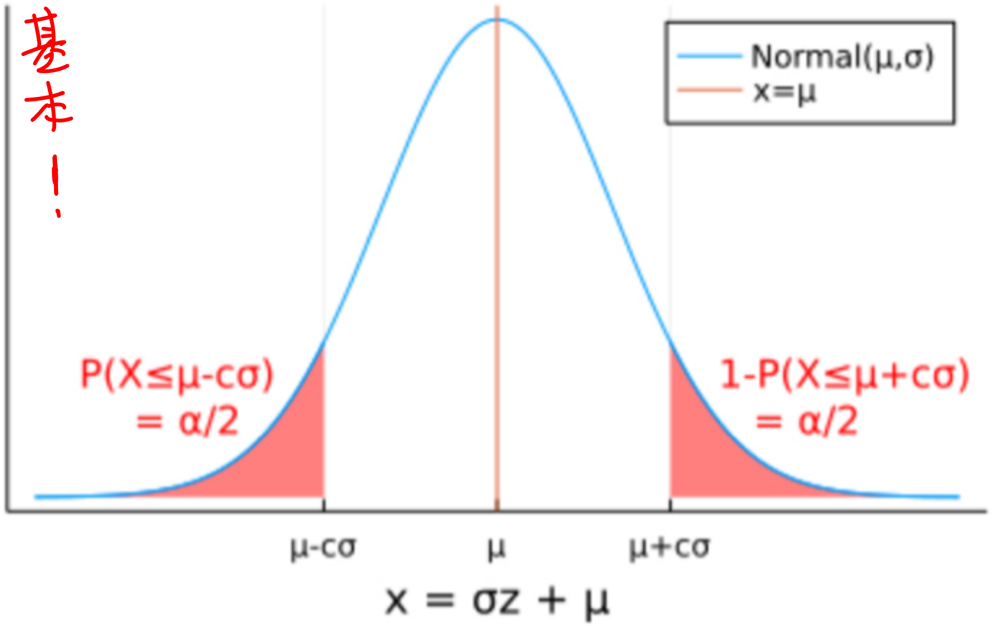
が $1 - \alpha$ に等しくなるような c を誤差函数

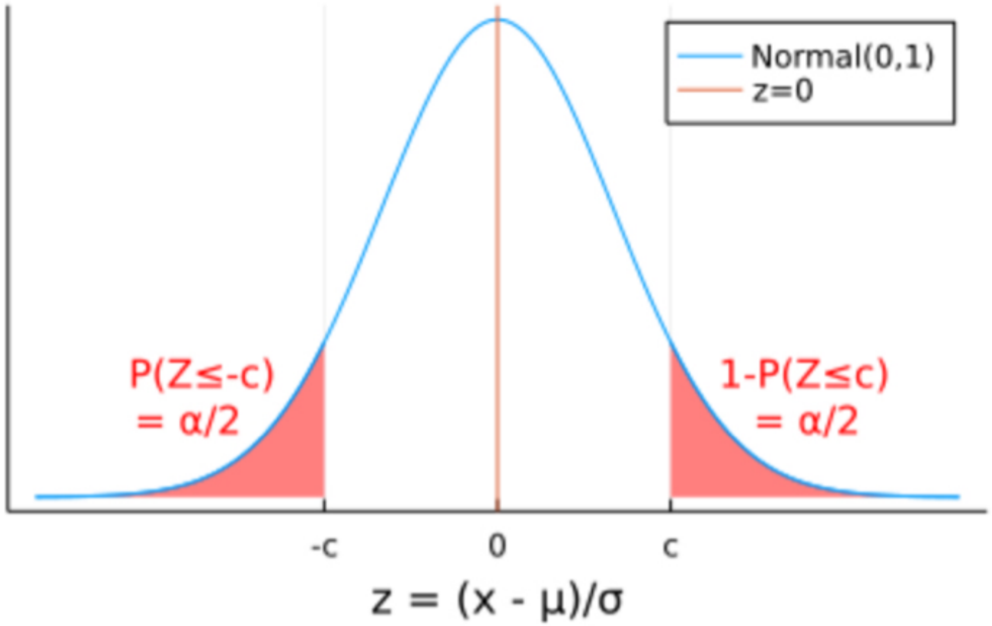
$$y = \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-u^2)\,du$$

の逆函数 $x = \mathrm{erfinv}(y)$ を使って表せ. (正規分布の定義に戻って地道に計算せよ.)

さらに, 上の確率が 95% になる c と 99% になる c を小数点以下第2桁目まで求めよ.

これだけは必ず解いておくこと





問題: 分散がχ^2分布の ν 分の1に従う正規分布は自由度 ν の t 分布になる

これの答えは
すぐに説明する.

$\nu > 0$ だと仮定する. 確率変数達 Y, Z は独立であるとし(同時確率密度函数がそれぞれの密度函数の積になる),

$$Z \sim \mathrm{Normal}(0,1), \quad Y \sim \mathrm{Chisq}(\nu) = \mathrm{Gamma}(\nu/2, 2)$$

と仮定する. $\mathrm{Chisq}(\nu)$ の期待値は $(\nu/2)2 = \nu$ になり, 分散は $(\nu/2)2^2 = 2\nu$ になり, 標準偏差は $\sqrt{2\nu}$ になるので, ν が 2 より大きければ, 標準偏差は期待値よりも小さくなる. このとき,

$$T = \frac{Z}{\sqrt{Y/\nu}} \sim \mathrm{TDist}(\nu)$$

となることを示せ.

標準正規分布は上の形でχ^2分布と組み合わせてt分布にして使うことがある.

$Y \sim \mathrm{Chisq}(\nu)$ のとき, $E[Y] = \nu$, $\mathrm{std}(Y) = \sqrt{2\nu}$ なので $E\left[\frac{Y}{\nu}\right] = 1$, $\mathrm{std}\left(\frac{Y}{\nu}\right) = \sqrt{\frac{2}{\nu}}$.

標準偏差 $= \sqrt{\text{分散}}$

もしも $\frac{Y}{\nu}$ が定数なら, $\frac{Z}{\sqrt{Y/\nu}}$ の分散は $\frac{1}{Y/\nu} = \frac{\nu}{Y}$ になる.

実際には $\frac{Y}{\nu}$ は確率変数なのでそう単純ではないが, $T = \frac{Z}{\sqrt{Y/\nu}}$ の分布は
1以外の分散を持つ正規分布達の混合分布になる.

解答例の準備

Z は標準正規分布に従う確率変数であるとする.

Y が確率密度函数 $p(y|\nu)$ を持つとき, $z = \sqrt{y/\nu}\,t$, $dz = \sqrt{y/\nu}\,dt$ とおくと,

$\vee$
0

$$E[f(Z/\sqrt{Y/\nu})] = \int_0^\infty \left(\int_{-\infty}^\infty f\left(z/\sqrt{y/\nu}\right)\frac{e^{-z^2/2}}{\sqrt{2\pi}}p(y|\nu)\,dz\right)dy$$

$\dfrac{Z}{\sqrt{Y/\nu}} = T$ の
密度函数の形を
知りたければ
$E[f(T)] = \int f(t)p(t)\,dt$
の $p(t)$ を求めればよい.

$$= \int_0^\infty \left(\int_{-\infty}^\infty f(t)\frac{e^{-yt^2/(2\nu)}}{\sqrt{2\pi}}p(y|\nu)\sqrt{\frac{y}{\nu}}\,dt\right)dy$$

$$= \int_{-\infty}^\infty f(t)\left(\frac{1}{\sqrt{2\pi\nu}}\int_0^\infty e^{-t^2y/(2\nu)}y^{1/2}p(y|\nu)\,dy\right)dt$$

なので, 確率変数 $T = Z/\sqrt{Y/\nu}$ は確率密度函数

$$p(t) = \frac{1}{\sqrt{2\pi\nu}}\int_0^\infty e^{-t^2y/(2\nu)}y^{1/2}p(y|\nu)\,dy$$

を持つことになる.

解答例: $Y \sim \mathrm{Chisq}(\nu) = \mathrm{Gamma}(\nu/2, 2)$ の確率密度函数は

$$p(y|\nu) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)} e^{-y/2} y^{\nu/2-1} \quad (y > 0)$$

これは χ^2分布の定義

になるので, 前節の結果より, T の確率密度函数が以下のように計算される:

$$
\begin{aligned}
p(t) &= \frac{1}{\sqrt{2\pi\nu}\, 2^{\nu/2}\Gamma(\nu/2)} \int_0^\infty e^{-t^2 y/(2\nu)} y^{1/2} e^{-y/2} y^{\nu/2-1}\, dy \\
&= \frac{1}{\sqrt{\nu}\, 2^{(\nu+1)/2} \underbrace{\sqrt{\pi}}_{=\Gamma(\frac{1}{2})}\, \Gamma(\nu/2)} \int_0^\infty \exp\left(-\frac{1+t^2/\nu}{2} y\right) y^{(\nu+1)/2-1}\, dy \\
&= \frac{1}{\sqrt{\nu}\, 2^{(\nu+1)/2} \Gamma(1/2)\, \Gamma(\nu/2)} \left(\frac{1+t^2/\nu}{2}\right)^{-(\nu+1)/2} \Gamma((\nu+1)/2) \\
&= \frac{\Gamma((\nu+1)/2)}{\sqrt{\nu}\,\Gamma(1/2)\Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2} \\
&= \frac{1}{\sqrt{\nu}\, B(1/2, \nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}
\end{aligned}
$$

キャンセル

$(c>0)$

$$\int_0^\infty e^{-cy} y^{\alpha-1}\, dy = c^{-\alpha}\Gamma(\alpha)$$

ガンマ函数は通常この形で使われる.

$$\Gamma\left(\tfrac{1}{2}\right)\Gamma\left(\tfrac{\nu}{2}\right) = B\left(\tfrac{1}{2}, \tfrac{\nu}{2}\right)\Gamma\left(\tfrac{\nu+1}{2}\right)$$

3つめの等号で $\Gamma(1/2) = \sqrt{\pi}$ とガンマ函数のよくある使用法を使い, 最後の等号で $\Gamma(\alpha)\Gamma(\beta) = \Gamma(\alpha+\beta)B(\alpha, \beta)$ を使った. 最後の式は t 分布の確率密度函数である. ゆえに $T \sim \mathrm{TDist}(\nu)$.

できた!

$\square$

t分布のグラフ

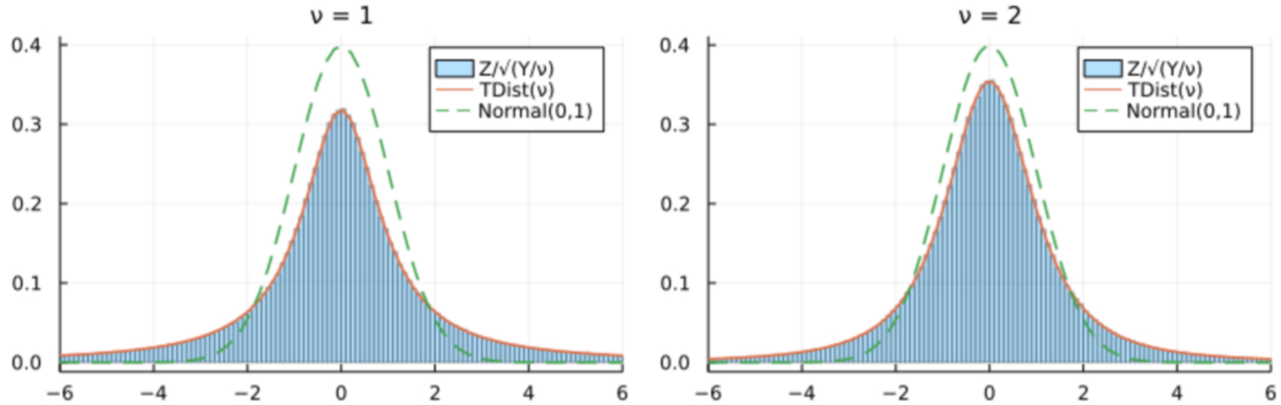
緑の破線は比較用の標準正規分布 Normal(0,1)

のヒストグラムは
Zとして Normal(0,1) の
Yとして Chisq(ν) の
乱数を大量に生成して
$T = \dfrac{Z}{\sqrt{Y/\nu}}$ を大量に計算する
ことによって作った。
(次ページで説明)

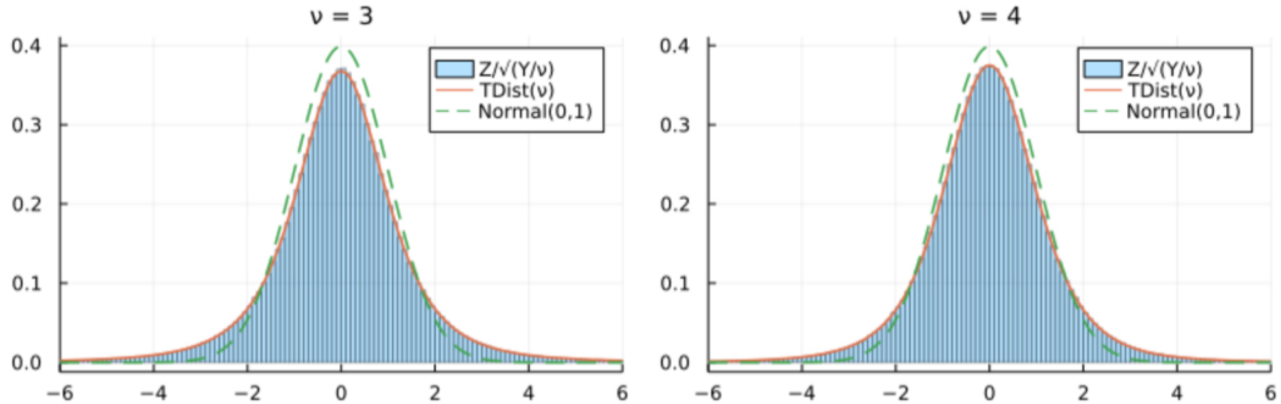
だいだい色の実線はt分布

自由度 $\nu \to \infty$ で
t分布 → 標準正規分布

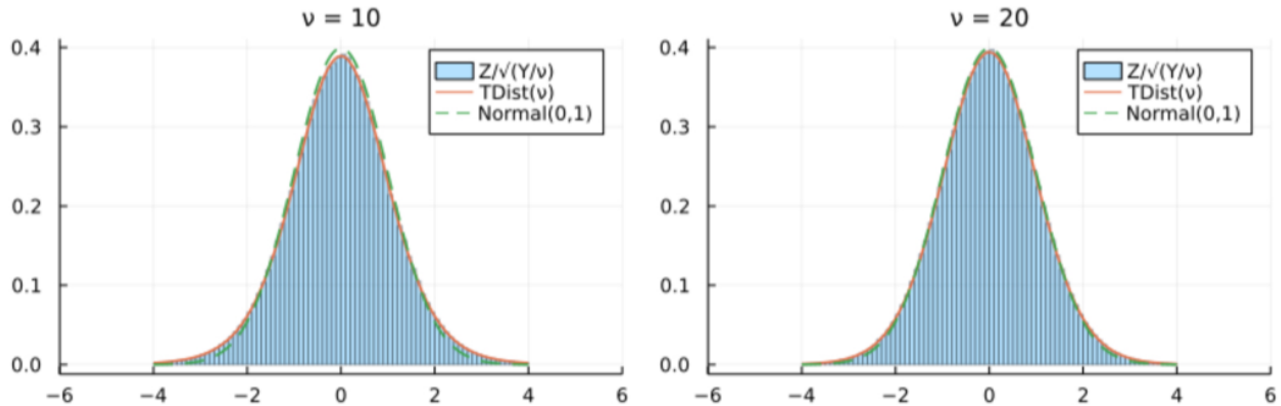
```
plot(plot_nct(1), plot_nct(2); size=(800, 250))
```




```
plot(plot_nct(3), plot_nct(4); size=(800, 250))
```




```
plot(plot_nct(10), plot_nct(20); size=(800, 250))
```



t分布の乱数の作り方

① 自由度 ν の χ^2 分布の乱数 $y>0$ を生成する.

② 標準正規分布の乱数 $-\infty<z<\infty$ を生成して
$\sqrt{y/\nu}$ で割る. これは分散 $\frac{\nu}{y}$ の正規分布の
乱数を生成していることになる.

この $t=\dfrac{z}{\sqrt{y/\nu}}$ が自由度 ν の t 分布の乱数になる.

統計学や機械学習では
シンプルな確率分布を組み合わせて,
複雑な確率分布を作って利用する.
その理解のためには,
その確率分布の乱数を
どのようにすれば作れるだろうか?
と考えることが極めて有効である!

$$\overbrace{p(t|\nu)}^{\text{自由度}\nu\text{の}t\text{分布}} = \int_0^\infty \overbrace{\frac{e^{-\frac{t^2}{2\nu/y}}}{\sqrt{2\pi\nu/y}}}^{\text{分散}\frac{\nu}{y}\text{の正規分布}} \overbrace{\frac{e^{-\frac{y}{2}}\,y^{\frac{\nu}{2}-1}}{2^{\nu/2}\,\Gamma(1/2)}}^{\text{自由度}\nu\text{の}\chi^2\text{分布}}\,dy$$

$$= \int_0^\infty p_{②}(t|y)\,p_{①}(y)\,dy \quad \text{の形}$$

① $p_{①}(y)$ の分布の乱数 y を生成,

② その y を使って,
$p_{②}(t|y)$ の分布の乱数 t を生成.

このパターンは非常によく出て来る!

重要自習問題

必ず解いて欲しい問題: t 分布における確率が95%または99%になる区間

$0 < \alpha < 1$, $\nu > 0$ であるとし, $T \sim \mathrm{TDist}(\nu)$ と仮定する. このとき, T が区間 $[-c, c]$ に含まれる確率

$$P(-c \le T \le c)$$

が $1-\alpha$ になるような c を **正則化された不完全ベータ函数**

$$y = I(x|\alpha,\beta) = I_x(\alpha,\beta) = \frac{\int_0^x v^{\alpha-1}(1-v)^{\beta-1}\,dv}{B(\alpha,\beta)}$$

の逆函数 $x = I^{-1}(y|\alpha,\beta)$ を使って表せ. (注意: $I_x(\alpha,\beta)$ は広く使われている標準的な記号法だが, $I(x|\alpha,\beta)$ と $I^{-1}(y|\alpha,\beta)$ はここだけの記号法である.)

さらに, $\nu = 9$ と $\nu = 19$ と $\nu = 29$ の場合に $P(-c \le T \le c)$ が 95% になる c と 99% になる c を小数点以下第2桁目まで求めよ.

ここだけは必ず解いておいてください,

t検定や
平均の信頼区間で使われる.

Poisson分布の定義

$\lambda > 0$ と仮定する. 確率質量函数

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

$$p(k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k!} \quad (k = 0, 1, 2, \ldots)$$

で定義される無限離散分布を **Poisson分布** (ポアソン分布)と呼び, 次のように表すことにする:

$$\mathrm{Poisson}(\lambda).$$

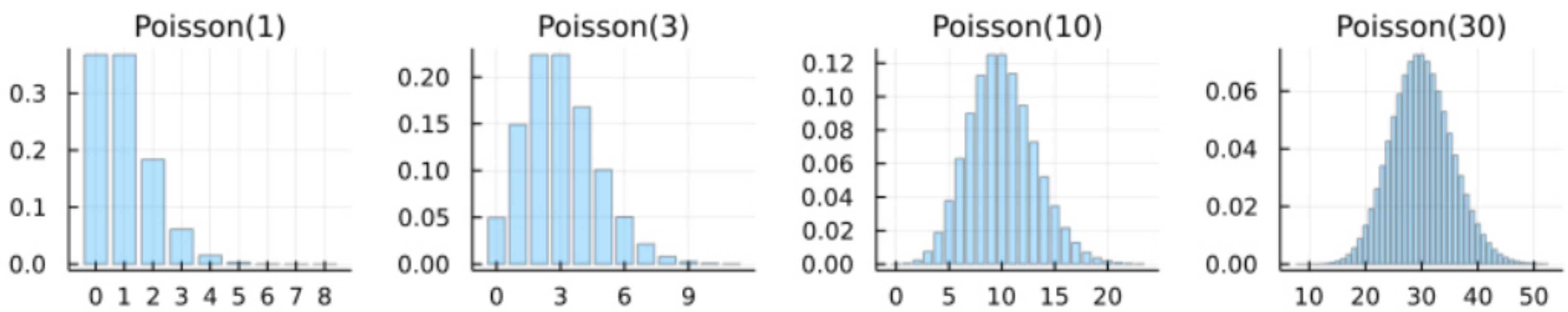
期待値 λ

分散 λ

同じ

Poisson分布は一定期間内に起こるイベントの回数の分布のモデル化としてよく使われている.

今日のテーマの1つ



以上はPoisson分布 $\mathrm{Poisson}(\lambda)$ の確率質量函数のグラフの例である. ここでは λ が整数の場合のみを扱ったが, λ は整数でなくてもよい. $\mathrm{Poisson}(30)$ のグラフは正規分布のグラフに近くなっている. パラメータ λ が大きなPoisson分布は正規分布で近似される.

再掲

$\boxed{2}$ Poisson分布とガンマ分布の以下の解釈について説明

1が出ることをイベント発生と解釈

離散時間の場合 (Bernoulli試行) $\xrightarrow{\text{連続時間極限}}$ 連続時間の場合

離散時間の場合 (Bernoulli試行)

単位時間にL回試行

単位時間のあいだに平均 λ 回 $\overbrace{\text{イベント}}^{\text{1が出る}}$ 発生

二項分布 $K \sim \mathrm{Binomial}(L, \lambda/L)$ $\xrightarrow{L \to \infty}$ Poisson分布 $K \sim \mathrm{Poisson}(\lambda)$

L回の試行中のイベント発生回数Kの分布

イベント発生の時間的間隔の期待値は $\theta = \frac{1}{\lambda}$

負の二項分布

$M \sim \mathrm{NegativeBinomial}(\alpha, \underbrace{1/(L\theta)}_{=\lambda/L})$

$N := \alpha + M$

$=$ (α回イベントが起こるまでの試行回数)

$\frac{N}{L} \xrightarrow{L \to \infty}$ (ちょうどα回イベントが起こるまでの時間) $= T$.

連続時間の場合

単位時間のあいだに平均 λ 回イベントが起こる

Poisson分布 $K \sim \mathrm{Poisson}(\lambda)$

単位時間のあいだに起こるイベントの回数Kの分布

イベント発生の時間的間隔の期待値は $\theta = \frac{1}{\lambda}$

ガンマ分布

$T \sim \mathrm{Gamma}(\alpha, \theta)$

$$\boxed{\mathrm{Binomial}(n, \lambda/n) \xrightarrow{n\to\infty} \mathrm{Poisson}(\lambda)}$$

$\left(n = 0, 1, 2, \dots, \quad \lambda > 0, \ \frac{\lambda}{n} \leqq 1\right)$

前ページのLをnと書いた。

$$\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \xrightarrow[n\to\infty]{} e^{-\lambda}\frac{\lambda^k}{k!}$$

$(k = 0, 1, \dots, n)$ $\qquad$ $(k = 0, 1, 2, 3, \dots)$

このように、Poisson分布は二項分布の極限になっている。

$\boxed{\text{証明}}$ $n\to\infty$ のとき、

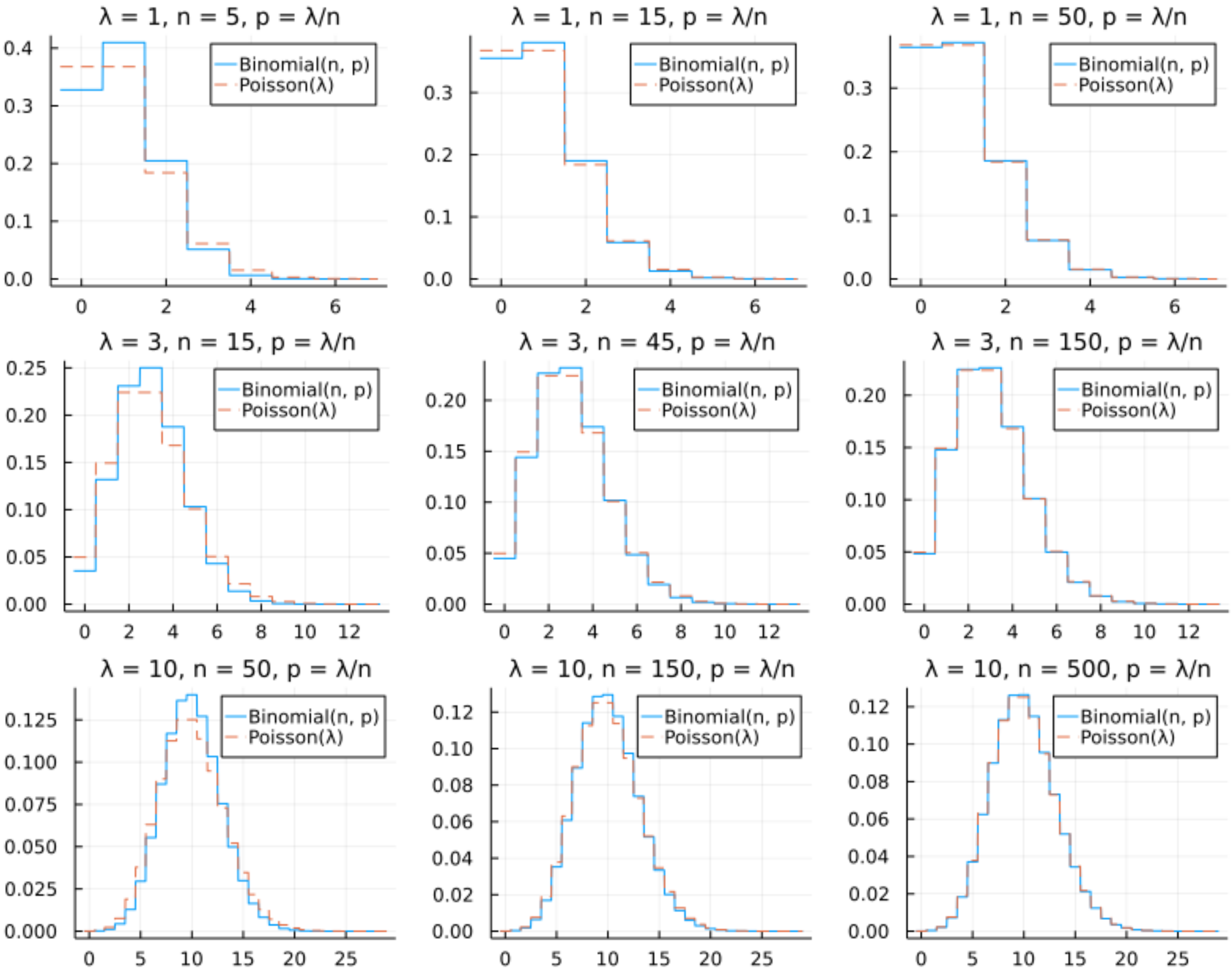
$$\binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!}\frac{\lambda^k}{n^k}\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \underbrace{\frac{n(n-1)\cdots(n-k+1)}{n\cdot n\cdot\cdots\cdot n}}_{\to 1}\underbrace{\left(1-\frac{k}{n}\right)^{-k}}_{\to 1}\underbrace{\left(1-\frac{\lambda}{n}\right)^{n}}_{\to e^{-\lambda}}\frac{\lambda^k}{k!} \xrightarrow[n\to\infty]{} e^{-\lambda}\frac{\lambda^k}{k!} \qquad \boxed{\text{q.e.d.}}$$

$$\left\{\begin{aligned} &\frac{n(n-1)\cdots(n-k+1)}{n\cdot n\cdot\cdots\cdot n} \\ &= \frac{n}{n}\,\frac{n-1}{n}\cdots\frac{n-k+1}{n} \xrightarrow{n\to\infty} 1 \end{aligned}\right.$$

$$\left\{\begin{aligned} \log\left(1-\frac{\lambda}{n}\right)^{n} &= L\log\left(1-\frac{\lambda}{n}\right) \\ &= n\left(-\frac{\lambda}{n}+\frac{\lambda^2}{2n^2}+\cdots\right) \\ &= -\lambda+\frac{\lambda^2}{2n}+\cdots \longrightarrow -\lambda \\ \therefore \left(1-\frac{\lambda}{n}\right)^{n} &\to e^{-\lambda} \end{aligned}\right.$$

n: 小 → 中 → 大


λ = 1, n = 5, p = λ/n
λ = 1, n = 15, p = λ/n
λ = 1, n = 50, p = λ/n
λ = 3, n = 15, p = λ/n
λ = 3, n = 45, p = λ/n
λ = 3, n = 150, p = λ/n
λ = 10, n = 50, p = λ/n
λ = 10, n = 150, p = λ/n
λ = 10, n = 500, p = λ/n
Binomial(n, p)
Poisson(λ)

以下, Bernoulli試行で1が出ることを「イベントが起こった」という.

離散時間		連続時間
$K \sim \mathrm{Binomial}(n, \lambda/n)$	$\longrightarrow$	$K \sim \mathrm{Poisson}(\lambda)$
単位時間のあいだに n 回の試行を行う.	$\xrightarrow[\text{連続時間極限}]{n \to \infty}$	連続単位時間のあいだにイベントが何回起こるかを見る.
確率 $p = \dfrac{\lambda}{n}$ でイベントが起こる試行を n 回行ったときに起こったイベントの回数 K の分布	$\longrightarrow$	単位時間のあいだに起こったイベントの回数 K の分布
$p = \dfrac{\lambda}{n}$ なので n 回の試行で起こるイベントの回数の期待値は $np = \lambda$ になる.	$\longrightarrow$	単位時間のあいだに起こるイベントの回数の期待値は λ になる.
分散 $np(1-p) = \lambda\left(1 - \dfrac{\lambda}{n}\right)$	$\xrightarrow{n \to \infty}$	分散 λ

離散時間

$M \sim \mathrm{NegativeBinomial}\left(\alpha, \frac{1}{L\theta}\right) \xrightarrow[L \to \infty]{T = \frac{N}{L} = \frac{\alpha + M}{L}}$

連続時間

$T \sim \mathrm{Gamma}(\alpha, \theta)$

確率 $p = \frac{1}{L\theta}$ でイベントが起こる試行を
単位時間あたり L 回ずつ，
イベントが α 回起こるまで続ける．
ちょうど α 回イベントが起こるまでの
試行回数 $N = \alpha + M$ の分布

連続時間極限

平均して時間間隔 θ で
起こるイベントが
α 回起こるまでにかかる
時間 T の分布

$\theta = \frac{1}{Lp} = \begin{pmatrix}\text{単位時間のあいだに} \\ \text{起こるイベントの回数} \\ \text{の期待値}\end{pmatrix}^{-1}$ なので

θ はイベントが起こる時間の間隔の
期待値になる．

$\begin{pmatrix}\text{ちょうど } \alpha \text{ 回イベントが} \\ \text{起こるまでにかかった時間}\end{pmatrix} = \frac{N}{L} = T$

このようにガンマ分布は
負の二項分布の連続時間極限
とみなされる．

証明は次ページより．

$\boxed{準備}$ $\alpha, \theta > 0$, $p = \dfrac{1}{L\theta}$, $n = Lt$ とおく、 $\Gamma(n) = (n-1)!$ を使うと,

$$\binom{n-1}{n-\alpha} = \frac{\Gamma(n)}{\Gamma(n-\alpha+1)\Gamma(\alpha)} = \frac{n\Gamma(n)}{n\Gamma(\alpha)\Gamma(n-\alpha+1)} = \frac{\Gamma(n+1)}{n\Gamma(\alpha)\Gamma(n-\alpha+1)} = \frac{1}{nB(\alpha, n-\alpha+1)}.$$

$\Gamma(\alpha)\Gamma(\beta) = B(\alpha,\beta) \times \Gamma(\alpha+\beta)$

(1) NegativeBinomial(α, p)におけるイベントがちょうどα回起こったときの試行回数$N = \alpha + M$の確率質量函数：

$$P(n-\alpha|\alpha,p) = \binom{n-1}{n-\alpha} p^\alpha (1-p)^{n-\alpha} = \frac{1}{nB(\alpha, n-\alpha+1)} p^\alpha (1-p)^{n-\alpha}.$$

(2) Gamma(α, θ)の確率密度函数：

$$p(t|\alpha,\theta) = \frac{1}{\theta^\alpha \Gamma(\alpha)} e^{-t/\theta} t^{\alpha-1} \quad (t>0).$$

(3) ベータ函数 $\longrightarrow$ ガンマ函数： $l \to \infty$ のとき,

ガンマ分布がベータ分布の極限になっていることも意味している.

$$l^\alpha B(\alpha, l+b) = l^\alpha \int_0^1 x^{\alpha-1}(1-x)^{l+b-1}\,dx$$

$x = \dfrac{t}{l}$

$$= \int_0^l t^{\alpha-1}\left(1-\frac{t}{l}\right)^{l+b-1} dt \longrightarrow \int_0^\infty t^{\alpha-1} e^{-t}\,dt = \Gamma(\alpha).$$

$$\boxed{N-\alpha \sim \mathrm{NegativeBinomial}\left(\alpha, \frac{1}{L\theta}\right) \text{ のとき, } \frac{N}{L} \xrightarrow[L\to\infty]{} T \sim \mathrm{Gamma}(\alpha, \theta)}$$

$\boxed{\text{証明}}$ $p = \frac{1}{L\theta}$, $n = Lt$ とおくと, $L\to\infty$ において,

$$\frac{1}{nB(\alpha, n-\alpha+1)} p^{\alpha}(1-p)^{n-\alpha} \leftarrow (1)$$

$$= \frac{1}{Lt\cdot B(\alpha, Lt-\alpha+1)} \frac{1}{L^{\alpha}\theta^{\alpha}} \overbrace{\left(1-\frac{1}{L\theta}\right)^{Lt-\alpha}}^{\to e^{-t/\theta}}$$

$$= \frac{1}{\underbrace{(Lt)^{\alpha} B(\alpha, Lt-\alpha+1)}_{(3)\ \to\ \Gamma(\alpha)}} \frac{L^{\alpha-1} t^{\alpha-1}}{L^{\alpha}\ \theta^{\alpha}} \left(1-\frac{1}{L\theta}\right)^{Lt-\alpha}$$

$$= \frac{1}{\theta^{\alpha}\underbrace{(Lt)^{\alpha} B(\alpha, Lt-\alpha+1)}} \left(1-\frac{1}{L\theta}\right)^{Lt-\alpha} t^{\alpha-1} \frac{1}{L}$$

$$= \frac{1}{\theta^{\alpha}\underbrace{\Gamma(\alpha)}} e^{-t/\theta}\, t^{\alpha-1} \frac{1}{L} \times (1+o(1)) \leftarrow (3)$$

$o(1)$: $L\to\infty$ で 0 に収束する量

$$\left(1-\frac{1}{L\theta}\right)^{Lt} = \left(1-\frac{t/\theta}{Lt}\right)^{Lt} \to e^{-t/\theta}$$

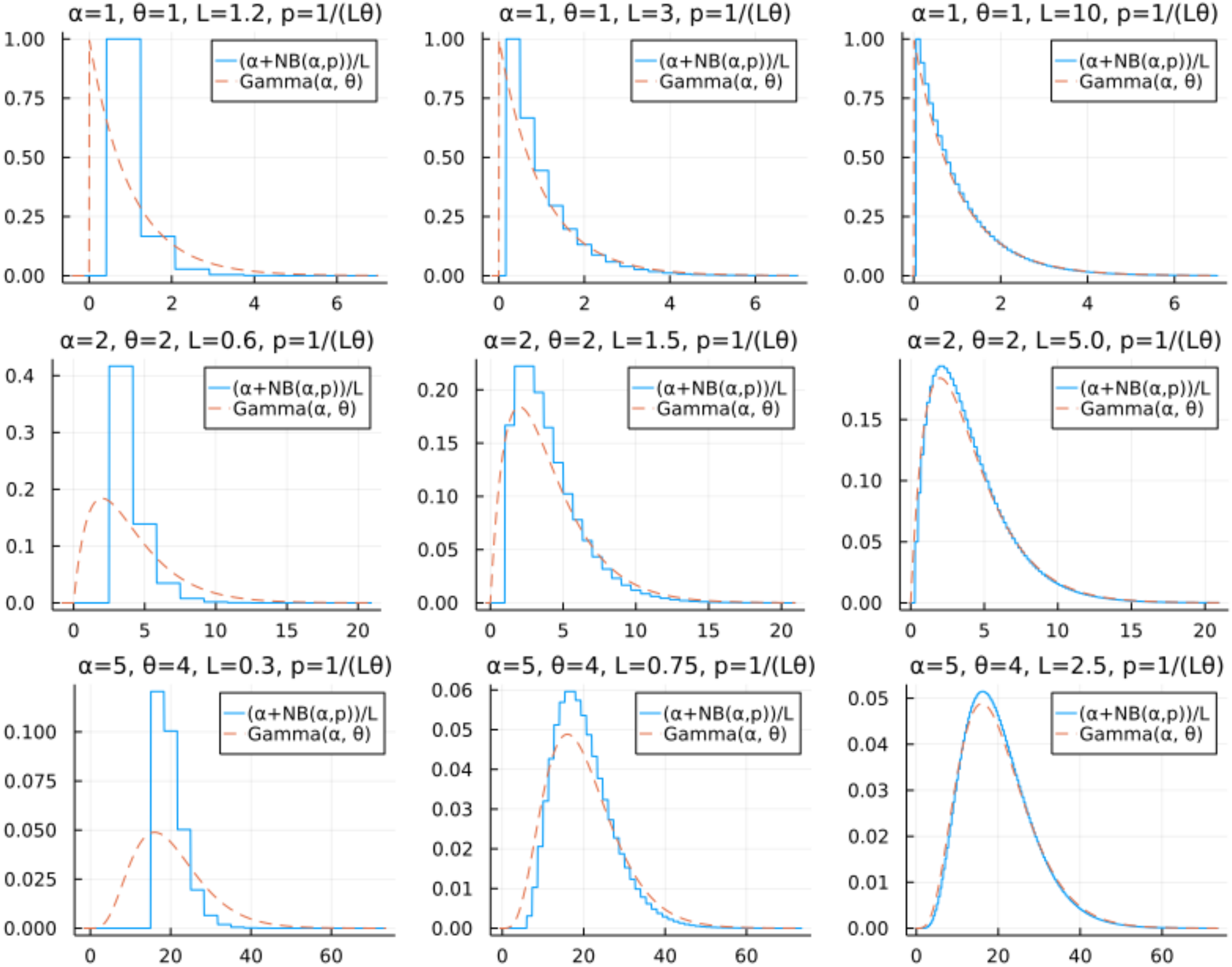
$$\uparrow$$

$$\left(1+\frac{x}{l}\right)^{l} \xrightarrow[l\to\infty]{} e^{x}$$

($n = Lt$ より "$dn = L\,dt$" ← これとキャンセル)

$\boxed{\text{q.e.d.}}$

L: 小 → 中 → 大


α=1, θ=1, L=1.2, p=1/(Lθ)
α=1, θ=1, L=3, p=1/(Lθ)
α=1, θ=1, L=10, p=1/(Lθ)
α=2, θ=2, L=0.6, p=1/(Lθ)
α=2, θ=2, L=1.5, p=1/(Lθ)
α=2, θ=2, L=5.0, p=1/(Lθ)
α=5, θ=4, L=0.3, p=1/(Lθ)
α=5, θ=4, L=0.75, p=1/(Lθ)
α=5, θ=4, L=2.5, p=1/(Lθ)
(α+NB(α,p))/L
Gamma(α, θ)

応用 Poisson分布における確率とガンマ分布における確率の関係

単位時間に平均して λ 回のイベントが起こると仮定する。

時間 τ のあいだに起こるイベントの回数: $\boxed{K_{\lambda\tau} \sim \mathrm{Poisson}(\lambda\tau)}$

平均して時間間隔 $\theta = \dfrac{1}{\lambda}$ でイベントが発生する。

そのとき、イベントが α 回起こるまでにかかる時間 $\boxed{T_{\alpha,\theta} \sim \mathrm{Gamma}(\alpha, \theta)}$.

(α は正の整数であると仮定する。)

$K_{\lambda\tau} < \alpha \iff$ 時間 τ のあいだに起こるイベントの回数が α 未満

$\iff$ イベントが α 回起こるまでにかかる時間が τ より大きい

$\iff T_{\alpha,\theta} > \tau$

自習問題 直接的に示せ。(ヒント: 両辺を τ で微分)

ゆえに、 $\boxed{P(K_{\lambda\tau} < \alpha) = P(T_{\alpha,\theta} > \tau)}$.

具体的な式で書くと、

$$\sum_{k=0}^{\alpha-1} \frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!} = \frac{\int_\tau^\infty e^{-t/\theta} t^{\alpha-1}\,dt}{\theta^\alpha \Gamma(\alpha)},$$

注 右辺の方が計算が"楽"

$P(\alpha, x) = \dfrac{\int_0^x \sim}{▨}$

注意 $Q(\alpha, x) = \dfrac{\int_x^\infty e^{-t} t^{\alpha-1}\,dt}{\Gamma(\alpha)}$ は正則化された(不完全)ガンマ函数と呼ばれ、コンピュータで効率的に計算できる。

重要自習問題

必ず解いて欲しい問題: χ²分布における確率が95%または99%になる範囲

$0 < \alpha < 1$, $\nu > 0$ であるとし, $Y \sim \mathrm{Chisq}(\nu)$ と仮定する. このとき, Y が c 以下になる確率

$$P(Y \le c)$$

が $1-\alpha$ になるような c を **正則化された(不完全)ガンマ函数** (regularized (incomplete) gamma function)

$$y = P(x|\alpha) = P(\alpha, x) = \frac{\int_0^x e^{-u} u^{\alpha-1}\,du}{\Gamma(\alpha)} \quad (x > 0)$$

の逆函数 $x = P^{-1}(y|\alpha)$ を使って表せ. (注意: $P(\alpha, x)$ は広く使われている標準的な記号法だが, $P(x|\alpha)$ と $P^{-1}(y|\alpha)$ はここだけの記号法である.)

さらに, $\nu = 1$ と $\nu = 2$ と $\nu = 3$ の場合に $P(Y \le c)$ が 95% になる c と 99% になる c を小数点以下第2桁目まで求めよ.

ここは必ず解いておくこと! (たくさんの種類がある χ^2 検定で使われる.)

解答例はGithubにある.

3 ベータ分布の一様乱数発生のくりかえしによる解釈

0〜1のあいだの
一様乱数生成を
n回くりかえした結果

T_iがp以下なら1
そうでないなら0に変換

Bernoulli試行
確率pで1を
確率$1-p$で0を生成

$$T_1 = \mathrm{rand}() \longmapsto X_1$$
$$T_2 = \mathrm{rand}() \longmapsto X_2$$
$$\cdots\cdots \qquad \vdots$$
$$T_n = \mathrm{rand}() \longmapsto X_n$$

重要

この中で下からk番目に
小さなものを$T_{(n,k)}$と書く。

$$\underbrace{T_{(n,k)} \sim \mathrm{Beta}(k, n-k+1)}$$

ポイント! 後で示す

p以下のT_iがk個以上

$$\Updownarrow \qquad \Updownarrow$$

$$T_{(n,k)} \leqq p \iff K_{n,p} \geqq k$$

この中に含まれる1の
個数を$K_{n,p}$と書く。

$$K_{n,p} \sim \mathrm{Binomial}(n, p)$$

$$\therefore\ P(T_{(n,k)} \leqq p) = P(K_{n,p} \geqq k).$$

さらに，一様乱数生成をくりかえして，1がちょうどk個出るまでの試行回数を$N_{k,p}$と書く。

$$N_{k,p} - k \sim \mathrm{NegativeBinomial}(k, p)$$

n回の試行で1となるX_iがk個以上 $\iff N_{k,p} \leqq n$

$$\Updownarrow \qquad \Updownarrow$$

n回の試行でp以下のT_iがk個以上

$$P(K_{n,p} \geqq k) = P(N_{k,p} \leqq n)$$
$$\| \qquad //$$
$$P(T_{(n,k)} \leqq p)$$

前ページの結論で $T_{(n,k)} \sim Beta(k, n-k+1)$ のみがまだ示されていない。

$\boxed{T_{(n,k)} \sim Beta(k, n-k+1) \text{ の証明}}$

{別証が Github にある資料にある
{別証の方が論理的にクリア。しかし計算はめんどう。

$T_{(n,k)} = (0\sim1$ の一様乱数達 $T_1, T_2, \ldots, T_n$ の中で k 番目に小さいもの$)$ であった。

$T_{(n,k)} \in [t, t+dt]$ となる確率は、
dt について2次以上の項を無視すると次のように近似される：

$$P\left(T_{(n,k)} \in [t, t+dt]\right) \approx \underbrace{\frac{n!}{(k-1)!\,1!\,(n-k)!}} \underbrace{t^{k-1}\,dt\,(1-t)^{n-k}}_{\simeq\, P(T_1,\ldots,T_{k-1} < t < T_k < t+dt < T_{k+1},\ldots,T_n)}$$

n 個の T_i を
$\underbrace{k-1 \text{ 個}}_{T_{(n,k)}\text{ より小}}$ と $\underbrace{1 \text{ 個}}_{= T_{(n,k)}}$ と $\underbrace{n-k \text{ 個}}_{T_{(n,k)}\text{ より大}}$
に分割する方法の数

そして、$\dfrac{n!}{(k-1)!\,(n-k)!} = \dfrac{\Gamma(n+1)}{\Gamma(k)\,\Gamma(n-k+1)} = \dfrac{1}{B(k, n-k+1)}$.

ゆえに、$T_{(n,k)}$ の確率密度函数は $\dfrac{t^{k-1}(1-t)^{n-k}}{B(k, n-k+1)} = \left(\begin{array}{l}Beta(k, n-k+1) \text{ の}\\ \text{確率密度函数}\end{array}\right)$ $\boxed{\text{q.e.d.}}$

応用　二項分布，負の二項分布，ベータ分布における確率の関係

すでに以下が示されている．

$T_{(n,k)} \sim \mathrm{Beta}(k, n-k+1)$，　$K_{n,p} \sim \mathrm{Binomial}(n,p)$，　$N_{k,p} - k \sim \mathrm{NegativeBinomial}(k,p)$ のとき，

$$P(T_{(n,k)} \leqq p) = P(K_{n,p} \geqq k) = P(N_{k,p} \leqq n).$$

具体的な式で書くと，

nが大きいとき少し重い計算になる．

$$\underbrace{\frac{\int_0^p t^{k-1}(1-t)^{n-k}\,dt}{B(k, n-k+1)}}_{A} = \underbrace{\sum_{i \geqq k} \binom{n}{i} p^i (1-p)^{n-i}}_{B} = \underbrace{\sum_{j \leqq n} \binom{j-1}{j-k} p^k (1-p)^{j-k}}_{C}.$$

自習問題　A＝B を直接的に示せ．（できれば A＝C も）

注意　$I_p(\alpha, \beta) = \dfrac{\int_0^p t^{\alpha-1}(1-t)^{\beta-1}\,dt}{B(\alpha,\beta)}$ は正則化された不完全ベータ函数と呼ばれ，コンピュータで効率的に計算できる．　□

二項分布

Binomial(n,p)

と

ベータ分布

Beta(k,n-k+1)

と

負の二項分布

Negative

Binomial(k,p)

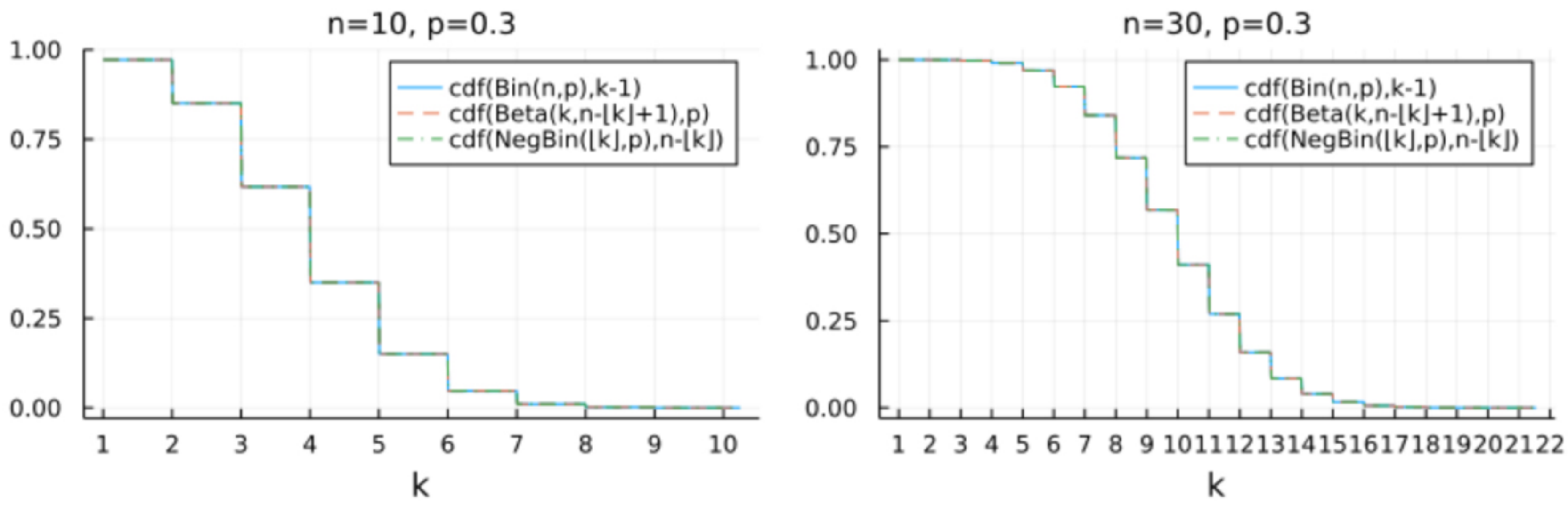
における

確率が

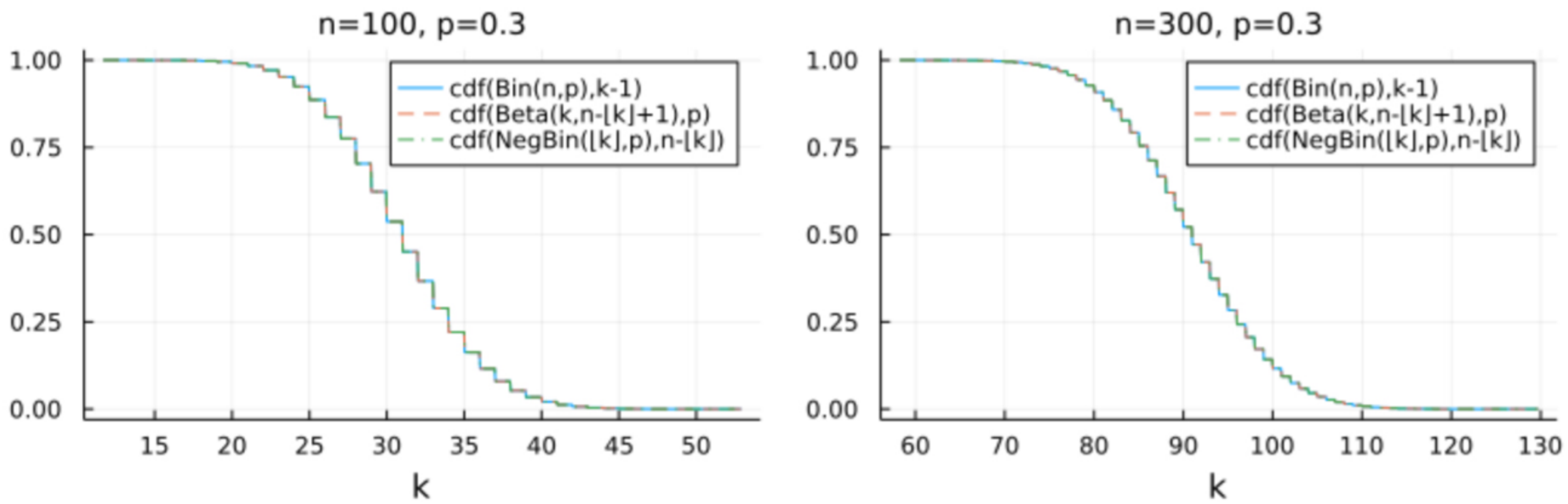
ぴったり

一致している。

```
plot(plot_cdf_bin_beta_nb(10, 0.3), plot_cdf_bin_beta_nb(30, 0.3);
    size=(800, 250), bottommargin=4Plots.mm)
```




```
plot(plot_cdf_bin_beta_nb(100, 0.3; xtick=0:5:100), plot_cdf_bin_beta_nb(300, 0.3; xtick=0:10:1000);
    size=(800, 250), bottommargin=4Plots.mm)
```



二項分布とベータ分布の関係のClopper-PearsonのP値函数への応用

重要な応用

二項分布モデルに基く比率 p (成功確率パラメータと呼んで来た)の **Clopper-Pearsonの信頼区間** を与えるP値函数(以下 **Clopper-PearsonのP値函数** と呼ぶ)の効率的な計算に，前節で述べた二項分布とベータ分布の関係が実際に使われている.

二項分布とベータ分布の関係はあとでパラメータ p のClopper-Pearsonの信頼区間の計算で使われる.

Clopper-PearsonのP値函数は二項分布モデル $\mathrm{Binomial}(n,p)$ と「n 回中 k 回成功した」というデータの整合性を測る函数である. Clopper-PearsonのP値函数は以下のように定義される.

二項分布 $\mathrm{Binomial}(n,p)$ の累積分布函数を，確率変数 $K \sim \mathrm{Binomial}(n,p)$ を使って

P値の計算にこれが必要

二項分布のcdf

$$F(k) = P(K \le k) = \sum_{j=0}^{k} \binom{n}{j} p^j (1-p)^{n-j}$$

と定義する. Clopper-PearsonのP値函数 $\mathrm{pvalue}_{\mathrm{CP}}(k|n,p)$ を次のように定める:

$$\mathrm{pvalue}_{\mathrm{CP}}(k|n,p) = \min(1, 2F(k), 2(1-F(k-1)) = \min(1, 2P(K \le k), 2P(K \ge k)).$$

$\min(P(K \le k), P(K \ge k))$ は二項分布におけるデータの値 k 以上に端側に偏る確率を意味し，その2倍がClopper-PearsonのP値である. P値が小さいほど，二項分布モデル $\mathrm{Binomial}(n,p)$ と「n 回中 k 回成功した」というデータの整合性が低いと考える.

Clopper-Pearsonの信頼区間は統計学入門における非常に基本的な題材である!

二項分布の累積分布函数はベータ分布の累積分布函数で書けるのであった. 確率変数 $T_k \sim \mathrm{Beta}(k, n-k+1)$ を使えば次が成立している:

$$F(k) = P(K \le k) = 1 - P(T_{k+1} \le p), \quad 1 - F(k-1) = P(K \ge k) = P(T_k \le p).$$

そして，右辺の計算で必要な $P(T_k \le p)$ は次のように書ける:

これらを使うとP値の計算でこれを使える.

ベータ分布のcdf

$$P(T_k \le p) = \frac{\int_0^p t^{k-1}(1-t)^{n-k}\,dt}{B(k, n-k+1}.$$

基本的な応用

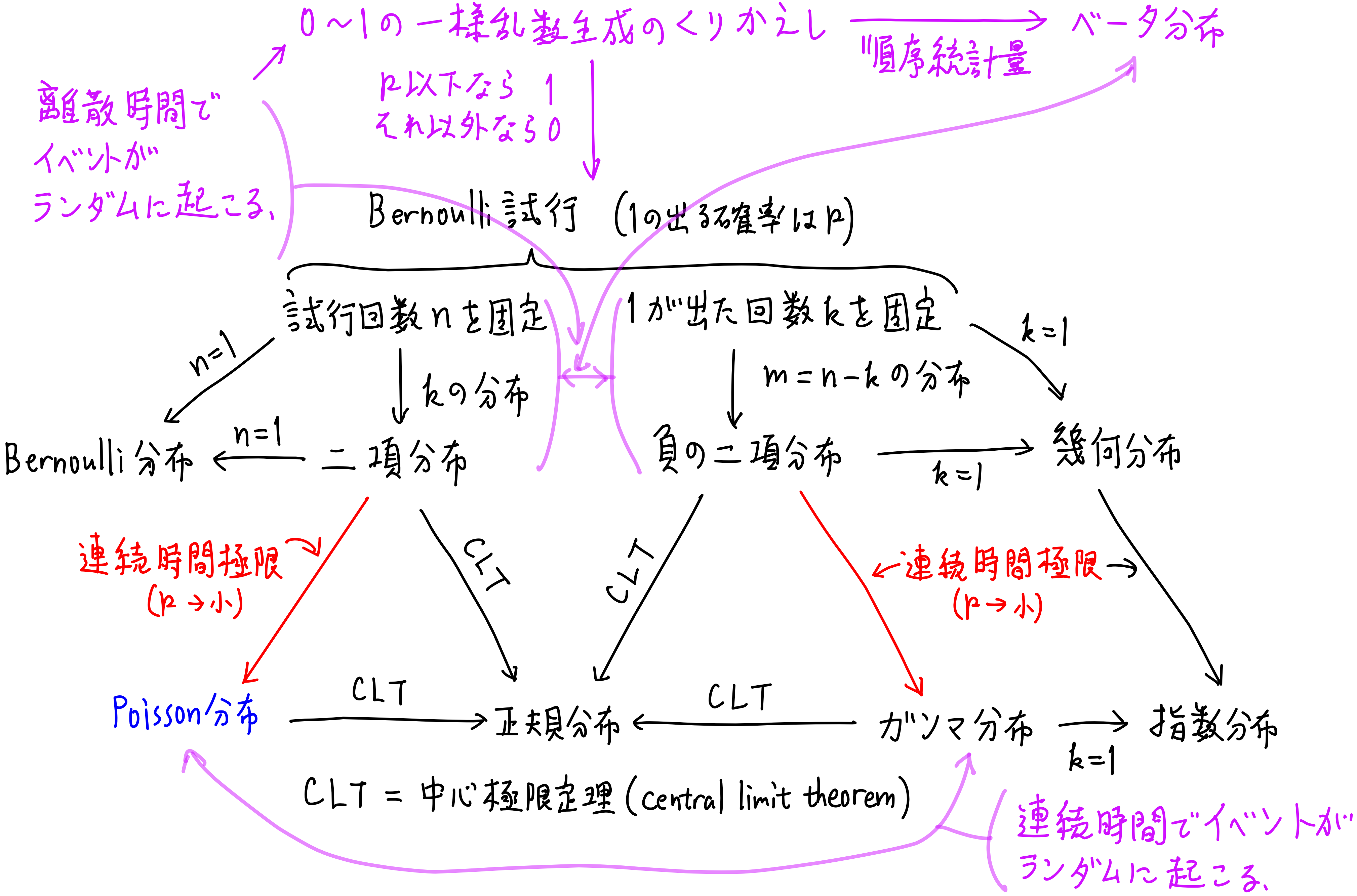
これの右辺は，基本特殊函数の1つである正則化された不完全ベータ函数になっており，コンピュータで効率的に計算できる.

注意: $0 < \alpha < 1$ のとき，「n 回中 k 回成功した」というデータが与える信頼度 $1-\alpha$ のClopper-Pearsonの信頼区間は $\mathrm{pvalue}_{\mathrm{CP}}(k|n,p) \ge \alpha$ を満たすパラメータ p の範囲と定義される:

$$\mathrm{CI}_{\mathrm{CP}}(k|n) = \{\, p \in [0,1] \mid \mathrm{pvalue}_{\mathrm{CP}}(k|n,p) \ge \alpha \,\}.$$

この辺りの話は後で詳しく説明することになる.

Bernoulli試行の関連確率分布の図



将来より進んだことの勉強に役に立つ問題

問題: Poisson分布の中心極限定理の直接証明

まだ説明していないこと: Poisson分布たちの積の条件付き確率分布でも二項分布や多項分布を作れる.

パラメータ $\lambda > 0$ を持つPoisson分布の確率質量函数の定義は

$$p(k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k!} \quad (k = 0, 1, 2, \dots)$$

であった. k は λ ごとに異なる値を取るものとし, 固定された x について

$$k = \lambda + \sqrt{\lambda}\,x + o(\sqrt{\lambda})$$

"$dk \approx \sqrt{\lambda}\,dx$"

を満たしていると仮定する. ここで $o(\sqrt{\lambda})$ は $\sqrt{\lambda}$ で割ると $\lambda \to \infty$ で 0 に収束する量を表す. このとき次が成立することを示せ:

$$p(k|\lambda) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}\frac{1}{\sqrt{\lambda}}(1 + o(1)).$$

$k \approx \lambda + \sqrt{\lambda}\,x$

$\Downarrow$

$$p(k|\lambda)\,dk \approx \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx$$

ここで $o(1)$ は $\lambda \to \infty$ で 0 に収束する量である.

Poisson分布の中心極限定理への理解

$\downarrow$

二項分布や多項分布の中心極限定理への理解

$\downarrow$

Kullback-Leibler情報量への理解

注意: 以下の解答例は二項分布の中心極限定理の証明のひな型になる.

ヒント

解答例: まず, 問題文よりも弱い次の条件を仮定する:

$$k = \lambda + o(\lambda) = \lambda(1 + o(1))$$

ここで $o(\lambda)$ は λ で割ると $\lambda \to \infty$ で 0 に収束する量である. このとき, $k!$ に Stirlingの公式を適用すると,

$$p(k|\lambda) = e^{-\lambda}\frac{\lambda^k}{k^k e^{-k}\sqrt{2\pi k}}(1 + o(1)) = \frac{1}{\sqrt{2\pi\lambda}}\left(\frac{k}{\lambda}\right)^{-k} e^{k-\lambda}(1 + o(1)).$$

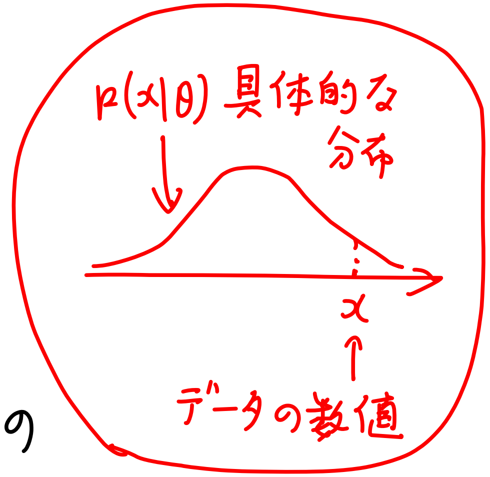
そして,

$$-\log\left(\left(\frac{k}{\lambda}\right)^{-k} e^{k-\lambda}\right) = k\log\frac{k}{\lambda} - (k - \lambda)$$

KL情報量の各項を与える.

詳しくはGithubにおいてある資料を見よ.

標本分布 ← データの生成のされ方の法則のモデル化



現実での仕事

複雑よくわからない現実

↓ 観測, 測定, 調査

データ = 数値の列や表
$x_1, x_2, \ldots, x_n$

↓

統計モデルと
データの数値の比較

↓

推定・推測・推論

↓

意思決定と行動

関連の数学的モデル

未知の現実の法則のモデル化としての
任意の確率分布 $q(x)$, $Q(x)$

iid の場合には $q(x_1)\cdots q(x_n)$ や $Q(x_1)\cdots Q(x_n)$

統計分析で使う具体的な確率分布
$p(x|\theta)$, $P(x|\theta)$

例 正規分布, 二項分布, etc.

統計分析で使う確率分布の標本分布
$p(x_1|\theta)\cdots p(x_n|\theta)$ や $P(x_1|\theta)\cdots P(x_n|\theta)$

と現実から得たデータの数値列を比較する。

標本分布

標本分布の定義の準備　$X_1, X_2, \ldots, X_n$ は確率変数であるとする。

同時確率分布

X_i 単独の確率分布だけではなく，$(X_1, \ldots, X_n)$ の全体の確率分布を考える。

$(X_1, \ldots, X_n)$ の函数 $f(X_1, \ldots, X_n)$ の期待値 $E[f(X_1, \ldots, X_n)]$ が定義されていれば $(X_1, \ldots, X_n)$ 全体の確率分布が得られたことになる。

離散　$0 \leqq P(x_1, \ldots, x_n) \leqq 1$, $\displaystyle\sum_{x_1, \ldots, x_n} P(x_1, \ldots, x_n) = 1$.

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

同時確率質量函数

連続　$0 \leqq p(x_1, \ldots, x_n) < \infty$, $\displaystyle\int \cdots \int p(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1$,

$$E[f(X_1, \ldots, X_n)] = \int \cdots \int f(x_1, \ldots, x_n) \underbrace{p(x_1, \ldots, x_n)}\, dx_1 \cdots dx_n,$$

同時確率密度函数

以上の設定のもとで，各 X_i 単独の確率質量（密度）関数は

$$P_i(x_i) = \sum_{x_i \text{以外の} x_1, \dots, x_n} P(x_1, \dots, x_n)$$

$$\left(p_i(x_i) = \int \cdots \int p(x_1, \dots, x_n)\, dx_1 \cdots \overset{\text{トル}}{\widehat{dx_i}} \cdots dx_n \right)$$ になる．

$\boxed{\text{証明}}$

$$E[f(X_i)] = \sum_{x_1, \dots, x_n} f(x_i)\, P(x_1, \dots, x_n) = \sum_{x_i} f(x_i) \overbrace{\left(\sum_{x_i \text{以外の} x_1, \dots, x_n} P(x_1, \dots, x_n) \right)}^{\substack{P_i(x_i) \\ \parallel}}$$

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

$\boxed{\text{q.e.d.}}$

独立性

確率変数達 $X_1, \dots, X_n$ が独立 (independent) であるとは、以下の互いに同値な条件が成立していることだと定める：

(1) $E[f_1(X_1)\cdots f_n(X_n)] = E[f_1(X_1)]\cdots E[f_n(X_n)]$. ← 計算ではこの形で使われる

(2) 離散の場合には $P(x_1, \dots, x_n) = P_1(x_1)\cdots P(x_n)$ で
連続の場合には $p(x_1, \dots, x_n) = \underbrace{p_1(x_1)\cdots p_n(x_n)}$.

$\boxed{\text{(2)の後者} \Rightarrow \text{(1)の証明}}$

$$
\begin{aligned}
E[f_1(X_1)\cdots f_n(X_n)] &= \int\cdots\int f_1(x_1)\cdots f_n(x_n)\,\overbrace{p_1(x_1)\cdots p_n(x_n)}\,dx_1\cdots dx_n \\
&= \int f_1(x_1)p_1(x_1)\,dx_1 \cdots \int f_n(x_n)p_n(x_n)\,dx_n \\
&= E[f_1(X_1)] \cdots E[f_n(X_n)].
\end{aligned}
$$

$\square$

$\boxed{\text{iid} = \text{i.i.d.} = \text{independent and identically distributed} = \text{独立同分布}}$

$X_1, \dots, X_n$ が独立同分布 (iid)

$\overset{\text{def.}}{\Longleftrightarrow}$ $X_1, \dots, X_n$ は独立でかつ X_i の各々は共通の同じ分布にしたがう.

$X_1, \dots, X_n$ が各 X_i が分布 D にしたがう独立同分布な確率変数達であるとき $(X_1, \dots, X_n)$ がしたがう同時確率分布を D^n と書き,

分布 D のサイズ n の標本分布 (distribution of size-n samples)

とよび, 以下のように書く:

- $X_1, \dots, X_n \sim D$ (independent),
- $(X_1, \dots, X_n) \sim D^n$,
- $(X_1, \dots, X_n)$ is a size-n sample of D.

例 試行回数 n の Bernoulli 試行の分布 $\mathrm{Bernoulli}(p)^n$ は

Bernoulli 分布 $\mathrm{Bernoulli}(p)$ のサイズ n の標本分布である:

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

ここで, $(x_1, \dots, x_n)$ は 1 と 0 からなる長さ n の数列.

$(X_1, \dots, X_n) \sim \mathrm{Bernoulli}(p)^n$ のとき, $K = X_1 + \cdots + X_n = (1\text{の個数})$ とおくと,

$P(K = k) = \binom{n}{k} p^k (1-p)^{n-k}$ なので, $K \sim \mathrm{Binomial}(n, p)$ (二項分布).

$n = 5$, $p = \frac{1}{3}$ のとき,

$$P((X_1, \dots, X_5) = (1,0,0,1,1)) = p^3(1-p)^2 = \frac{1^3\, 2^2}{3^5} = \frac{4}{243}$$

$$P(K = 3) = \binom{5}{3} p^3 (1-p)^2 = 10\, p^3 (1-p)^2 = \frac{40}{243} \approx 16.46\%$$

このモデルは以下のモデル化に使われる:

- 薬を使った n 人の患者が○○日以内に死んだか死なずにすんだか. (何人死んだか)
- 広告を見た n 人が商品を買ったか買わないか. (何人買ったか)

$\boxed{\text{例}}$ 平均 μ, 分散 σ^2 を持つ正規分布 Normal(μ, σ) のサイズ n の標本分布の(同時)確率密度函数の形は次のようになる:

$$p(x_1, \ldots, x_n \mid \mu, \sigma) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2} \right) \right) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right).$$

μ と σ は i によらない

このモデルは, たとえば,

- 無作為抽出された千葉の n 人の中3男子の身長の数値 $(x_1, \ldots, x_n)$ の分布のモデル化に使える。(身長の分布は正規分布にかなり近いことが知られている。)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \ (\text{標本平均}), \quad s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \ (\text{不偏分散})$$ とおくと,

$n-1$ である理由はあとで説明する。

$$\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} \left((x_i - \bar{x}) + (\bar{x} - \mu) \right)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + 2(\bar{x} - \mu) \underbrace{\sum_{i=1}^{n} (x_i - \bar{x})}_{=0} + n(\bar{x} - \mu)^2$$
$$= (n-1)s^2 + n(\bar{x} - \mu)^2$$

より,

$$p(x_1, \ldots, x_n \mid \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \left((n-1)s^2 + n(\bar{x} - \mu)^2 \right) \right).$$

$\bar{x}$ と s^2 だけで書ける! $\square$

$\bar{x}$ と s^2 は正規分布モデルの十分統計量

標本分布内での確率的ゆらぎの大きさの例

n回中K回成功 → pを $\hat{p}=\frac{K}{n}$ で推定

$K \sim \mathrm{Binomial}(n,p)$ のとき，$E[K]=np$，$\mathrm{std}(K)=\sqrt{\mathrm{Var}(K)}=\sqrt{np(1-p)}$ であったので，

$\hat{p}=\frac{K}{n}$ とおくと，$E[\hat{p}]=\frac{1}{n}E[K]=p$，$\underline{\mathrm{std}(\hat{p})}=\frac{1}{n}\mathrm{std}(K)=\sqrt{\frac{p(1-p)}{n}}$.

($\hat{p}$はpの不偏推定量)　これをSEと書いて，<u>標準誤差</u> (standard error) と呼ぶ．

以上は二項分布内の様子．(数学的フィクション)

二項分布は，毎回独立に<u>未知の</u>確率pで当たりが出るルーレットをn回まわしたときの当たりの回数kの分布のモデル化になっている．

そのようなルーレットをn回まわしたとき，k回当たりが出たとする．← $\boxed{\text{データ}}$

そして，未知である当たりの確率pの推定値を$\frac{k}{n}$で求めることにしよう (最尤(ゆう)法)．

二項分布モデルを使って，推定値の$\frac{k}{n}$と真の値pのちがいの大きさを推定したい．(つづく)

$\boxed{\text{ポイント}}$ 未知の値pの推定値をデータ「n回中k回当たりが出た」から$\frac{k}{n}$と求めても，<u>誤差がどの程度であるか</u>を見積もらないと意味のある統計分析にならない！

ここで確率論的な計算を使う．

誤差の大きさの指標

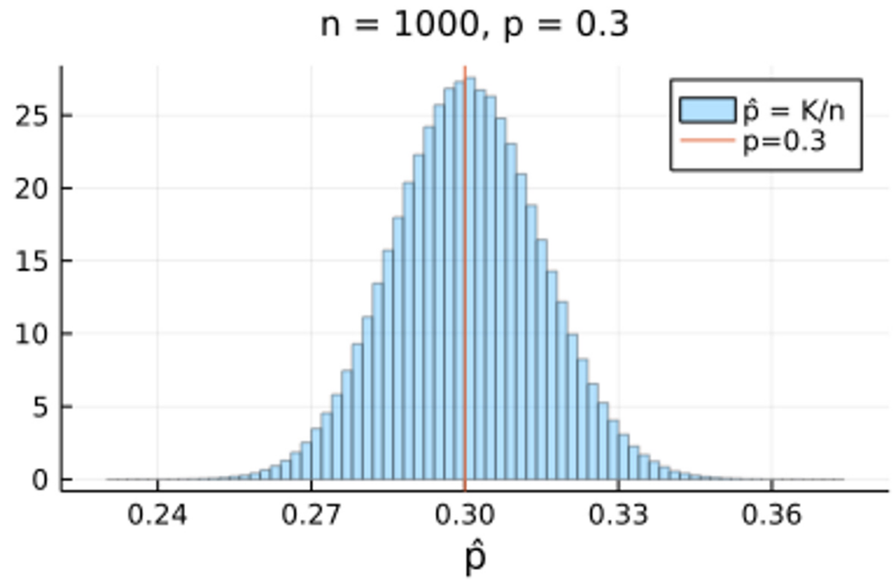
二項分布内での $\hat{p} = \dfrac{K}{n}$ の真の $\overbrace{\text{標準偏差}}$ の式は $\sqrt{\dfrac{p(1-p)}{n}}$ で未知の値 p を含む。

その p にデータから求めた推定値 k/n を代入して得られる $\sqrt{\dfrac{(k/n)(1-k/n)}{n}}$ を

真の標準偏差の推定値として使ったりする。(k=0, nのような場合はあぶないやり方。)

$\boxed{\text{数値例}}$ $n=1000$, $p=0.3$ のとき,

$$\mathrm{std}(\hat{p}) = \mathrm{std}\left(\frac{K}{n}\right) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.3\times 0.7}{1000}} \approx \underbrace{0.015}$$



$\hat{p} = \dfrac{K}{n}$ の分布

約2倍

$\hat{p} = \dfrac{K}{n}$ は

真の値の $p = 0.30$ から

$\pm\overbrace{0.03}$ 程度の範囲に95%以上が

分布している。

超重要必修自習問題

必修問題: 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).

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計算するのがこの問題の内容である. (注意・警告: ただし, 二項分布モデルの適用がこの場合に妥当であるかどうかは度外視する. 妥当であるか否かは目的によって異なる.)

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

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

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

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

- $P(K \ge 694844) = 2.5\%$ になるようなパラメータ p の値 p_L,
- $P(K \le 694844) = 2.5\%$ 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:

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

p_L, p_U の L, U は lower, upper の頭文字のつもりである.

確率やパラメータの数値は有効桁4桁まで求めよ. 0.00000000000000000001234 のように 0 を沢山含む表示は見難いので,

$$1.234 \times 10^{20} = 0.\underbrace{00000000000000000000}_{20}1234$$

のように書かずに,

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

のように書くこと. 1.234E−20, 1.234e−20 のどちらでもよい.

ヒント: $K \sim \mathrm{Binomia}(n, p)$ のとき,

$$P(K \ge k) = 1 - \mathrm{cdf}(\mathrm{Binomial}(n, p), k-1) = \mathrm{cdf}(\mathrm{Beta}(k, n-k+1), p),$$
$$P(K \le k) = \mathrm{cdf}(\mathrm{Binomial}(n, p), k) = 1 - \mathrm{cdf}(\mathrm{Beta}(k+1, n-k), p).$$

これらの公式から, 与えられた n と k について $P(K \ge k) = \alpha/2$ を満たす $p = p_L$ と $P(K \le k) = \alpha/2$ を満たす $p = p_U$ はそれぞれ次のように表されることがわかる:

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

WolframAlphaでは Beta は `BetaDistribution` と書く.

ヒント

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(モデルのデータへの適合度の指標の一種)を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).

リアルなデータが
二項分布モデルを通して見ると
どのように見えるか?

GitHubにある資料に答えと
詳しい解説がある.

X, Y と X, Z と Y, Z の各々の対が独立でも X, Y, Z 全体が独立でない例

右の $P(x,y,z)$ が定める分布にしたがう X, Y, Z はそのような例.

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

$P(x,y)=P(x,y,1)+P(x,y,0)=\frac{1}{4}$　($(x,y)=(1,1),(1,0),(0,1),(0,0)$ の4通り)

$P(x,z)=P(x,1,z)+P(x,0,z)=\frac{1}{4}$　((x,z) も4通り)

$P(y,z)=P(1,y,z)+P(0,y,z)=\frac{1}{4}$　((y,z) も4通り)

$P(x)=P(x,y=1)+P(x,y=0)=\frac{1}{2}$　($x=0,1$ で2通り), 同様にして, $P(y)=P(z)=\frac{1}{2}$.

ゆえに, $P(x,y)=P(x)P(y)$, $P(x,z)=P(x)P(z)$, $P(y,z)=P(y)P(z)$.

X, Y, Z は対ごとに独立.

しかし, $P(1,1,1)=\frac{1}{4}$ と $P(x=1)P(y=1)P(z=1)=\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\frac{1}{8}$ は等しくない.

X, Y, Z の全体は独立ではない.

つづく

つづき,

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.

$y = 1 =$ 効果有	$y = 0 =$ 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.

$z = 1 =$ 女性	$z = 0 =$ 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っ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)確率質量函数 $P(x, y)$ の表

	$y = 1 =$ 効果有	$y = 0 =$ 効果無
$x = 1 =$ 薬A	$P(x = 1, y = 1) = 1/4$	$P(x = 1, y = 0) = 1/4$
$x = 0 =$ 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函数 $P(x, z)$ の表

	$z = 1 =$ 女性	$z = 0 =$ 男性
$x = 1 =$ 薬A	$P(x = 1, z = 1) = 1/4$	$P(x = 1, z = 0) = 1/4$
$x = 0 =$ 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函数 $P(y, z)$ の表

	$y = 1 =$ 効果有	$y = 0 =$ 効果無
$z = 1 =$ 女性	$P(y = 1, z = 1) = 1/4$	$P(y = 0, z = 1) = 1/4$
$z = 0 =$ 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函数 $P(x, y, z)$ $P(x, y, z)$ の表

	$z = 1 =$ 女性	
	$y = 1 =$ 効果有	$y = 0 =$ 効果無
$x = 1 =$ 薬A	$P(1, 1, 1) = 1/4$	$P(1, 0, 1) = 0$
$x = 0 =$ 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函数 $P(x, y, z)$ の表

	$z = 0 =$ 男性	
	$y = 1 =$ 効果有	$y = 0 =$ 効果無
$x = 1 =$ 薬A	$P(1, 1, 0) = 0$	$P(1, 0, 0) = 1/4$
$x = 0 =$ 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.

確率変数達の共分散と相関係数

X, Y は確率変数とし，$\mu_X = E[X]$，$\sigma_X^2 = \mathrm{var}(X)$，$\mu_Y = E[Y]$，$\sigma_Y^2 = \mathrm{var}(Y)$ とおく．

共分散 (covariance) $\longleftrightarrow$ ベクトルの内積

自分で確認せよ．(GitHubに答えがある．)

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

Cauchy-Schwarz の不等式 $E[XY]^2 \leqq E[X^2]E[Y^2]$ より

$$|\mathrm{cov}(X,Y)| \leqq \mathrm{std}(X)\,\mathrm{std}(Y) \quad (\sigma_{XY} \leqq \sigma_X \sigma_Y).$$

GitHub に証明あり．線形代数の場合と同じ．

相関係数 (correlation coefficient)

$$\rho_{XY} = \mathrm{cor}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\mathrm{std}(X)\,\mathrm{std}(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \longleftrightarrow \frac{(\text{内積})}{(\text{ノルム})(\text{ノルム})} = \cos\theta.$$

無相関 (uncorrelated)

$\rho_{XY} = \mathrm{cor}(X,Y) = 0$ のとき，X と Y は<u>無相関</u>であるという．

(数学的には $\sigma_{XY} = \mathrm{cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = 0$ と同値．
ただし，$\sigma_{XY} \approx 0$ であっても $\rho_{XY} \neq 0$ のことがあるので注意．)

GitHubに答えがある．

例を挙げよ．

$X_1, \dots, X_n$ は，$i \neq j$ ならば X_i と X_j が無相関なとき，<u>対ごとに無相関</u>であるという．

$\boxed{\text{独立} \Rightarrow \text{無相関}}$

証明 X, Y は独立であると仮定する。

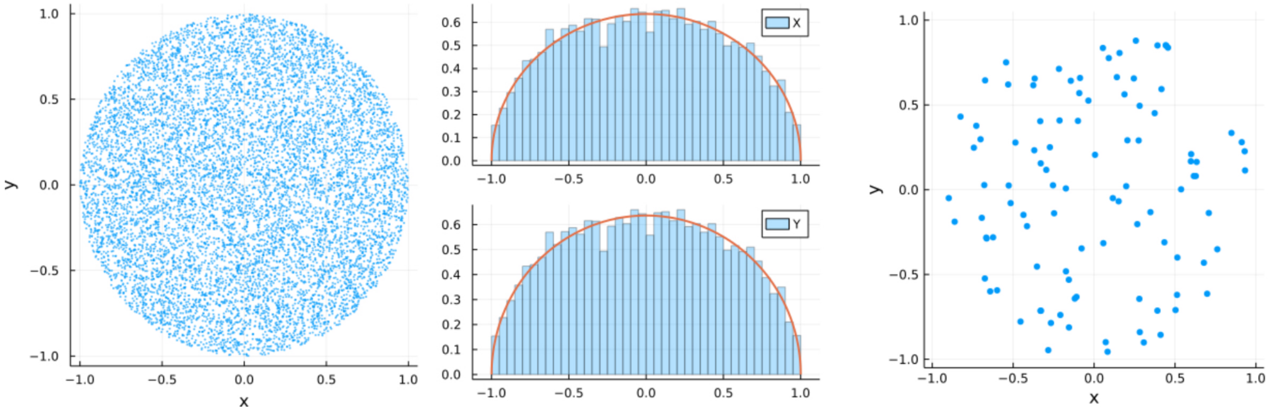
このとき、$E[f(X)g(Y)] = E[f(X)]E[g(Y)]$ なので、

$$\mathrm{cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] = E[X-\mu_X]E[Y-\mu_Y] = 0\cdot 0 = 0.$$

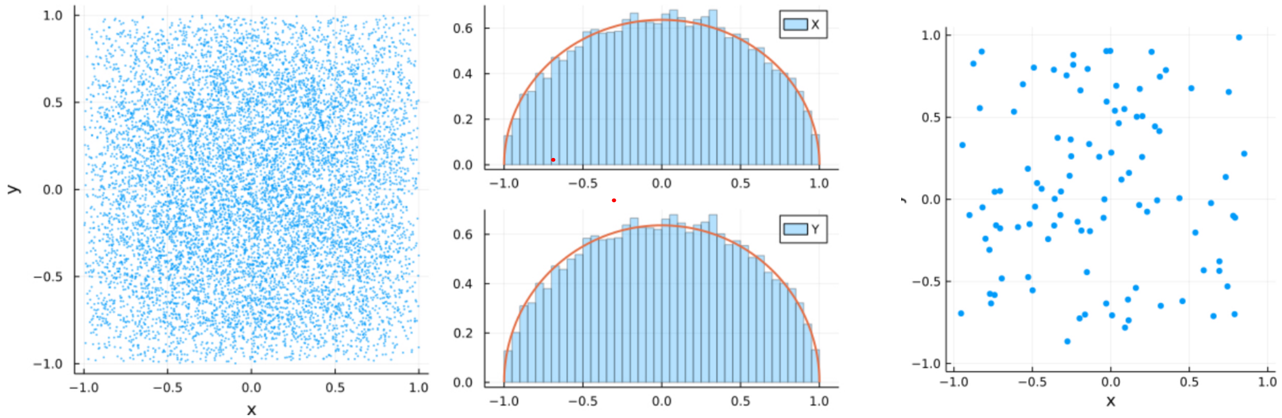
$\boxed{\text{q.e.d.}}$

$\boxed{\text{逆は成立しない。}}$

以下のとはちがう例を自分で作ってみよ。← 答えは GitHub にある。



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.

$\sigma_{XX} = \mathrm{cov}(X, X) = \mathrm{var}(X) = \sigma_X^2$ なので共分散の概念は分散の概念を含む.

$i \neq j$ のとき X_i と Y_j が無相関 $\Rightarrow$ $\mathrm{cov}\left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathrm{cov}(X_i, Y_i)$,

特に独立

$\left(i \neq j \text{ のとき } X_i \text{ と } X_j \text{ が無相関} \Rightarrow \mathrm{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{var}(X_i). \right)$

特に独立

← これ以後空気のごとく使われる.

証明 $\mu_{X_i} = E[X_i]$, $\mu_{Y_i} = E[Y_i]$ とおくと, $i \neq j$ のとき, X_i と Y_j が無相関であることより

$$i \neq j \Rightarrow \mathrm{cov}(X_i, Y_j) = E[(X_i - \mu_{X_i})(Y_j - \mu_{Y_j})] = 0.$$

ゆえに,

$$\mathrm{cov}\left(\sum_{i=1}^n X_i, \sum_{i=1}^n Y_i\right) = E\left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_{X_i}\right)\left(\sum_{j=1}^n Y_j - \sum_{j=1}^n \mu_{Y_j}\right)\right]$$

$\sum_i$ を $\sum_j$ に書き変えた.

$$= E\left[\sum_{i=1}^n (X_i - \mu_{X_i}) \sum_{j=1}^n (Y_j - \mu_{Y_j})\right]$$

$$= \sum_{i=1}^n \sum_{j=1}^n E[(X_i - \mu_{X_i})(Y_j - \mu_{Y_j})]$$ ← これは $i \neq j$ のとき 0

$$= \sum_{i=1}^n E[(X_i - \mu_{X_i})(Y_i - \mu_{Y_i})] = \sum_{i=1}^n \mathrm{cov}(X_i, Y_i).$$

q.e.d.

特に $X_1, \ldots, X_n$ が独立なとき,

$$E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i], \quad \mathrm{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{var}(X_i)$$

← 非常によく使われる.

医学研究でよく見るメタアナリシスに興味がある人向けの自習問題

答えは
GitHubにある.

2.11 問題(メタアナリシスの出発点): 共通の期待値と異なる分散を持つ確率変数の荷重平均

$X_1, \dots, X_n$ は対ごとに無相関な確率変数達であるとし, X_i 達は共通の期待値 μ と共通でない分散 $\sigma_i^2 > 0$ を持つと仮定する:

$$E[X_1] = \cdots = E[X_n] = \mu, \quad E[X_i] = \sigma_i^2 \quad (i = 1, \dots, n)$$

このとき, 共通の期待値 μ の推定量 $\hat{\mu}$ を荷重 w_i による荷重平均

$$\hat{\mu} = \sum_{i=1}^n w_i X_i, \quad \sum_{i=1}^n w_i = 1$$

の形で定義したい. 条件 $\sum_{i=1}^n w_i = 1$ より, 不偏性 $E[\hat{\mu}] = \mu$ がただちに導かれる. このとき, $\hat{\mu}$ の分散 $\operatorname{var}(\hat{\mu})$ を最小化する荷重 w_i は

$$w_i = \frac{1/\sigma_i^2}{\sum_{i=1}^n 1/\sigma_i^2}, \quad \text{i.e.} \quad \hat{\mu} = \frac{\sum_{i=1}^n X_i/\sigma_i^2}{\sum_{i=1}^n 1/\sigma_i^2}$$

になることを示せ. さらに, このとき

$$\operatorname{var}(\hat{\mu}) = \frac{1}{\sum_{i=1}^n 1/\sigma_i^2} < \min(\sigma_1^2, \dots, \sigma_n^2)$$

となることも示せ.

注意: この結果はことなる分散を持つ複数の推定量をうまくまとめることによって, 分散がより小さな推定量を作れることを意味している. これは同一の量に関する異なる研究の結果をまとめるメタアナリシスの出発点になる.

標本(サンプル,データ)の平均, 分散, 共分散, 相関係数

$E[X]$, $\mathrm{var}(X)$, $\mathrm{cov}(X,Y)$, $\mathrm{cor}(X,Y)$ とはちがうものであることに注意!

現実から数値データの場合

数値のデータ $(x_1, y_1), \dots, (x_n, y_n)$ について,

標本平均 $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\left(\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2,\ \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i,\ \text{etc.}\right)$

(不偏補正されていない)標本分散 $\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \overline{x^2} - (\bar{x})^2$

不偏分散 $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}\left(\overline{x^2} - (\bar{x})^2\right).$

どうして $n-1$ でわるかは後で説明する.

証明せよ.

不偏共分散 $s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{n}{n-1}\left(\overline{xy} - \bar{x}\,\bar{y}\right)$

相関係数 $r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{1/2}\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)^{1/2}}$ $\leftrightarrow \cos\theta$

確率変数の場合

以下は $E[X]$, $var(X)$, $cov(X,Y)$, $cor(X,Y)$ とはちがうものであることに注意!

確率変数達 $(X_1, Y_1), \dots, (X_n, Y_n)$ は以下をみたしていると仮定する:

(1) $E[X_i] = \mu_X$, $E[Y_i] = \mu_Y$ (共通の期待値 (μ_X, μ_Y) を持つ)

(2) $var(X_i) = \sigma_X^2$, $var(Y_i) = \sigma_Y^2$, $cov(X_i, Y_i) = \sigma_{XY}$ (共通の分散と共分散を持つ)

(3) $i \neq j$ のとき, $cov(X_i, X_j) = cov(Y_i, Y_j) = cov(X_i, Y_j) = 0$ (無相関)

特にiidならば成立

標本平均 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$, $\overline{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$, etc.

不偏分散 $S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{n}{n-1}\left(\overline{X^2} - (\overline{X})^2\right)$. 証明せよ.

どうして $n-1$ で割るかは後で説明する.

不偏共分散 $S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \frac{n}{n-1}\left(\overline{XY} - \overline{X}\,\overline{Y}\right)$

相関係数 $R_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\left(\sum_{i=1}^{n}(X_i - \overline{X})^2\right)^{1/2}\left(\sum_{i=1}^{n}(Y_i - \overline{Y})^2\right)^{1/2}}$ $\longleftrightarrow$ $\cos\theta$

これらは確率変数 X_i, Y_i の函数なので確率変数になる.

$E[X]$, $var(X)$, $cov(X,Y)$, $cor(X,Y)$ (←これらは数値になる) とはちがうものである!

標本平均達の期待値, 分散, 共分散

$$E[\bar{X}] = \mu_X, \quad E[\bar{Y}] = \mu_Y, \quad \mathrm{Var}(\bar{X}) = \frac{\sigma_X^2}{n}, \quad \mathrm{Var}(\bar{Y}) = \frac{\sigma_Y^2}{n}, \quad \mathrm{cov}(\bar{X},\bar{Y}) = \frac{\sigma_{XY}}{n},$$

① ② ③ ④ ⑤

$\boxed{\text{証明}}$ ①のXがYの場合⇒②で, ⑤のX=Yの場合⇒③,④なので①,⑤を示せば十分.

$$E[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E[X_i]}_{=\mu_X} = \mu_X.$$

$$\mathrm{cov}(\bar{X},\bar{Y}) = E[(\bar{X}-\mu_X)(\bar{Y}-\mu_Y)] = E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_X)\cdot\frac{1}{n}\sum_{j=1}^{n}(Y_j-\mu_Y)\right]$$

$$= \frac{1}{n^2}\sum_{i,j=1}^{n} \underbrace{E[(X_i-\mu_X)(Y_j-\mu_Y)]}_{i=j\text{ のとき }\sigma_{XY},\ i\neq j\text{ のとき }0} = \frac{1}{n^2}\, n\sigma_{XY} = \frac{\sigma_{XY}}{n}.$$

$\boxed{\text{q.e.d.}}$

$\boxed{\text{注意}}$ $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ の標準偏差 $\mathrm{std}(\bar{X}) = \sqrt{\mathrm{var}(\bar{X})} = \frac{\sigma_X}{\sqrt{n}}$ は,

$\bar{X}$ の実現値で値 μ_X を推定するときの誤差の大きさの指標になる.

($SE = \mathrm{std}(\bar{X})$ と書いて, μ_X の推定量 $\bar{X}$ の<u>標準誤差</u> (standard error) と呼ぶ.)

$\bar{X}$ の標準偏差は $n\to\infty$ で 0 に収する!

□

不偏(共)分散の定義でどうして $n-1$ で割るか

$\boxed{\text{答え}}$ $E[S_X^2]=\sigma_X^2$, $E[S_{XY}]=\sigma_{XY}$ となるから.

(このとき, S_X^2, S_{XY} はそれぞれ σ_X^2, σ_{XY} の不偏推定量であるという.)

$\boxed{\text{証明}}$ $E[S_X^2]=\sigma_X^2$ は $E[S_{XY}]=\sigma_{XY}$ の $X=Y$ の場合なので後者のみを示せばよい.

$$\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})=\sum_{i=1}^{n}\left((X_i-\mu_X)-(\bar{X}-\mu_X)\right)\left((Y_i-\mu_Y)-(\bar{Y}-\mu_Y)\right)$$

$$=\sum_{i=1}^{n}(X_i-\mu_X)(Y_i-\mu_Y)-\Big(\underbrace{\sum_{i=1}^{n}(X_i-\mu_X)}_{=n(\bar{X}-\mu_X)}\Big)(\bar{Y}-\mu_Y)-(\bar{X}-\mu_X)\Big(\underbrace{\sum_{i=1}^{n}(Y_i-\mu_Y)}_{=n(\bar{Y}-\mu_Y)}\Big)+n(\bar{X}-\mu_X)(\bar{Y}-\mu_Y)$$

$$=\underbrace{\sum_{i=1}^{n}(X_i-\mu_X)(Y_i-\mu_Y)}_{\text{これの期待値は }\sum_{i=1}^{n}\underbrace{\mathrm{cov}(X_i,Y_i)}_{=\sigma_{XY}}=n\sigma_{XY}}-n\underbrace{(\bar{X}-\mu_X)(\bar{Y}-\mu_Y)}_{\text{これの期待値は }\mathrm{cov}(\bar{X},\bar{Y})=\frac{\sigma_{XY}}{n}}$$

ゆえに, $\underbrace{E\left[\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})\right]=n\sigma_{XY}-n\frac{\sigma_{XY}}{n}=(n-1)\sigma_{XY}}$.

(これが $n-1$ でわる理由)

したがって, $E[S_{XY}]=E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})\right]=\frac{1}{n-1}(n-1)\sigma_{XY}=\sigma_{XY}$.

$\boxed{\text{q.e.d.}}$

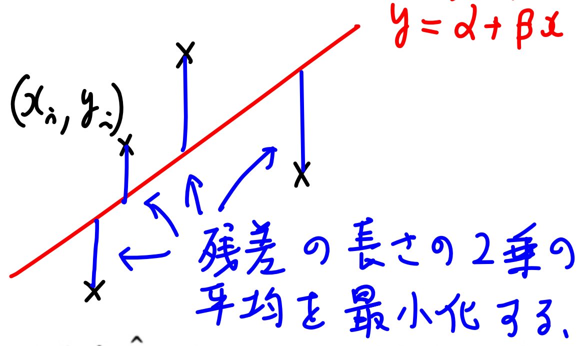
4 最小二乗法による線形回帰

$\boxed{\bar{x}, \bar{y}, s_x^2, s_y^2, s_{xy} \text{ の応用}}$

最小二乗法による単純な線形回帰について説明しよう. 以下, $x_1, \dots, x_n$ のうち2つは互いに異なると仮定する.

与えられたデータ $(x_1, y_1), \dots, (x_n, y_n)$ における y_i 達の値達を x_i 達の値の共通の一次函数 $\alpha + \beta x_i$ (α, β は i によらない)で近似することを考える. そのとき,

$$\varepsilon_i = y_i - (\alpha + \beta x_i)$$



を **残差** と呼ぶ. このとき, **平均二乗残差**

$$f(\alpha, \beta) = \frac{1}{n} \sum_{i=1}^{n} (y_i - (\alpha + \beta x_i))^2$$

を最小化することを **最小二乗法** と呼び, 最小二乗法で求めた係数 α, β の値を **回帰係数** と呼び, $\hat{\alpha}, \hat{\beta}$ と書き, それに対応する一次函数

$$y = \hat{\alpha} + \hat{\beta} x$$

を **回帰直線** (regression line)と呼ぶ. このとき, x を **説明変数** や **独立変数** などと呼び, y を **目的変数**, **従属変数**, **反応変数** などと呼ぶ.

これは一次函数によるデータの近似になっているので, **線形回帰** (ordinary least squares, OLS, linear regression)と呼ばれる. この場合には説明変数が x の1つだけなので **単回帰** (simple regression, simple linear regression)と呼ぶこともある. 説明変数が多変数になった場合には **重回帰** (multiple regression)と呼ぶ.

以下ではさらに, **最小平均二乗残差** (もしくは単に **平均二乗残差**)の値を次のように書くことにする:

$$\hat{\sigma}^2 = f(\hat{\alpha}, \hat{\beta}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - (\hat{\alpha} + \hat{\beta} x_i))^2$$

注意: 最小二乗法は説明変数 x と目的変数 y について対称ではないので, それらの変数の立場を逆転して作った回帰直線は元の回帰直線に一致しない. 下の方で紹介するGalton (1886)の回帰直線の節も参照せよ.

GitHubの方を見よ.

最小二乗法による線形回帰の解は $\bar{x}, \bar{y}, s_x^2, s_y^2, s_{xy}$ で書ける!

4.1 問題: 最小二乗法による線形回帰の公式の導出

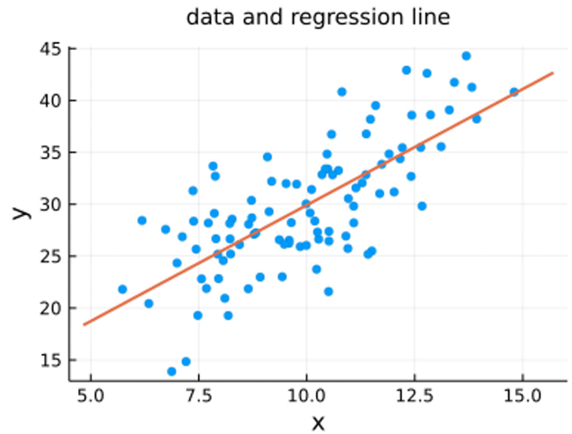
データ x_i の標本平均と不偏分散をそれぞれ $\bar{x}$, s_x^2 と書き, y_i の標本平均と不偏分散をそれぞれ $\bar{y}$, s_y^2 と書き, x_i と y_i の不偏共分散を s_{xy} と書くとき, 以上の $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ は次のように表されることを示せ:

$$\hat{\beta} = \frac{s_{xy}}{s_x^2}, \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \quad \hat{\sigma}^2 = \frac{n-1}{n}\frac{s_x^2 s_y^2 - s_{xy}^2}{s_x^2}.$$

```
α̂ = 7.562346117161635
β̂ = 2.235682111246986
σ̂ = 4.148269982085684
```

最小二乗法の線形回帰の例 ⟶

linear regression



以下ではこの公式をみとめて使って良い。

① $\hat{\beta} = \dfrac{s_{xy}}{s_x^2}$, $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ で回帰直線 $y = \hat{\alpha} + \hat{\beta}x$ が求まる。

② $\hat{\sigma}^2 = \dfrac{n-1}{n}\dfrac{s_x^2 s_y^2 - s_{xy}^2}{s_x^2}$ で最小平均二乗残差が求まる。

$\boxed{\text{自習問題}}$ 上の公式が成立する理由を知りたい人は自分で証明してみよ!

$\boxed{\text{ヒント}}$ $f(\alpha, \beta)$ を α, β で偏微分して 0 とおく。

$\boxed{\text{ヒント}}$ $f(\alpha, \beta)$ を「平方完成形」に変形してみよ。

GitHubに答えがある。

重要必修自習問題

$\bar{x}, \bar{y}, s_x^2, s_y^2, s_{xy}$ の計算の例

4.2 必修問題: 最小二乗法の計算例 (Anscombe's quartet)

答えはGitHubにある.

以下に示した $(x_1, y_1), \ldots, (x_n, y_n)$ 型の4つのデータのそれぞれについて,

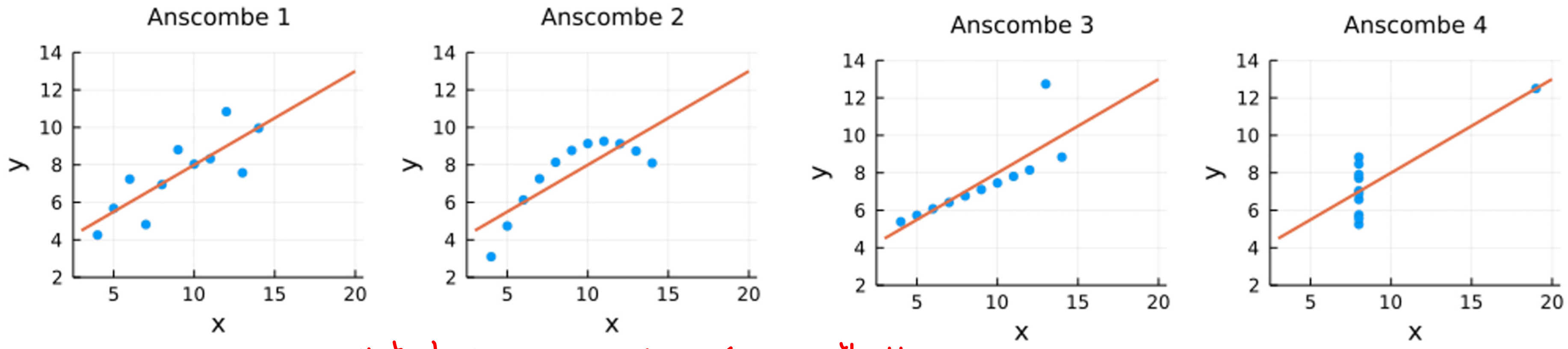
- x_i 達の標本平均 $\bar{x}$
- y_i 達の標本平均 $\bar{y}$
- x_i 達の不偏分散 s_x^2,
- y_i 達の不偏分散 s_y^2,
- x_i 達と y_i 達の不偏共分散 s_{xy},
- 回帰直線 $y = \hat{\alpha} + \hat{\beta}x$ の係数 $\hat{\alpha}, \hat{\beta}$ と最小平均二乗残差の平方根 $\hat{\sigma}$

アンスコムの例

データ1のみが最小二乗法による線形回帰が適切

を求めよ. それぞれについて小数点以下第1桁まで求めよ. $n=11$

1. (10, 8.04), (8, 6.95), (13, 7.58), (9, 8.81), (11, 8.33), (14, 9.96), (6, 7.24), (4, 4.26), (12, 10.84), (7, 4.82), (5, 5.68)
2. (10, 9.14), (8, 8.14), (13, 8.74), (9, 8.77), (11, 9.26), (14, 8.1), (6, 6.13), (4, 3.1), (12, 9.13), (7, 7.26), (5, 4.74)
3. (10, 7.46), (8, 6.77), (13, 12.74), (9, 7.11), (11, 7.81), (14, 8.84), (6, 6.08), (4, 5.39), (12, 8.15), (7, 6.42), (5, 5.73)
4. (8, 6.58), (8, 5.76), (8, 7.71), (8, 8.84), (8, 8.47), (8, 7.04), (8, 5.25), (19, 12.5), (8, 5.56), (8, 7.91), (8, 6.89)



できればこの図も自分で描いてみてください.

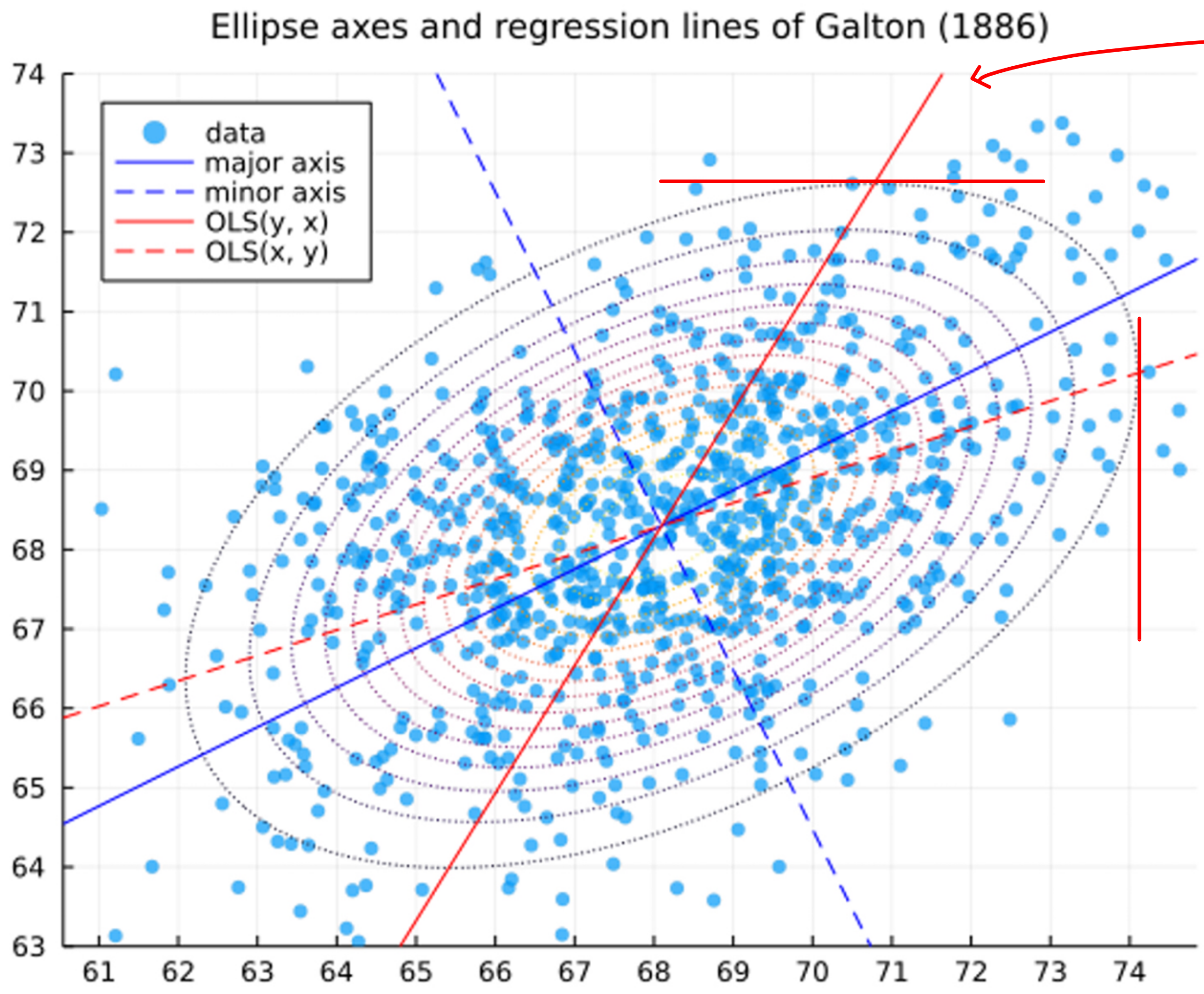
ゴルトン氏のグラフの再現

タテ軸：親の身長，ヨコ軸：成人した子の身長

Galton, Francis. Regression towards mediocrity in hereditary stature.
Journal of the Anthropological Institute, Vol. 5 (1886), 246-263.
https://galton.org/bib/JournalItem.aspx_action=view_id=157

楕円とx軸に平行な直線の接点を通る

回帰直線 $x=\hat{\alpha}'+\hat{\beta}'y$



楕円の長軸

回帰直線

$y=\hat{\alpha}+\hat{\beta}x$

楕円とy軸に平行な直線の接点を通る

詳しくはGitHubを参照

独立同分布な確率変数達の標本平均と不偏分散の分布

(中心極限定理に向けて)

$X_1, X_2, \dots, X_n$ は各々が分布Dにしたがう独立同分布な確率変数達とする.

X_i たち共通の期待値と分散をそれぞれ μ, σ^2 と書く. このとき,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i \quad (\text{標本平均}), \quad S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2 \quad (\text{不偏分散})$$

も確率変数になり, 次が成立することをすでに示してある:

$$E[\bar{X}] = \mu, \quad E[S^2] = \sigma^2, \quad \mathrm{var}(\bar{X}) = \frac{\sigma^2}{n}. \longleftarrow \iff \mathrm{std}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

分布Dの歪度 (わいど, skewness) $\bar{\kappa}_3$ と尖度 (せんど, kurtosis) $\bar{\kappa}_4$ が次のように定義される:

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

(D= Normal(μ, σ) のとき $\bar{\kappa}_3 = \bar{\kappa}_4 = 0$ となる.)

このとき, $\bar{X}$ と S^2 の共分散と S^2 の分散は以下のように表される:

n が大きいとき,

$$\mathrm{cov}(\bar{X}, S^2) = \sigma^3 \frac{\bar{\kappa}_3}{n}, \quad \mathrm{var}(S^2) = \sigma^4\left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}\right) \approx \sigma^4 \frac{\bar{\kappa}_4 + 2}{n}$$

証明はGitHubの方を見よ. めんどうな長い計算が必要.

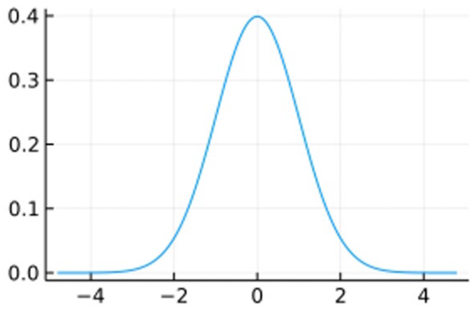
$(\bar{X}, S^2)$ について基本的な次の5つの値が分布Dからどのように決まるかがわかった:

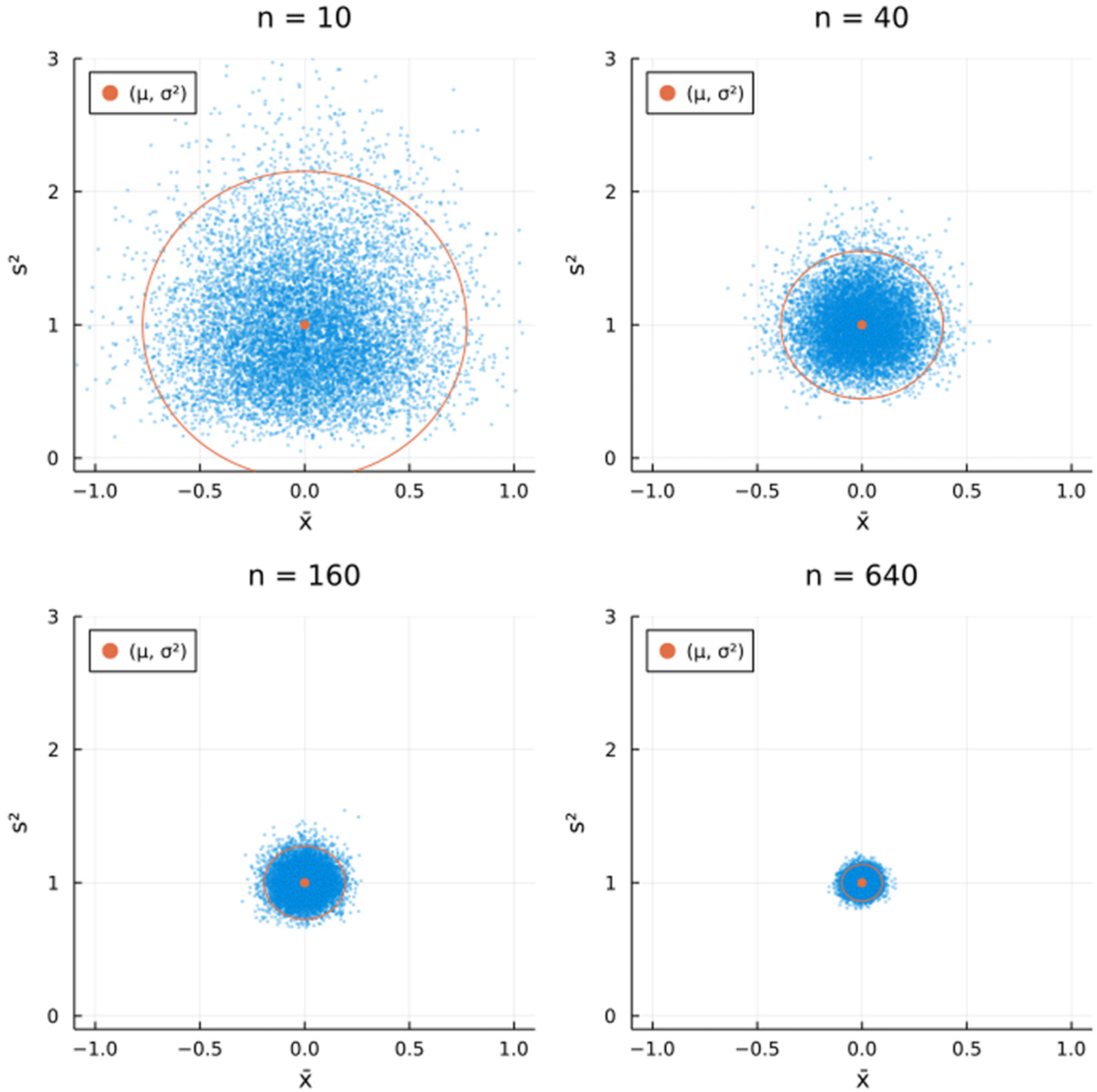
$$E[\bar{X}] = \mu, \quad E[S^2] = \sigma^2, \quad \mathrm{var}(\bar{X}) = \frac{\sigma^2}{n}, \quad \mathrm{cov}(\bar{X}, S^2) = \sigma^3\frac{\bar{\kappa}_3}{n}, \quad \mathrm{var}(S^2) = \sigma^4\left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}\right).$$

前ページのような式を見せられてもイメージがわかないと思うので、
$(\bar{x}, s^2)$ の乱数を1万個生成してプロットしてみよう。

理解のためにはこれ重要!

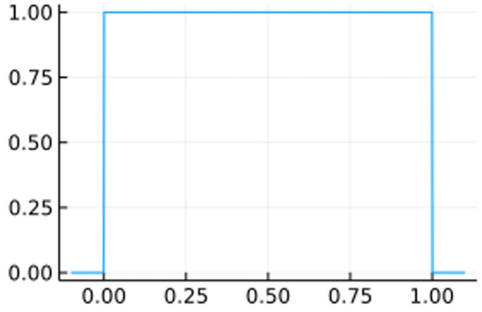
```
Normal{Float64}(μ=0.0, σ=1.0)
μ = 0.0
σ² = 1.0
skewness = 0.0
kurtosis = 0.0
```

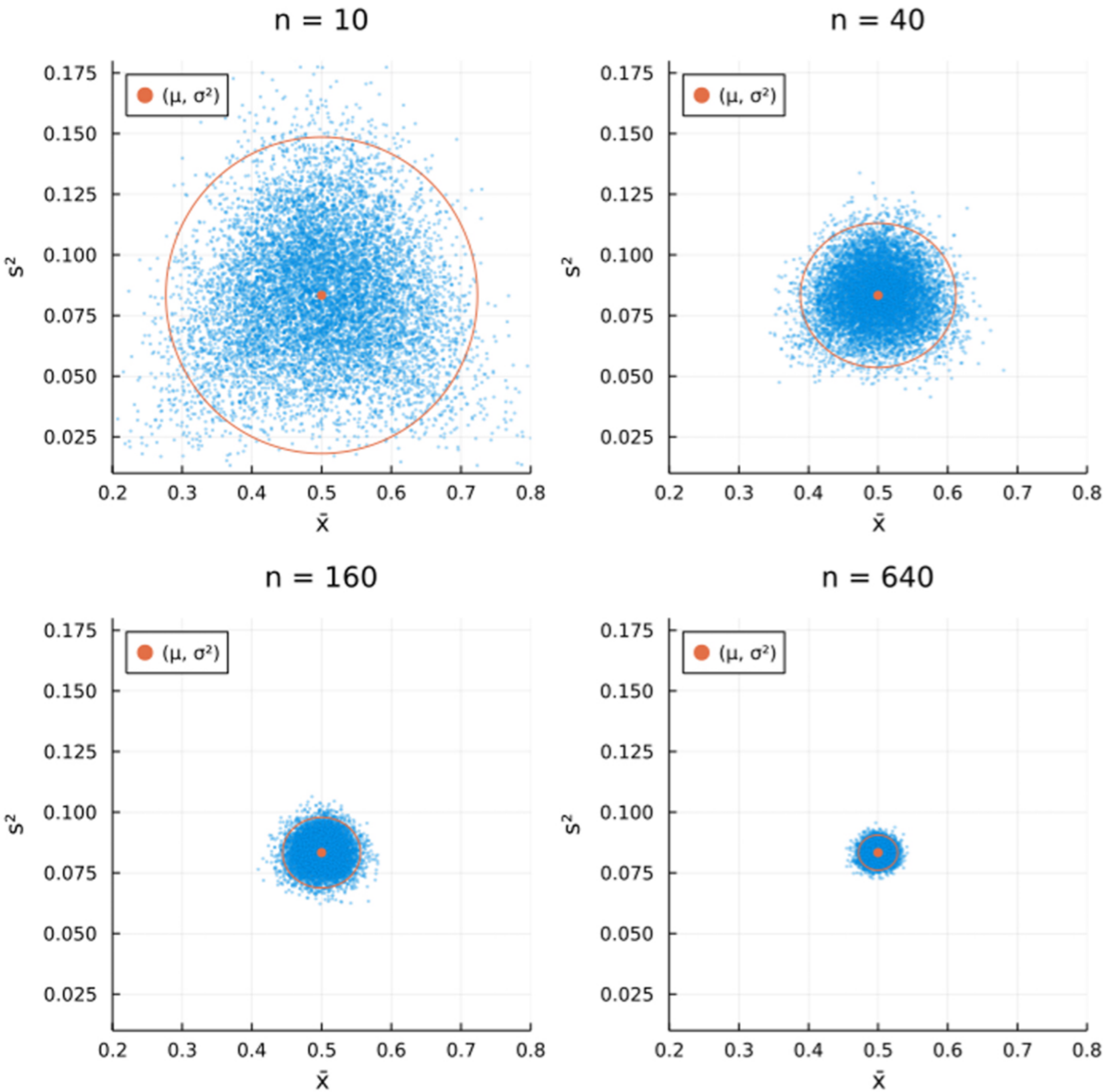




正規分布の場合。

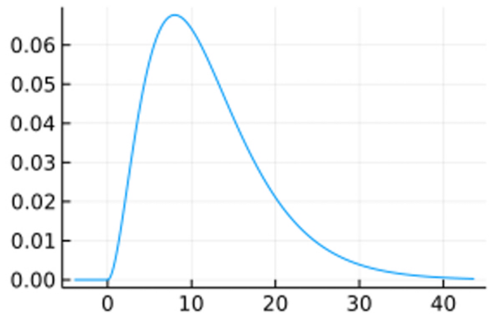
```
Uniform{Float64}(a=0.0, b=1.0)
μ = 0.5
σ² = 0.08333333333333333
skewness = 0.0
kurtosis = -1.2
```

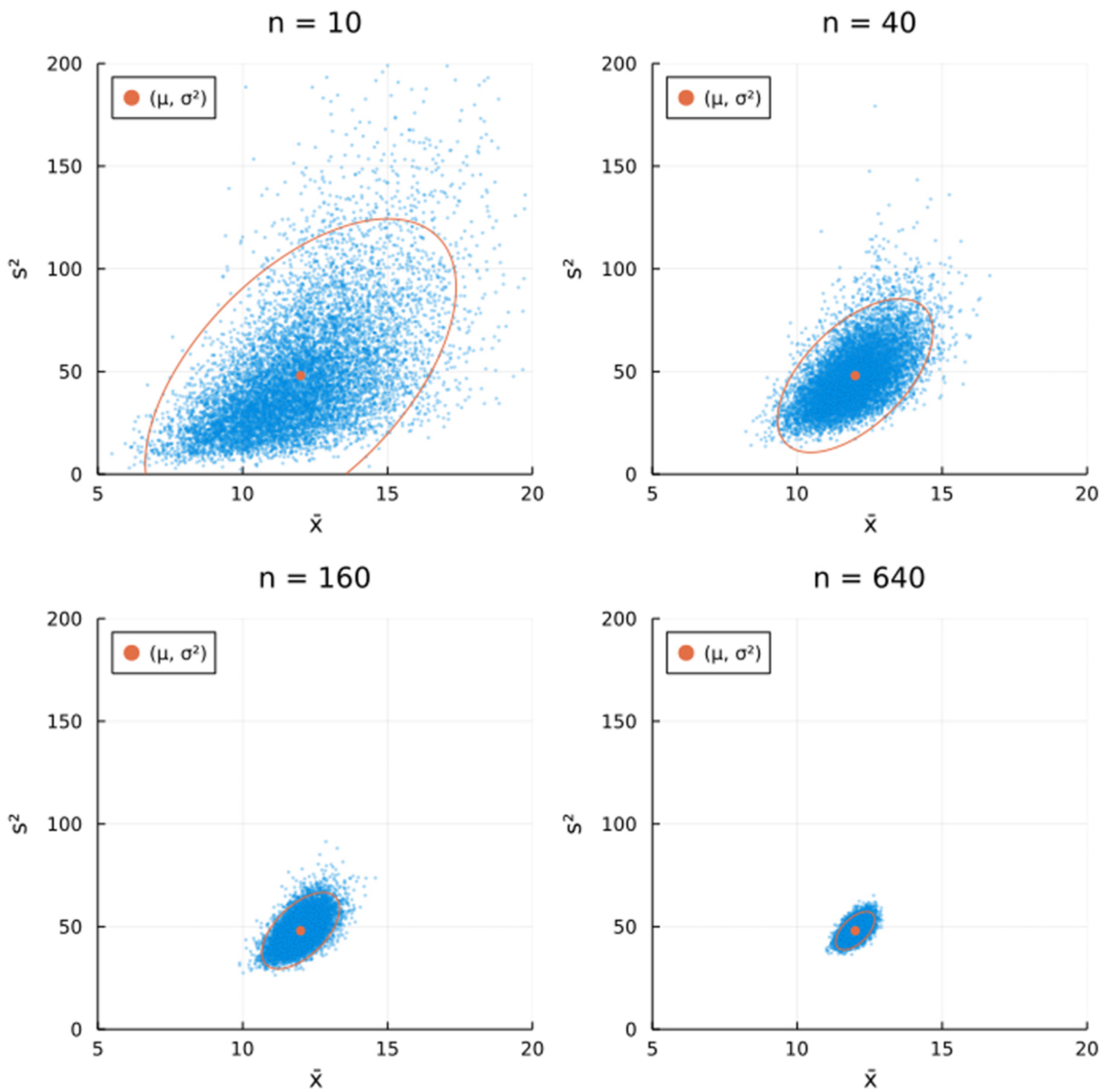




正規分布の場合に
非常に近い

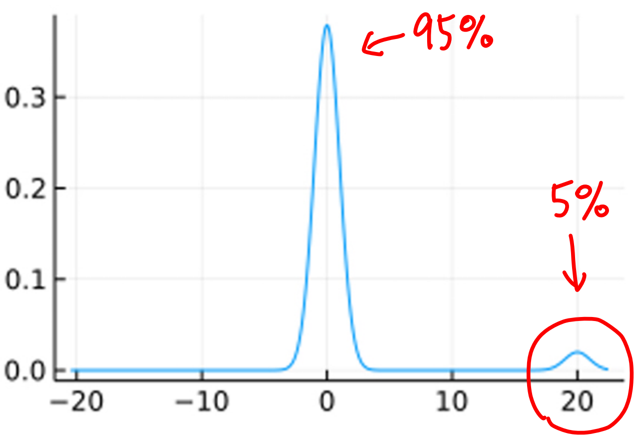
```
Gamma{Float64}(α=3.0, θ=4.0)
μ = 12.0
σ² = 48.0
skewness = 1.1547005383792517
kurtosis = 2.0
```

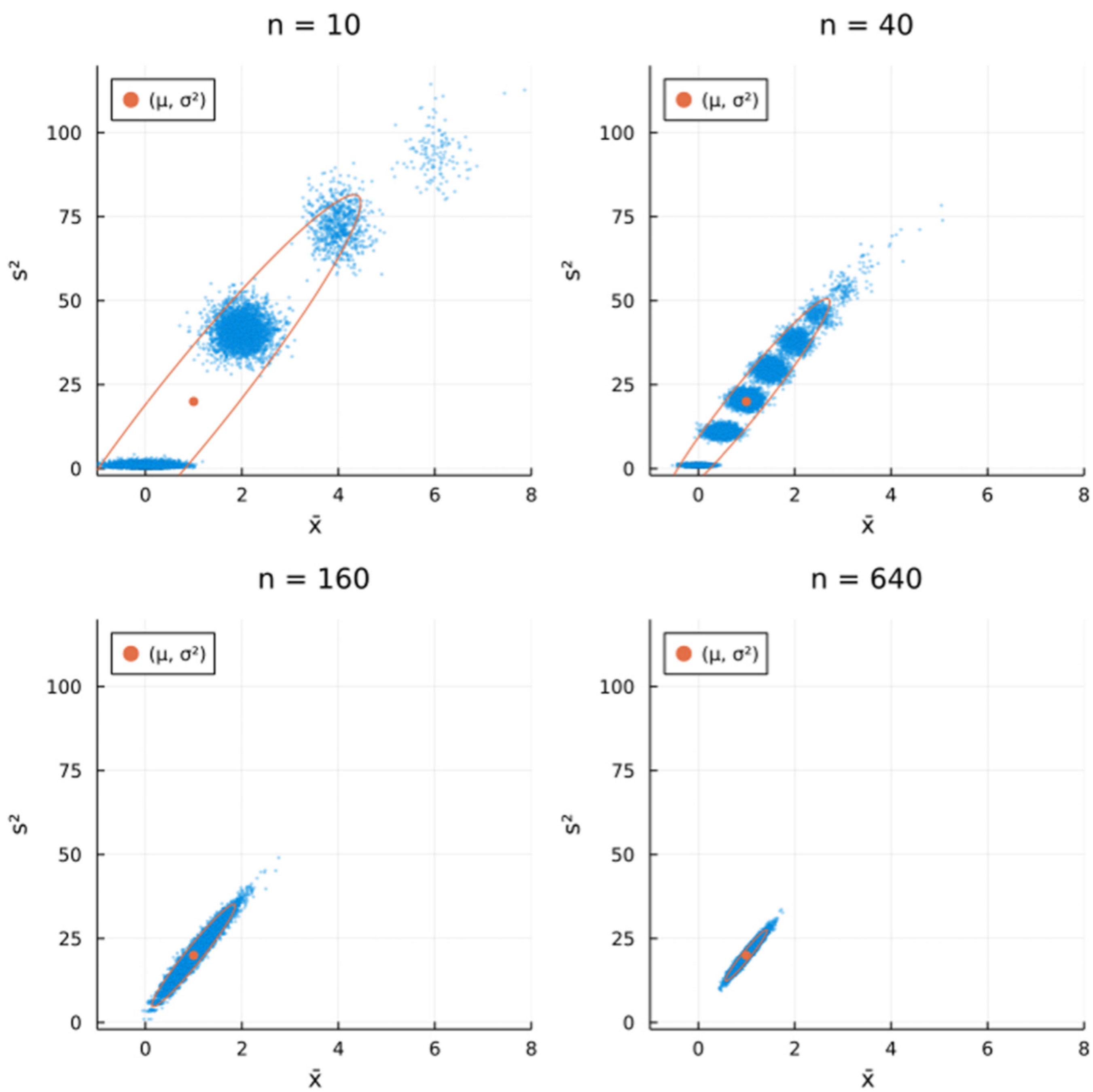




左右非対称な分布の標本分布における$(\bar{X}, S^2)$の分布は左右非対称になる。

```
μ = 1.0
σ² = 20.0
skewness = 3.823676241524649
kurtosis = 13.58499999999097
```





95%の標準正規分布と
5%の別の正規分布を
混ぜた混合正規分布

この5%のせいで
非常に「面白いこと」
になってしまっている。

GitHubの方には他の例もあるので見ておいてください。

$\boxed{\text{正規分布の場合}}$ (結果のみ, 証明はGitHubの資料にある.)

$X_1, X_2, \dots, X_n$ が平均 μ, 分散 σ^2 の正規分布のサイズ n の標本分布にしたがう確率変数達であるとき,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ (\text{標本平均}), \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}) \ (\text{不偏分散})$$

は独立になり,

$$\bar{X} \sim \mathrm{Normal}\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = \mu + \sqrt{\frac{\sigma^2}{n}}\,\mathrm{Normal}(0,1)$$

$$S^2 \sim \mathrm{Gamma}\left(\frac{n-1}{2}, \frac{2\sigma^2}{n-1}\right) = \frac{\sigma^2}{n-1}\mathrm{Gamma}\left(\frac{n-1}{2}, 2\right) = \frac{\sigma^2\,\mathrm{Chisq}(n-1)}{n-1}$$

が成立する. ゆえに,

$$Z := \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}, \quad Y = \frac{(n-1)S^2}{\sigma^2}, \quad T = \frac{Z}{\sqrt{Y/(n-1)}} = \frac{Z}{\sqrt{S^2/n}}$$

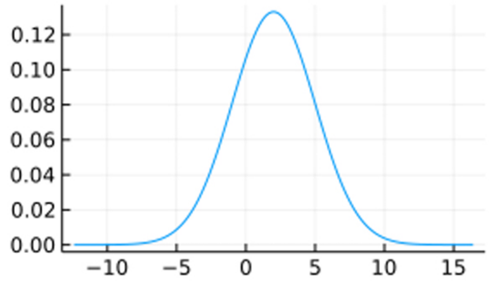
とおくと, Z と Y は独立で, 次が成立する:

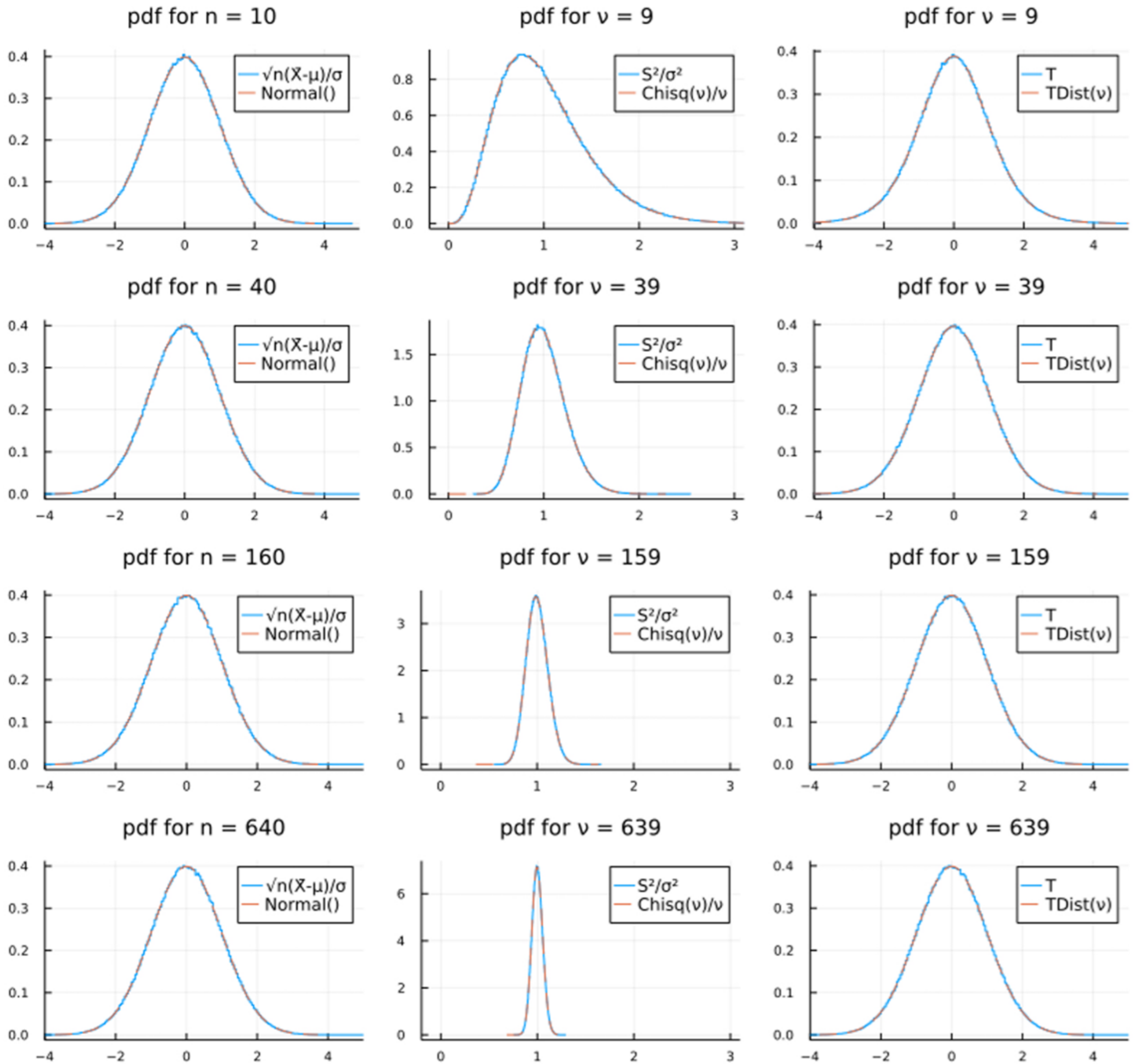
$$Z \sim \mathrm{Normal}(0,1), \quad Y \sim \mathrm{Chisq}(n-1), \quad T \sim \mathrm{TDist}(n-1) \qquad \square$$

以上の結果には再度ふれることになると思われる.

前ページのように式だけで説明されてもイメージがわかない。グラフを描いてみよう！ これ大事！

```
Normal{Float64}(μ=2.0, σ=3.0)
μ = 2.0
σ² = 9.0
skewness = 0.0
kurtosis = 0.0
```





$Z = \dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma}$ の分布は Normal(0,1) に一致。

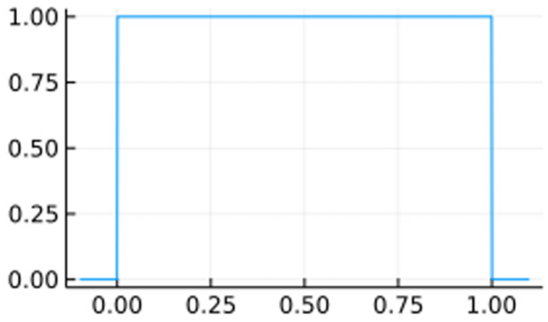
$S^2/(n-1)$ の分布は Chisq(n-1)/(n-1) に一致。

$T = \dfrac{Z}{\sqrt{S^2/n}}$ の分布は TDist(n-1) に一致

一致しすぎてイメージがわきにくい。

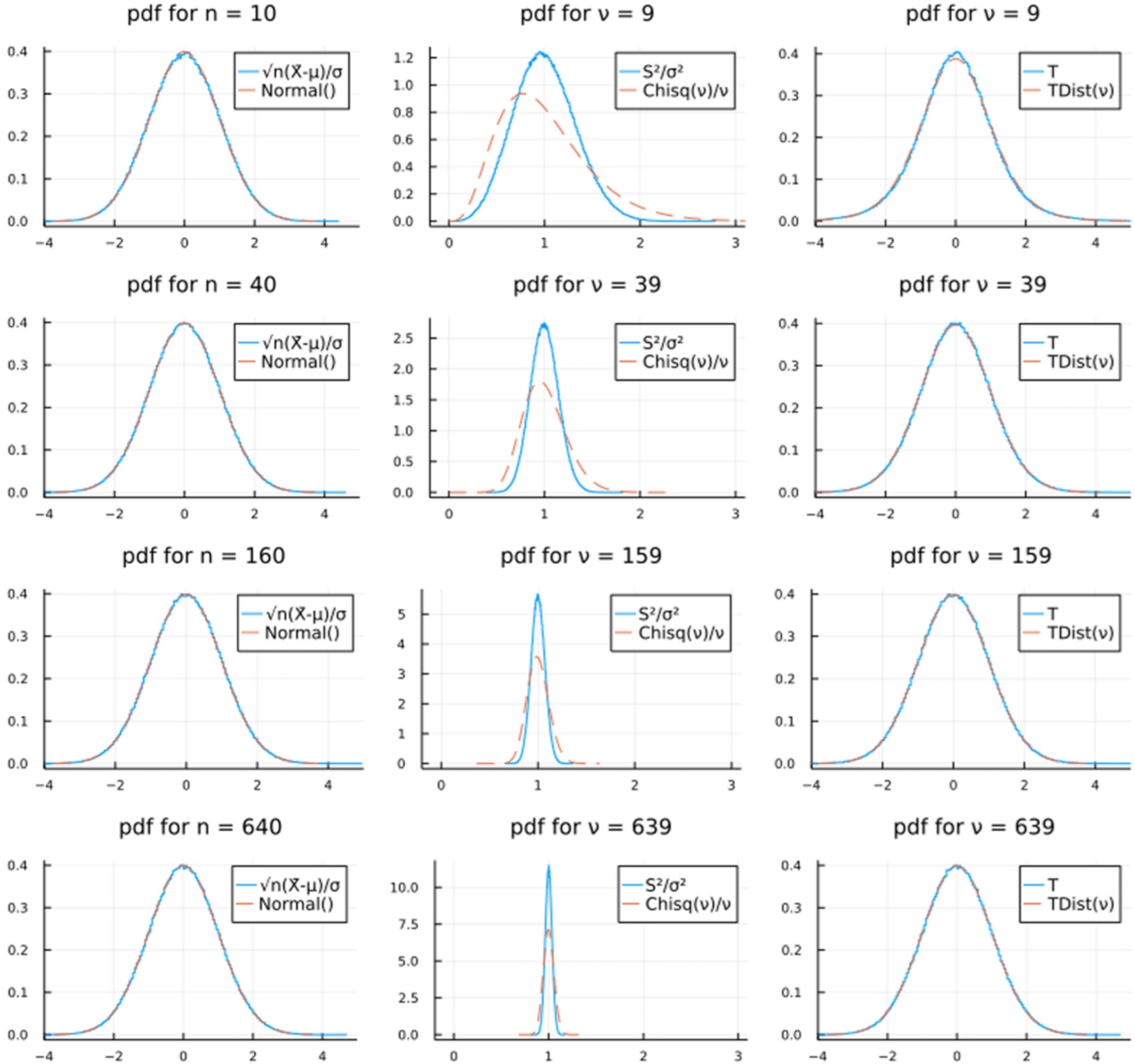
正規分布以外の標本分布で Z, S^2/σ^2, T の分布をプロットしてみよう。

```
Uniform{Float64}(a=0.0, b=1.0)
μ = 0.5
σ² = 0.08333333333333333
skewness = 0.0
kurtosis = -1.2
```

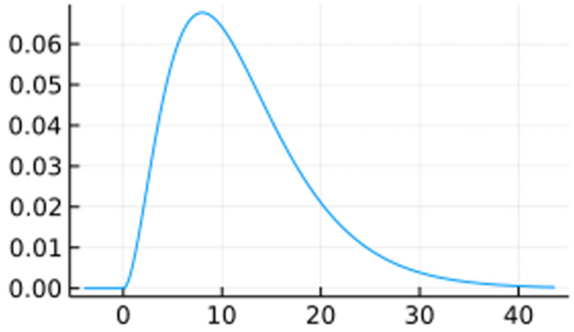


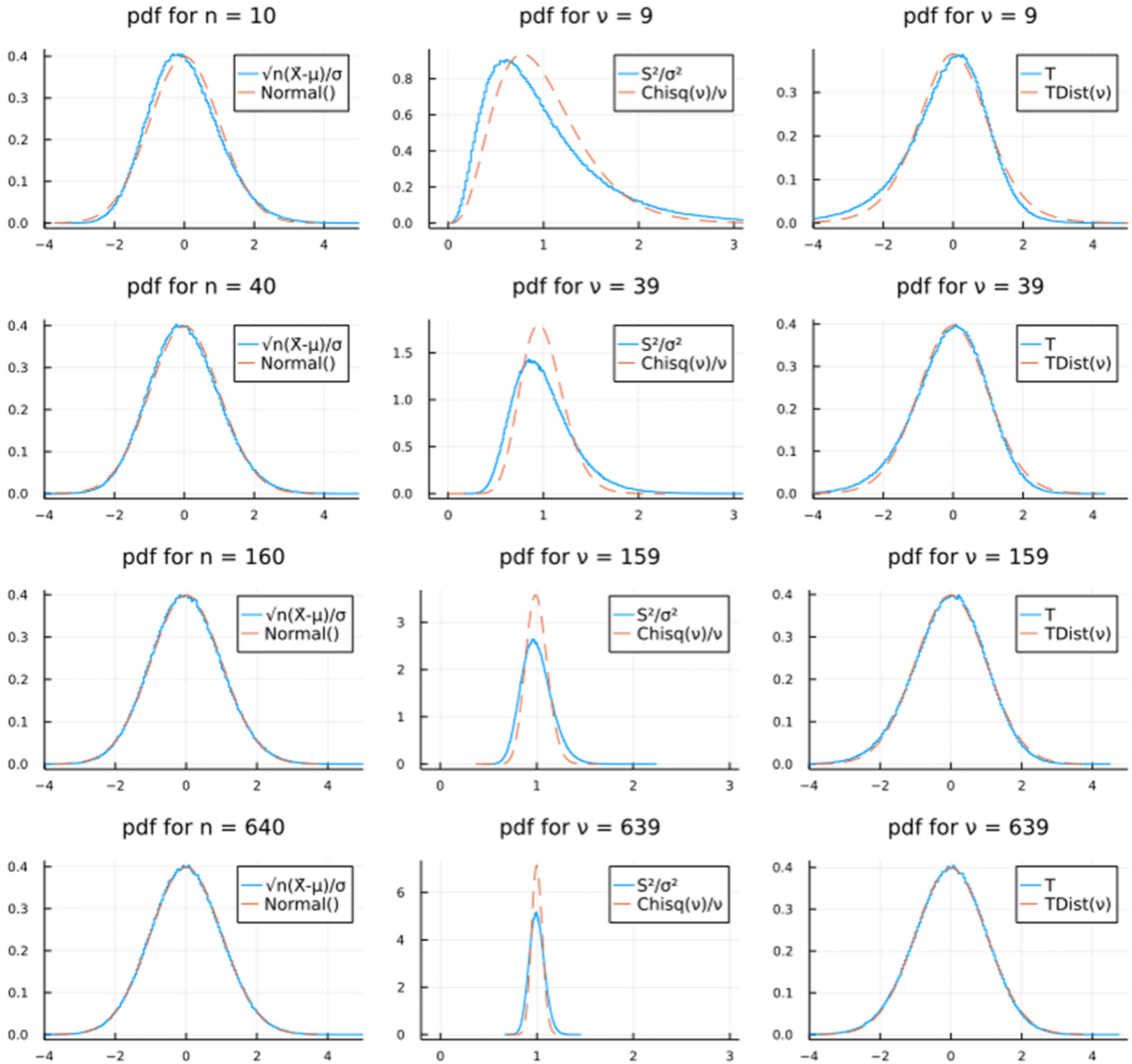
ZとTは正規分布の場合と同様に Normal(0,1), TDist(n−1) とよく一致している。

S^2/σ^2 は一致していない。



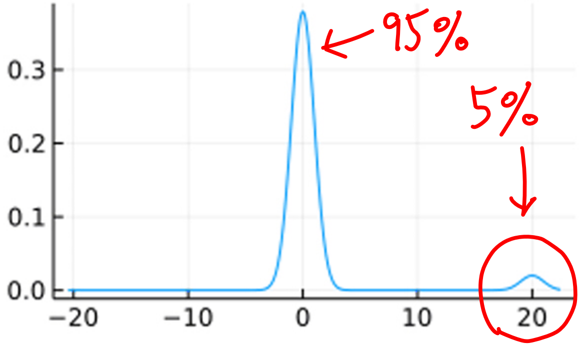
```
Gamma{Float64}(α=3.0, θ=4.0)
μ = 12.0
σ² = 48.0
skewness = 1.1547005383792517
kurtosis = 2.0
```





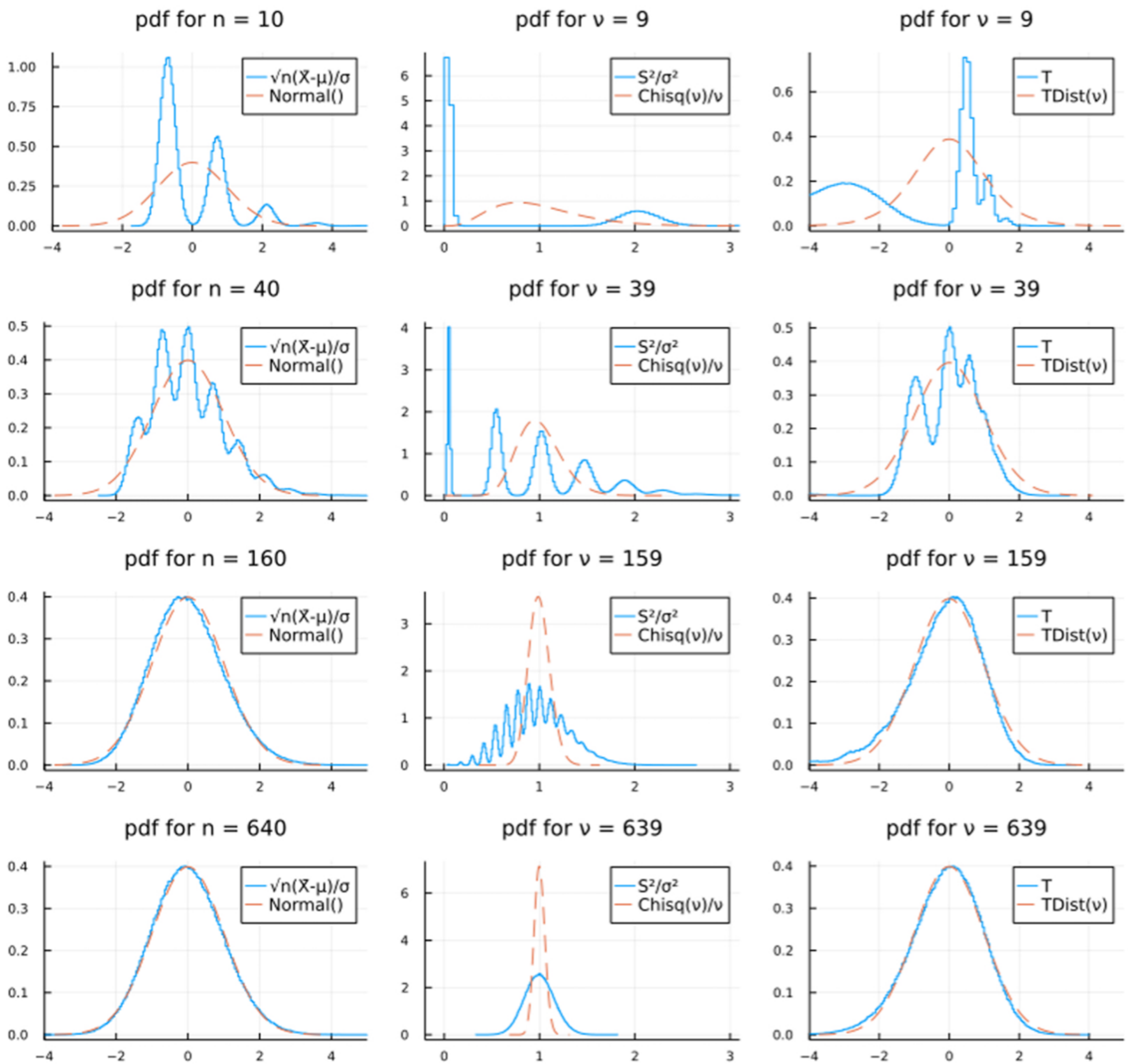
どのように一致していないか
に注目せよ、

```
μ = 1.0
σ² = 20.0
skewness = 3.823676241524649
kurtosis = 13.58499999999097
```



5%の小さな山のせいで

「面白いこと」になっている!



大数の法則と中心極限定理

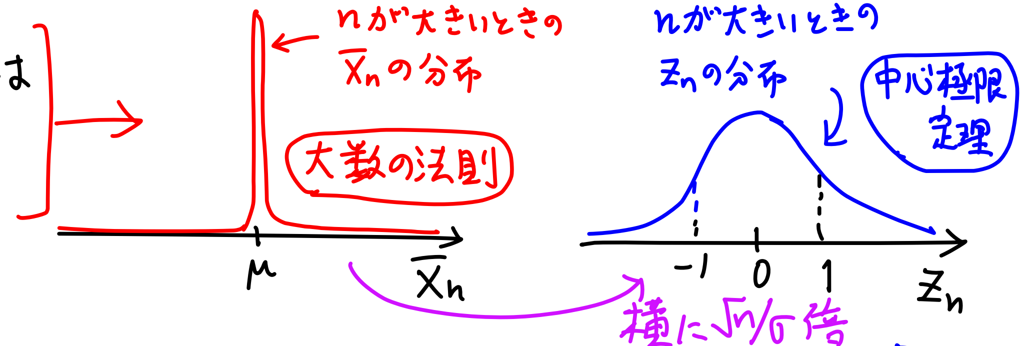
$\dfrac{X_1+\cdots+X_n}{n}$ や $\dfrac{X_1+\cdots+X_n}{\sqrt{n}}$ や $X_1+\cdots+X_n$ がみたす法則

まとめ $X_1, X_2, X_3, \ldots$ は独立同分布確率変数列であるとする。

大数の法則 各 X_i がしたがう分布の期待値がうまく定義されていると仮定する。

(すなわち, $E[|X_i|]<\infty$ が成立していると仮定する.) X_i たち共通の期待値を $\mu=E[X_i]$ と書く.

このとき, 確率変数 $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ の分布は n を大きくすると μ に集中して行く.



中心極限定理 (正規分布による近似定理)

$E[|X_i|]<\infty$, $\mu=E[X_i]$, $\sigma^2=E[(X_i-\mu)^2]<\infty$ であると仮定する. このとき, n を大きくすると,

(1) $Z_n = \dfrac{1}{\sqrt{n}}\sum_{i=1}^{n}\dfrac{X_i-\mu}{\sigma}$ の分布は標準正規分布 Normal$(0,1)$ で近似される.

(2) $X_1+\cdots+X_n$ の分布は正規分布 Normal$(n\mu, \sqrt{n}\,\sigma)$ で近似される.

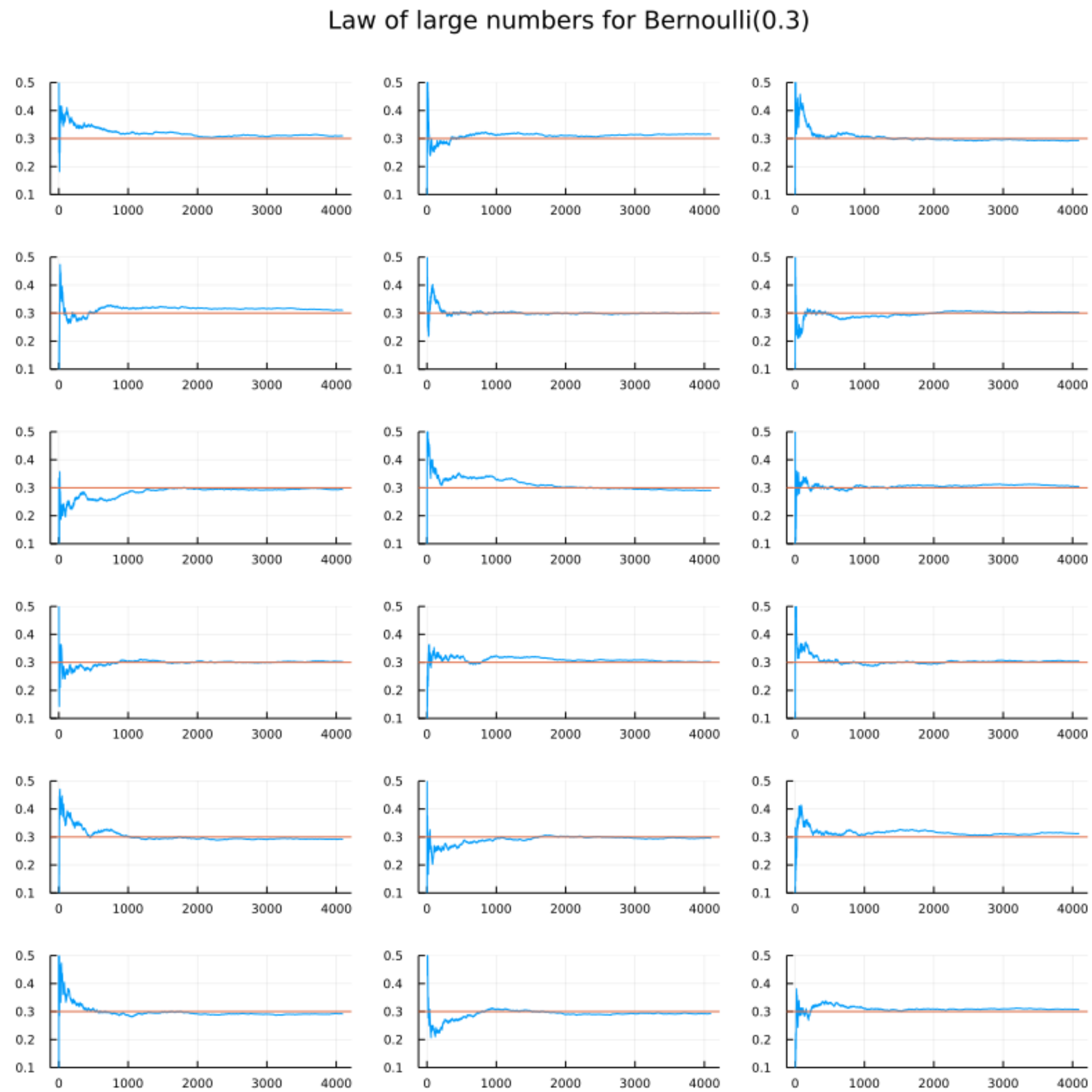
(3) $\overline{X}_n = \dfrac{X_1+\cdots+X_n}{n}$ の分布は正規分布 Normal$\left(\mu, \underbrace{\dfrac{\sigma}{\sqrt{n}}}\right)$ で近似される.　$n\to\infty$ で 0 に近付く

(4) $\sigma Z_n = \dfrac{(X_1-\mu)+\cdots+(X_n-\mu)}{\sqrt{n}}$ の分布は正規分布 Normal$(0,\sigma)$ で近似される.

n によらない

(1)〜(4)は同値.

大数の法則のデモンストレーション



Bernoulli試行の場合

確率0.3で1に

確率0.7で0になる

独立試行の結果

$$X_1, X_2, \dots, X_N \quad (N=2^{12})$$

を18個ランダムに生成し,

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

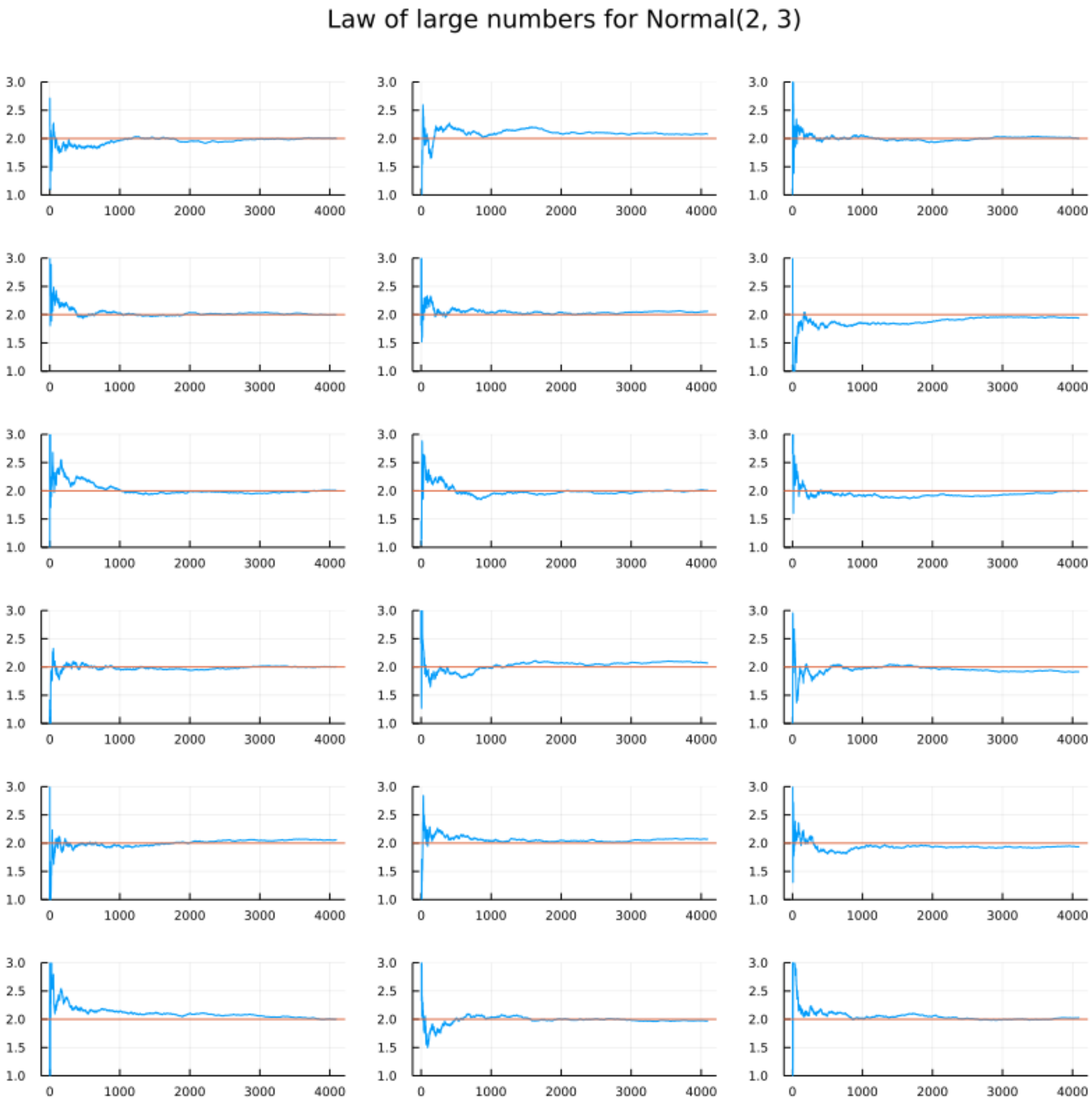
のグラフを18個描いた.

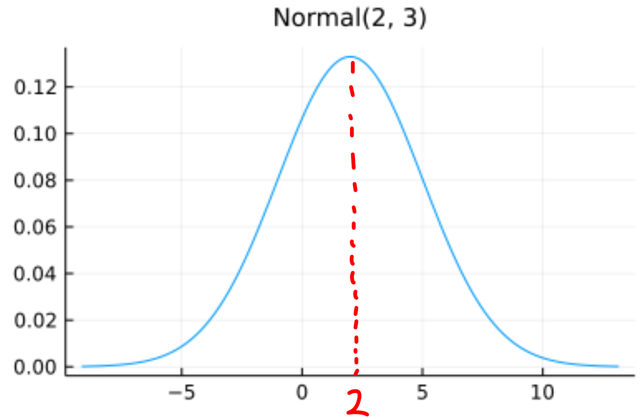
ヨコ軸がnで, タテ軸が$\bar{X}_n$,

nを大きくすると,

$\bar{X}_n$は0.3に近付く.

正規分布の大数の法則





上の正規分布の乱数

$X_1, X_2, \ldots, X_N$

を $N = 2^{12}$ 個生成することを

18回行った。

各々について、

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

を $n = 1, 2, \ldots, N$ について

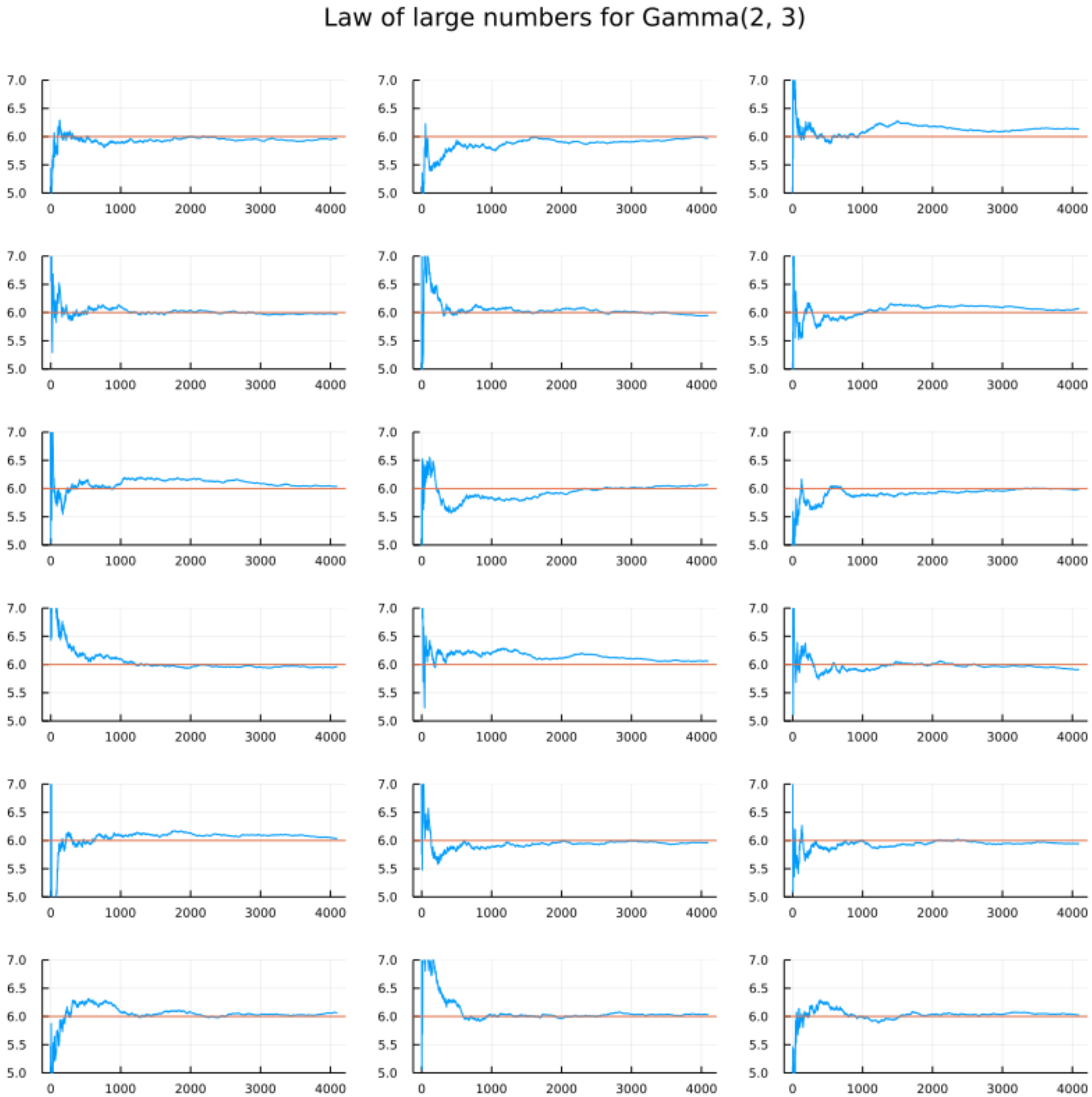
計算してプロット。

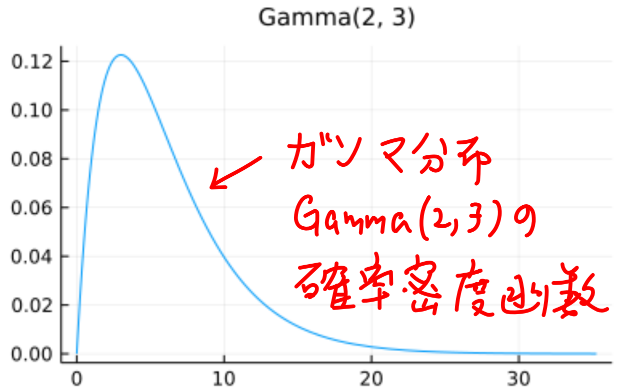
$\overline{X}_n$ は $n \to$ 大で

分布 Normal(2,3) の平均値

の 2 に収束する。

ガンマ分布の大数の法則





一般に $\mathrm{Gamma}(\alpha, \theta)$ の

平均 μ と分散 σ^2 は

$\mu = \alpha\theta$, $\sigma^2 = \alpha\theta^2$.

ガンマ分布の乱数列

$X_1, X_2, \dots, X_N \quad (N = 2^{12})$

を18回生成して,

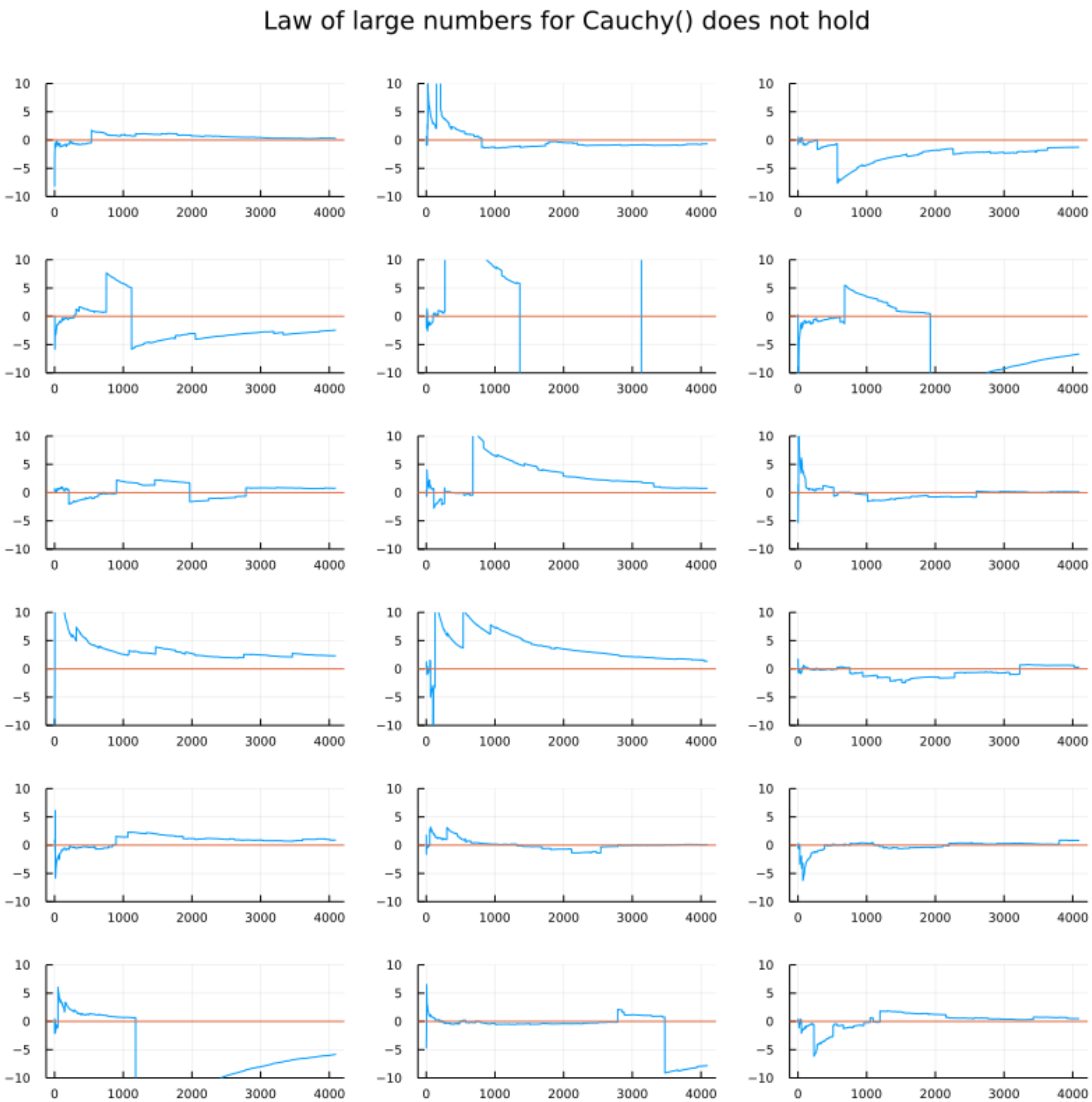
その各々について,

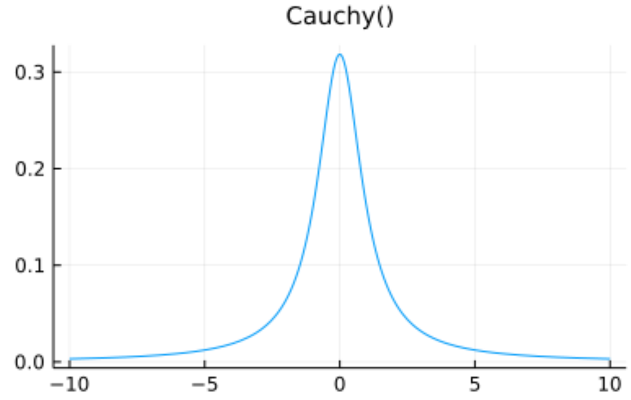
$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

を計算してプロット.

$\bar{X}_n$ は $\underset{2\times 3}{\overset{}{\underset{\|}{6}}}$ に収束する.

大数の法則が成立しない場合の例（Cauchy分布）





$$p(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

$X \sim$ Cauchy() のとき,

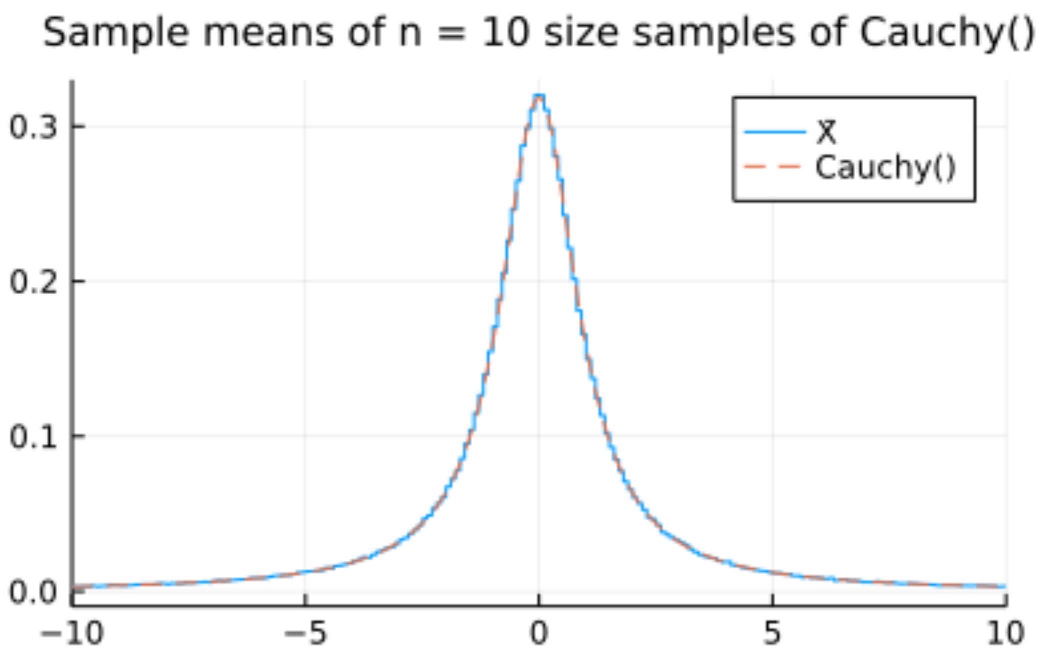
$$E[|X|] = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{|x|}{1+x^2}dx = \infty$$

X の期待値 $E[X]$ はうまく定義されない.

Cauchy分布にしたがう乱数では極端に絶対値が大きな値が出易いせいで,大数の法則は成立しない.

← $n \to$ 大で 0 に近付くように見える場合もあるがたまに出る極端な値がそれを台なしにする.

Cauchy分布の再生性



$X_1, X_2, \ldots, X_n$ は

それぞれが Cauchy分布

$$p(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

にしたがう独立同分布な

確率変数(乱数)たちであるとする

このとき,

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

も同じCauchy分布にしたがう.

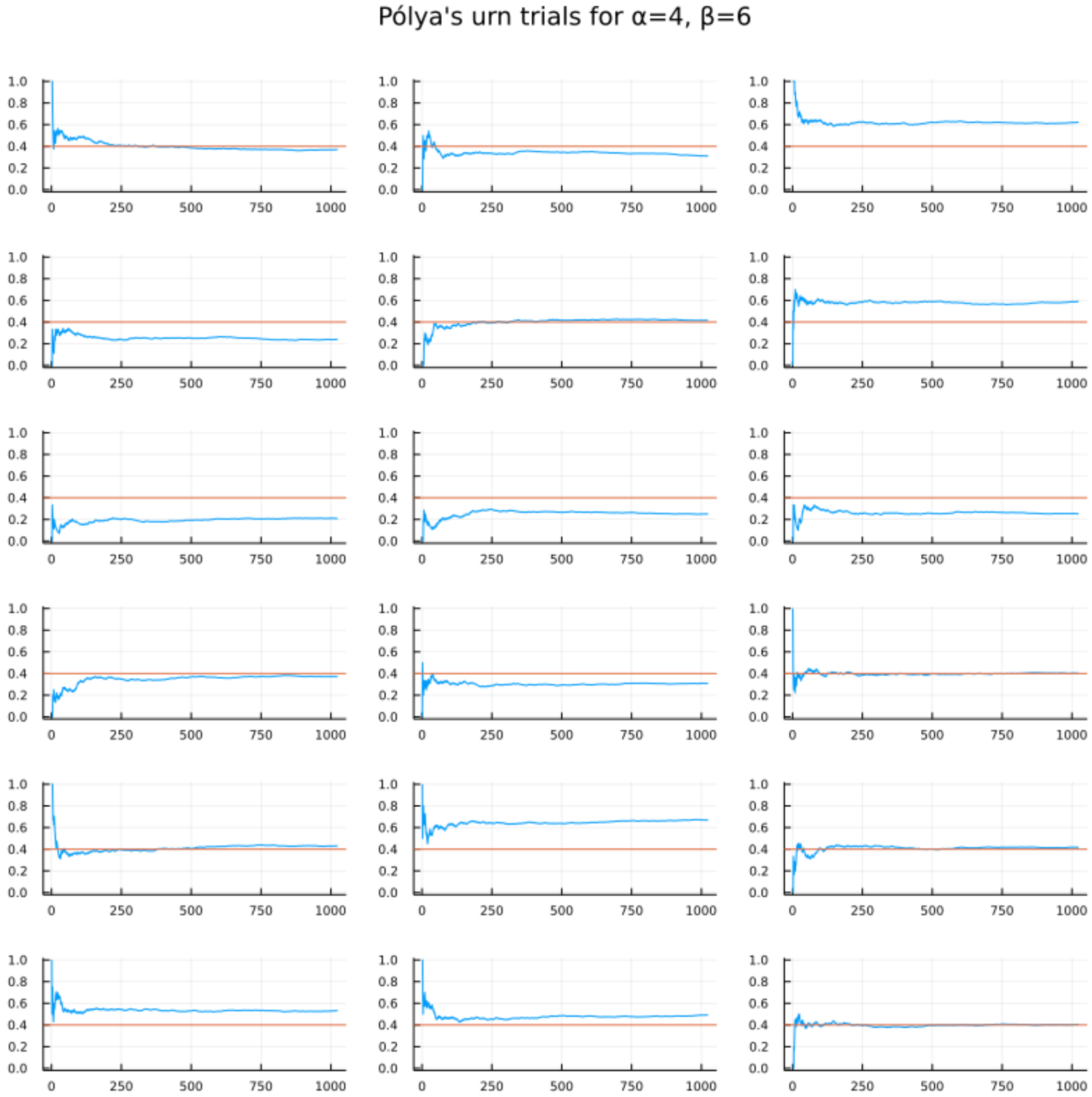
青の線は $n=10$ 場合の $\bar{X}_n$ の分布

---の線は Cauchy分布

ぴったり一致している.

独立性が成立していないせいで大数の法則が成立しない例 Pólyaのつぼ

つぼの中には最初赤い玉がα個, 白い玉がβ個入っている.



つぼからランダムに玉をとりだし,
赤なら赤い玉をつぼに2個もどし,
白なら白い玉をつぼに2個もどす,
つぼの中の玉の個数は
毎回1個ずつ増えて行く.

$$\begin{cases} \text{赤} \Rightarrow X_i = 1 \\ \text{白} \Rightarrow X_i = 0 \end{cases}$$

← $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ のグラフ

$\bar{X}_n$ は毎回ちがう値に収束する.

Polyaのつぼの試行の確率分布の記述

__例__ (赤白白赤白の順で出る確率) $= \dfrac{\alpha}{\alpha+\beta}\dfrac{\beta}{\alpha+\beta+1}\dfrac{\beta+1}{\alpha+\beta+2}\dfrac{\alpha+1}{\alpha+\beta+3}\dfrac{\beta+2}{\alpha+\beta+4}$.

　1 0 0 1 0

一般に $(x_1, x_2, \ldots, x_n)$ の中の k 個が赤 $=1$ のとき,

$$((x_1, x_2, \ldots, x_n)\text{の順で出る確率}) = \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)\cdot\beta(\beta+1)\cdots(\beta+n-k-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+n-1)}$$

別の表示 (GitHubにある「確率分布の解釈」のノートを参照せよ.)

$$\left(\begin{aligned}\Gamma(\alpha+k) &= (\alpha+k-1)\Gamma(\alpha+k-1) = (\alpha+k-1)(\alpha+k-2)\Gamma(\alpha+k-2)\\ &= \cdots = (\alpha+k-1)\cdots(\alpha+1)\alpha\Gamma(\alpha)\end{aligned}\right) \quad \therefore \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)} = \alpha(\alpha+1)\cdots(\alpha+k-1)$$

$$((x_1, x_2, \ldots, x_n)\text{の順で出る確率}) = \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\frac{\Gamma(\beta+n-k)}{\Gamma(\beta)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)} = \frac{B(\alpha+k, \beta+n-k)}{B(\alpha, \beta)}$$

$$= \int_0^1 p^k(1-p)^{n-k}\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\,dp = \int_0^1 \underbrace{\prod_{i=1}^n\left(p^{x_i}(1-p)^{1-x_i}\right)}_{\text{Bernoulli試行}}\underbrace{\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}}_{\text{ベータ分布}}\,dp$$

大数の法則の証明

(大数の弱法則のみを示す。)

$X_1, X_2, \ldots$ は独立同分布確率変数列であると仮定する。

$E[|X_i|] < \infty$, $\mu = E[X_i]$, $\sigma^2 = E[(X_i - \mu)^2] < \infty$ の場合の大数の弱法則のみを示す。

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ とおく。このとき、

$$E[\overline{X}_n] = \mu, \quad Var(\overline{X}_n) = \frac{\sigma^2}{n}.$$

$n \to \infty$ で 0 に収束する。

これを後で使う。

$\boxed{\text{証明}}$ E[ ] の線形性と規格化条件より、

$$E[\overline{X}_n] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu.$$

$$\overline{X}_n - \mu = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)$$

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

$$Var(\overline{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{E[(X_i - \mu)^2]}_{=\sigma^2} + \frac{1}{n^2}\sum_{i \neq j}\underbrace{E[(X_i - \mu)(X_j - \mu)]}_{\substack{= E[X_i - \mu]E[X_j - \mu] \\ = 0 \times 0}}$$

$$= \frac{\sigma^2}{n}$$

$\boxed{\text{q.e.d.}}$

Chebyshevの不等式 平均 μ, 分散 σ^2 を持つ確率変数 X と $\varepsilon > 0$ について,

$$P(|X-\mu| \geqq \varepsilon) \leqq \frac{\sigma^2}{\varepsilon^2}.$$

← X が μ から離れる確率は σ^2 を使って上からおさえられる.

証明 $|X-\mu| \geqq \varepsilon$ という条件は $1 \leqq \frac{(X-\mu)^2}{\varepsilon^2}$ と同値である. ゆえに,

$$\begin{aligned}
P(|X-\mu| \geqq \varepsilon) &= P\left(1 \leqq \frac{(X-\mu)^2}{\varepsilon^2}\right) \\
&= E\left[\underbrace{1 \leqq \frac{(X-\mu)^2}{\varepsilon^2} \text{ のとき } 1,\ \text{それ以外のとき } 0}_{\leqq \frac{(X-\mu)^2}{\varepsilon^2}}\right] \\
&\leqq E\left[\frac{(X-\mu)^2}{\varepsilon^2}\right] \\
&= \frac{E[(X-\mu)^2]}{\varepsilon^2} \\
&= \frac{\sigma^2}{\varepsilon^2}.
\end{aligned}$$

確率は期待値で表わされる.

期待値の方が計算しやすい.

E[ ]の単調性

E[ ]の線形性

(確率変数 X とは $f(X)$ の期待値 $E[f(X)]$ が定義されているものである.)

q.e.d.

$\boxed{\text{大数の弱法則}}$ $X_1, X_2, \ldots$ は独立同分布確率変数列であるとし,

$$E[|X_i|] < \infty,\quad \mu = E[X_i],\quad \sigma^2 = E[(X_i - \mu)^2] < \infty,\quad \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

と仮定する. このとき, 任意の $\varepsilon > 0$ に対して,

$$P(|\bar{X}_n - \mu| \geqq \varepsilon) \leqq \frac{\sigma^2}{\varepsilon^2}\frac{1}{n} \longrightarrow 0 \quad (\text{as } n \to \infty).$$

$n \to \infty$ で $\bar{X}_n$ の分布が μ に集中するという意味

すなわち,

$$1 \geqq P(|\bar{X}_n - \mu| < \varepsilon) = 1 - P(|\bar{X}_n - \mu| \geqq \varepsilon) \geqq 1 - \frac{\sigma^2}{\varepsilon^2}\frac{1}{n} \to 1 \quad (\text{as } n \to \infty).$$

$\boxed{\text{証明}}$ $E[\bar{X}_n] = \mu$, $\mathrm{var}(\bar{X}_n) = \dfrac{\sigma^2}{n}$ と Chebyshev の不等式より,

$$P(|\bar{X}_n - \mu| \geqq \varepsilon) \leqq \frac{\mathrm{var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2}\frac{1}{n} \longrightarrow 0 \quad (\text{as } n \to \infty).$$

$\boxed{\text{q.e.d.}}$

証明は一瞬で完了する!

$\boxed{\text{注意}}$ 大数の強法則 $P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$ を示すには $\sum_{n=1}^{\infty} P(|\bar{X}_n - \mu| \geqq \varepsilon) < \infty$ を示せば十分だが, $\sum_{n=1}^{\infty} \frac{1}{n} = \infty$ なので弱法則から強法則は出て来ない. □

自習問題

問題: 大数の法則に関する誤解

当たりとはずれが毎回独立ににどちらも 50% の確率で出るルーレットを回し続ける状況を考える. 当たりが出たら, 胴元から1万円もらい, はずれが出たら1万円支払うというルールになっているとする. この賭け事に参加しているある人がこう言った:

> はずれの方が沢山出続けてたくさん負けてしまった! でも, 確率論の大数の法則によって, これからは当たりの方が出易くなって, 負けは帳消しになるだろう.

この発言が間違っていることについて説明せよ.

不偏分散の分布も $n\to\infty$ で σ^2 に集中する。 (おまけ)

不偏分散の $n\to\infty$ での挙動

$X_1, X_2, X_3, \dots$ は独立同分布確率変数の列であるとし, その最初の n 個の標本平均と不偏分散をそれぞれ次のように書くことにする:

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

以下では $\mu = E[X_i]$ がうまく定義されており, $\sigma^2 = E[(X_i-\mu)^2] < \infty$, $E[|X_i-\mu|^3] < \infty$, $E[(X_i-\mu)^4] < \infty$ となっていると仮定する.

「標本分布について」のノートで以下を示していたのであった. X_i 達が共通に従う分布の歪度(わいど, skewness)と尖度(せんど, kurtosis)をそれぞれ

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

と書くと,

$$\begin{aligned} &E[\bar{X}_n] = \mu, \quad E[S_n^2] = \sigma^2, \\ &\operatorname{var}(\bar{X}_n) = \frac{\sigma^2}{n}, \quad \operatorname{cov}(\bar{X}_n, S_n^2) = \sigma^3\frac{\bar{\kappa}_3}{n}, \quad \operatorname{var}(S_n^2) = \sigma^4\left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}\right). \end{aligned}$$

ゆえに, Chebyshevの不等式を X, μ, σ^2 がそれぞれ S_n^2, $E[S_n^2] = \sigma^2$, $\operatorname{var}(S_n^2) = \sigma^4(\bar{\kappa}_4/n + 2/(n-1))$ の場合に適用すると, 任意の $\varepsilon > 0$ について,

尖度(せんど)

$$\boxed{P(|S_n^2 - \sigma^2| \geq \varepsilon) \leq \frac{\sigma^4}{\varepsilon^2}\left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}\right) \to 0 \quad \text{as } n\to\infty.}$$

このように, 不偏分散 S_n^2 についても大数の弱法則が成立しており, S_n^2 の分布は $n\to\infty$ で σ^2 の近くに集中して行く.

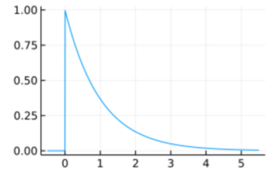
以下は前回の資料より

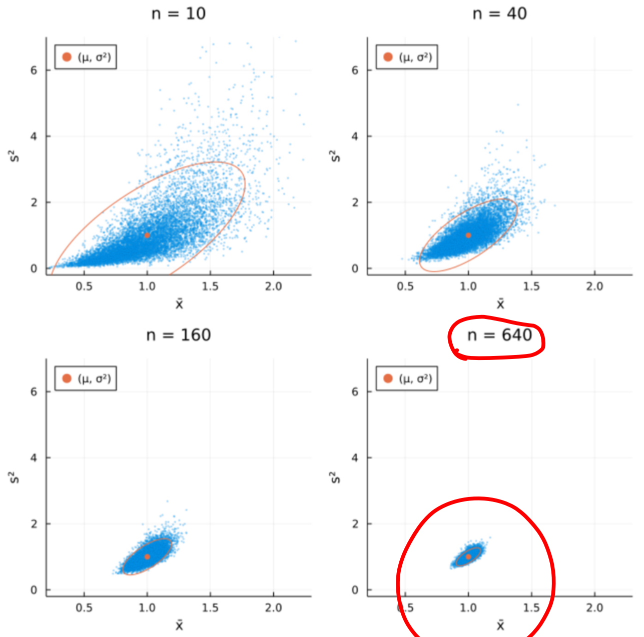
$$(\bar{X}_n, S_n^2) = \left(\frac{1}{n}\sum_{i=1}^n X_i,\ \frac{1}{n-1}\sum_{i=1}^n (X_i-\bar{X}_n)^2\right) \text{ の分布}$$

指数分布の標本分布における $\bar{X}, S^2$ の同時分布

```
plot_X̄_and_SX²_2x2(Exponential(); xlim=(0.2, 2.3), ylim=(-0.2, 7))
```

```
Exponential{Float64}(θ=1.0)
μ = 1.0
σ² = 1.0
skewness = 2.0
kurtosis = 6.0
```

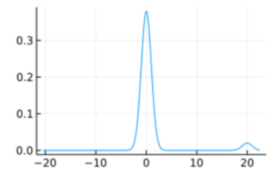


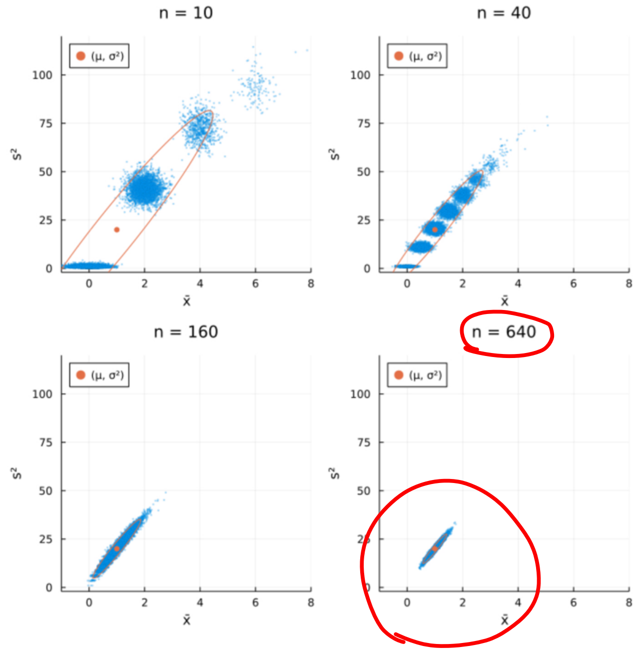


非対称な2つ山の混合正規分布の標本分布における $\bar{X}, S^2$ の同時分布

```
plot_X̄_and_SX²_2x2(MixtureModel([Normal(), Normal(20)], [0.95, 0.05]);
    xlim=(-1, 8), ylim=(-2, 120))
```

```
MixtureModel{Normal{Float64}}(K = 2)
components[1] (prior = 0.9500): Normal{Float64}(μ=0.0, σ=1.0)
components[2] (prior = 0.0500): Normal{Float64}(μ=20.0, σ=1.0)

μ = 1.0
σ² = 20.0
skewness = 3.823676241524649
kurtosis = 13.58499999999097
```

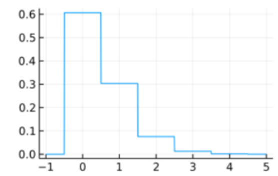


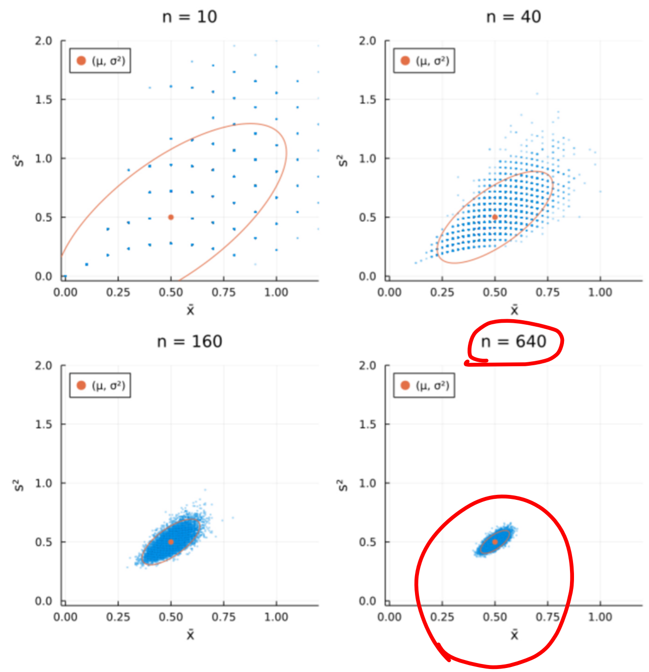


Poisson分布の標本分布における $\bar{X}, S^2$ の同時分布

```
plot_X̄_and_SX²_2x2(Poisson(0.5); xlim=(-0.02, 1.2), ylim=(-0.04, 2.0))
```

```
Poisson{Float64}(λ=0.5)
μ = 0.5
σ² = 0.5
skewness = 1.414213562373095
kurtosis = 2.0
```

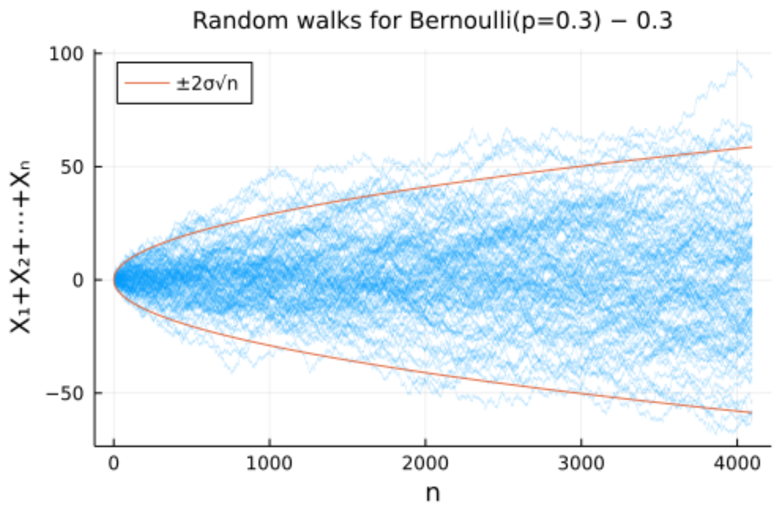


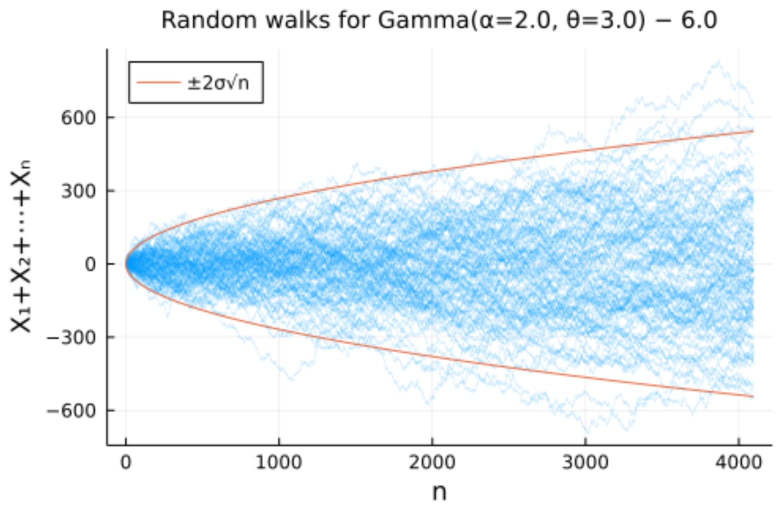


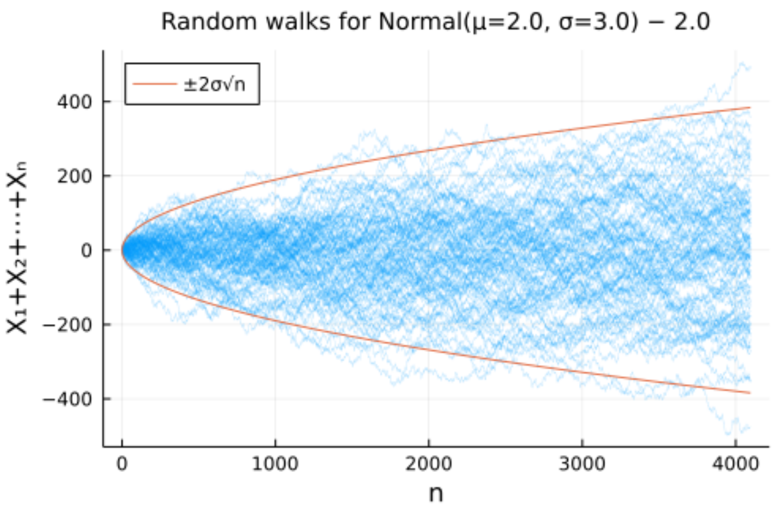
nを大きくすると $(\bar{X}_n, S_n^2)$ の分布は (μ, σ^2) に集中して行く。

$n \to$ 大で $(\bar{X}_n, S_n^2)$ のちらばり方の大きさは $\frac{1}{\sqrt{n}}$ に比例する。

$E[X_i]=0$ となる X_i たちの和 $X_1+\cdots+X_n$ の動き方を100個同時プロット







$\frac{\sqrt{n}}{n} \to 0$

$\Downarrow$

大数の法則

$\frac{\sqrt{n}}{\sqrt{n}} \to 1$

$\Downarrow$

中心極限定理

独立同分布確率変数列 $X_1, X_2, \ldots$ について,
$E[X_i]=0$, $\mathrm{var}(X_i)=\sigma^2<\infty$ が成立しているとき,

$$n \mapsto X_1+\cdots+X_n$$

のグラフを大量にプロットすると,
$X_1+\cdots+X_n$ は $\pm 2\sigma\sqrt{n}$ のあいだに
大部分が含まれる. $\longrightarrow$ 中心極限定理へ!

$\boxed{\text{中心極限定理}}$ $\dfrac{X_1+\cdots+X_n}{\sqrt{n}\,\sigma}$ の分布は標準正規分布に近付く!

中心極限定理

$X_1, X_2, \ldots$ は独立同分布確率変数列であるとし,

$$E[|X_i|] < \infty,\quad \mu = E[X_i],\quad \sigma^2 = E[(X_i-\mu)^2] < \infty,\quad E[|X_i-\mu|^3] < \infty$$

と仮定する。このとき,

注意 $E[Z_n]=0$, $\mathrm{Var}(Z_n) = E[Z_n^2] = \frac{n}{\sigma^2}E[(\bar{X}_n-\mu)^2] = 1.$

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i,\quad Z_n = \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{X_i-\mu}{\sigma},\qquad Z \sim \mathrm{Normal}(0,1)$$

とおくと,

$E\left[\frac{X_i-\mu}{\sigma}\right]=0$, $\mathrm{Var}\left(\frac{X_i-\mu}{\sigma}\right)=1$

$$\lim_{n\to\infty} E[f(Z_n)] = E[f(Z)] = \int_{-\infty}^{\infty} f(z)\,\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz.$$

n が大きなとき, Z_n の分布は標準正規分布で近似されるという意味

同値な別の表現

n が大きいとき,

(1) 近似的に $Z_n = \dfrac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} = \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^n \dfrac{X_i-\mu}{\sigma} \sim \mathrm{Normal}(0,1).$

(2) 近似的に $\sigma Z_n = \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^n (X_i-\mu) \sim \mathrm{Normal}(0,\sigma)$

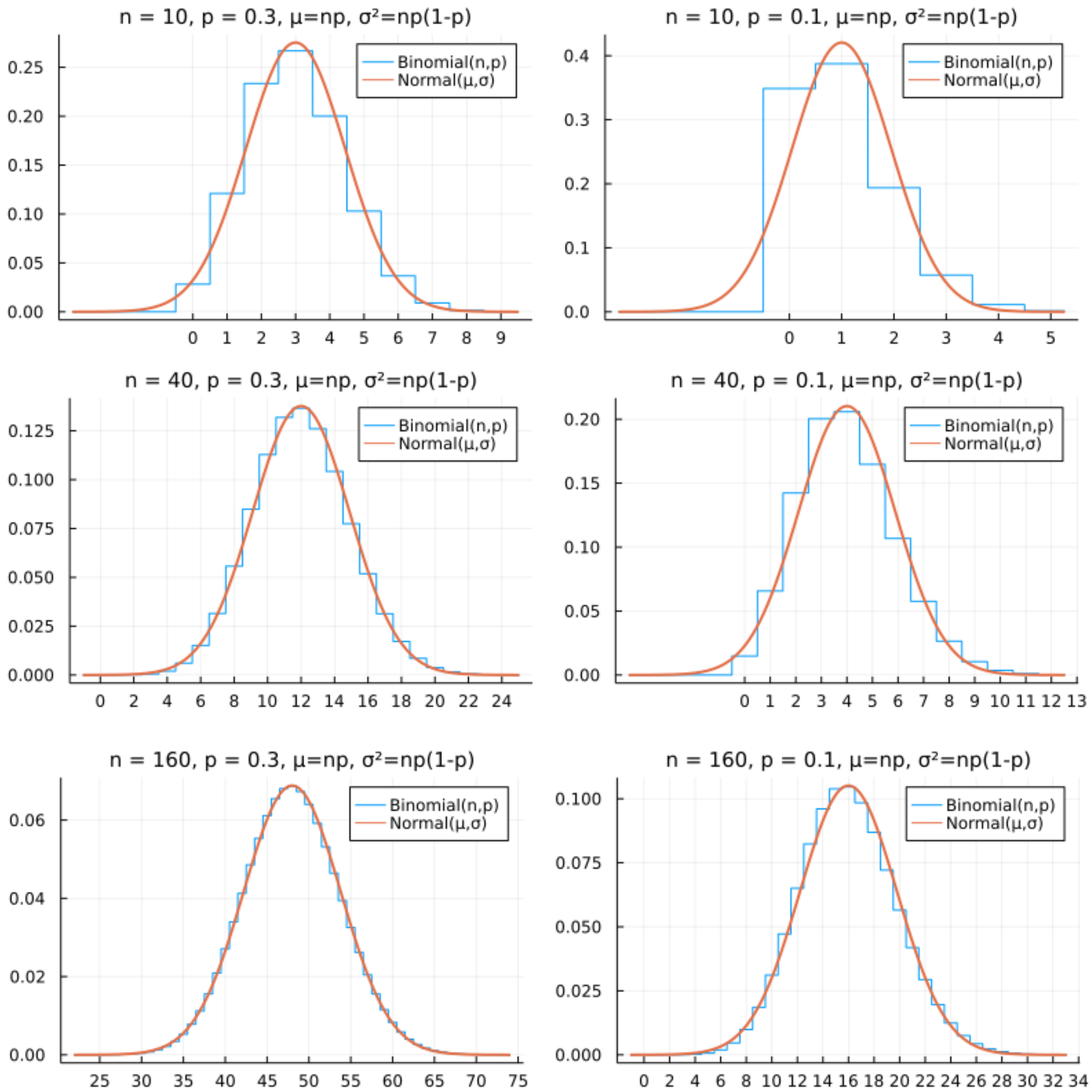
(3) 近似的に $\bar{X}_n = \mu + \dfrac{\sigma}{\sqrt{n}} Z_n \sim \mathrm{Normal}\left(\mu, \dfrac{\sigma}{\sqrt{n}}\right)$

(4) 近似的に $n\bar{X}_n = X_1+\cdots+X_n \sim \mathrm{Normal}(n\mu, \sqrt{n}\,\sigma).$

ゆるい仮定のもとで独立同分布な確率変数たちの和は近似的に正規分布にしたがう。

二項分布の中心極限定理

npとn(1-p)が大きいと精度が高くなる。



$X_1, X_2, \dots$ を各々が分布 Bernoulli(p) にしたがう独立同分布確率変数列とすると、

$$K_n = X_1 + \cdots + X_n \sim \mathrm{Binomial}(n, p),$$

$$E[K_n] = np =: \mu,$$

$$\mathrm{var}(K_n) = np(1-p) =: \sigma^2$$

nが大きなとき、近似的に

$$K_n \sim \mathrm{Normal}\left(np, \sqrt{np(1-p)}\right)$$

証明はGitHub資料にある。
かなりおもしろい。
Kullback-Leibler情報量に関係する。

必修自習問題

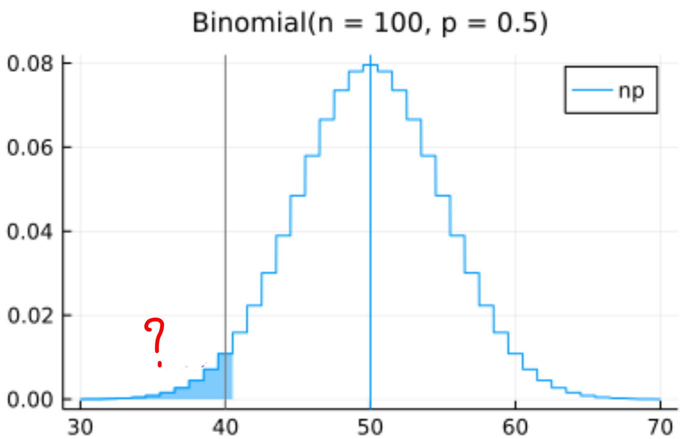
答えはGitHubにある.

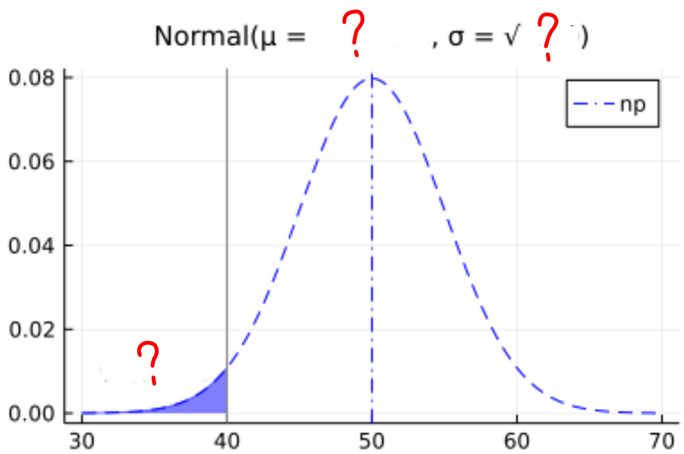
2.3 必修重要問題: 二項分布の正規分布近似を使った確率の近似計算1

確率変数 K は二項分布 Binomial(100, 1/2) に従っていると仮定する.

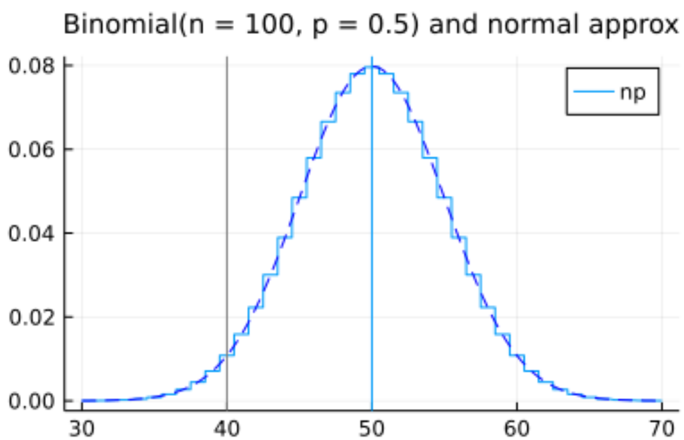
(1) 確率 $P(K \le 40)$ を小数点以下第5桁まで求めよ.

(2) 二項分布 Binomial(100, 1/2) と同じ期待値と分散を持つ正規分布に従う確率変数 X について確率 $P(X \le 40)$ を小数点以下第5桁まで求めよ.





左の図の4つの?を求める問題.



← このように二項分布の正規分布近似が非常にうまくいっている.

必修自習問題

2.4 必修重要問題: 二項分布の正規分布近似を使った確率の近似計算2

$0 < p < 1$ であると仮定し, K は二項分布 $\mathrm{Binomial}(100, p)$ に従う確率変数であるとする.

(1) $P(K \le 40) = 0.025$ を満たす p の値 p_U を小数点以下第4桁まで求めよ.

(2) $P(K \ge 40) = 0.025$ を満たす p の値 p_L を小数点以下第4桁まで求めよ.

$\hat{p} = 40/100 = 0.4$ とおく. X は K と同じ平均 np と推定された分散 $n\hat{p}(1-\hat{p})$ を持つ正規分布に従う確率変数であるとする.

(3) $P(X \le 40) = 0.025$ を満たす p の値 q_U を小数点以下第4桁まで求めよ.

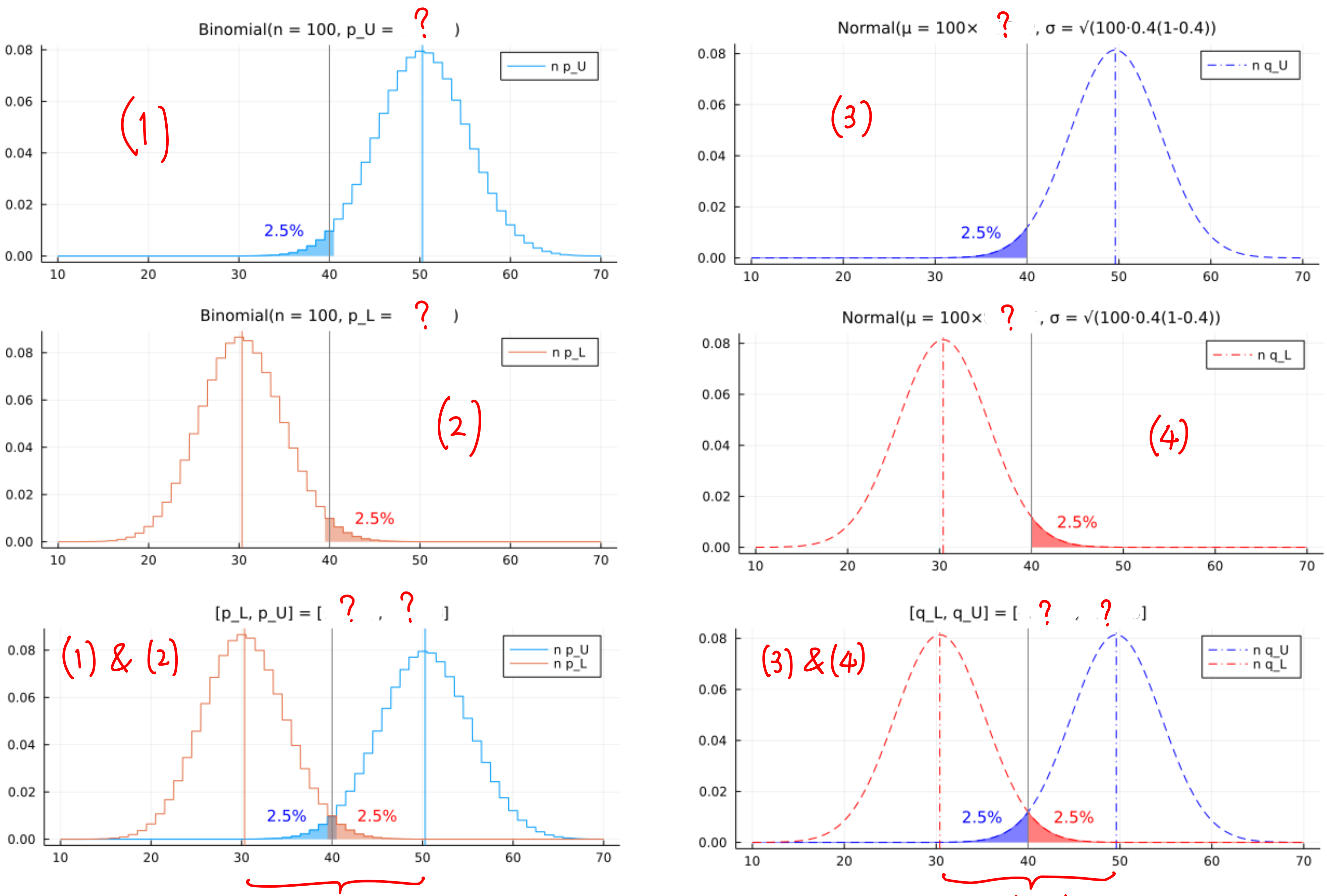
(4) $P(X \ge 40) = 0.025$ を満たす p の値 q_L を小数点以下第4桁まで求めよ.

注意: これらによって計算される区間 $[p_L, p_U]$ と $[q_L, q_U]$ はデータ「100 回中 40 回当たり」というデータから得られる p の 95% 信頼区間と呼ばれる. 前者は **Clopper-Pearsonの信頼区間** と, 後者は **Waldの信頼区間** と呼ばれる.

こちらの問題の方が前ページの問題よりずっとむずかしい.
単位をとるためのがんばりどころ!

次ページに図による問題の説明がある.

(1)と(2)について ↘　　　　　(3)と(4)について ↘



データ k = 40 に対する Clopper-Pearson の信頼区間 × 100　　　　Wald の信頼区間 × 100

離散分布を正規分布のような連続分布で近似するときの注意

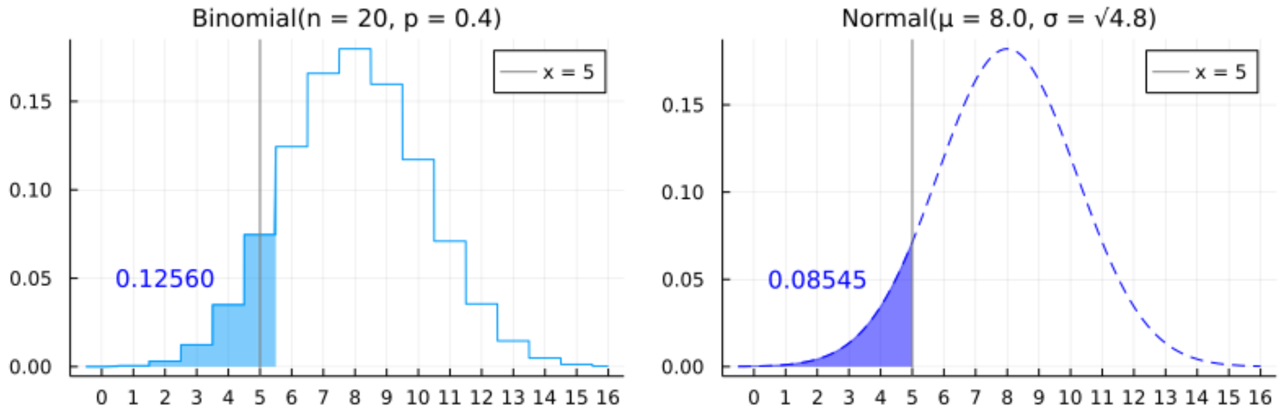
$n = 20,\ p = 0.4$

$\mu = np = 8,$

$\sigma^2 = np(1-p) = 20 \times 0.4 \times 0.6 = 4.8$

$K \sim \mathrm{Binomial}(n, p)$

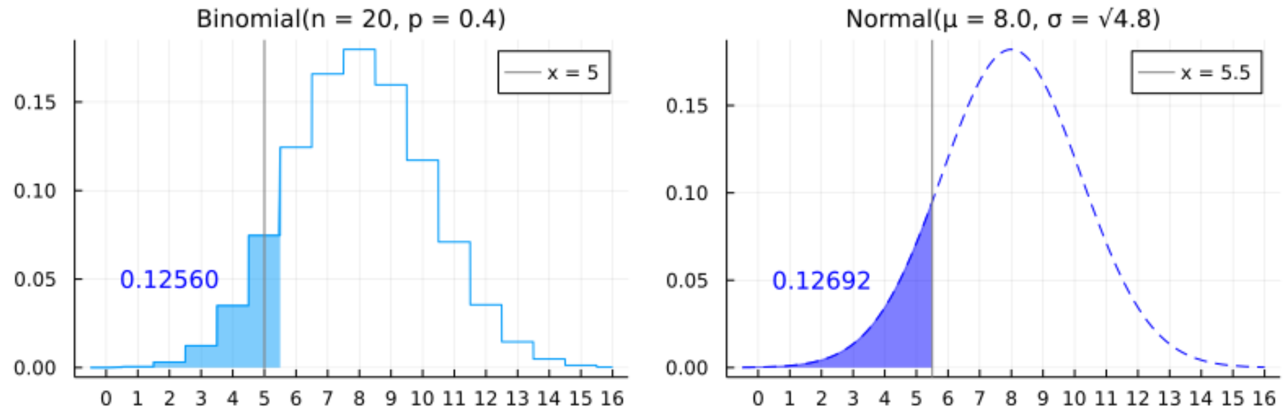
$X \sim \mathrm{Normal}(\mu, \sigma)$



$P(K \le 5) \approx 0.12560$ と

$P(X \le 5) \approx 0.08545$

の値が大きくちがう。

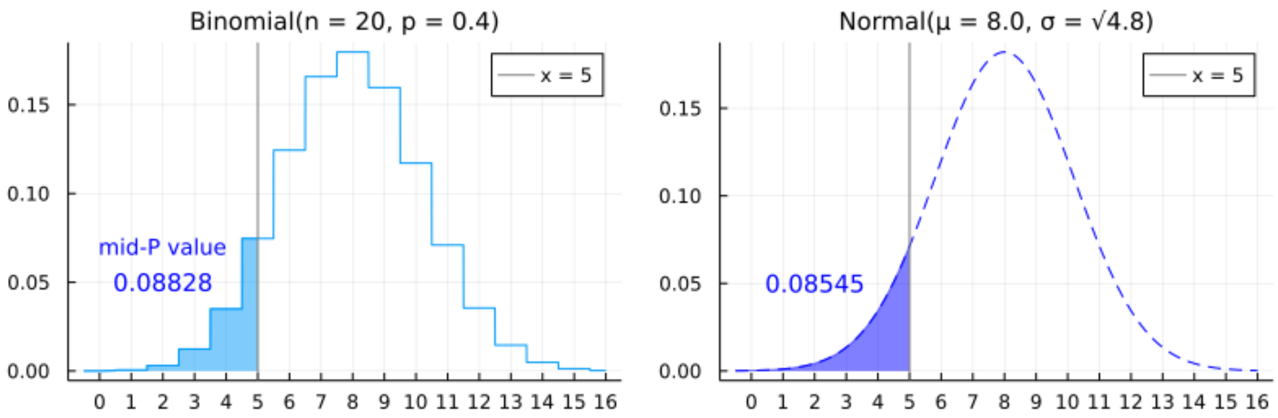


$P(K \le 5) \approx 0.12560$

$P(X \le 5.5) \approx 0.12692$

近い

連続性補正



$P(X \le 5) \approx 0.08545$

近い

$\dfrac{P(K \le 4) + P(K \le 5)}{2} \approx 0.08828$

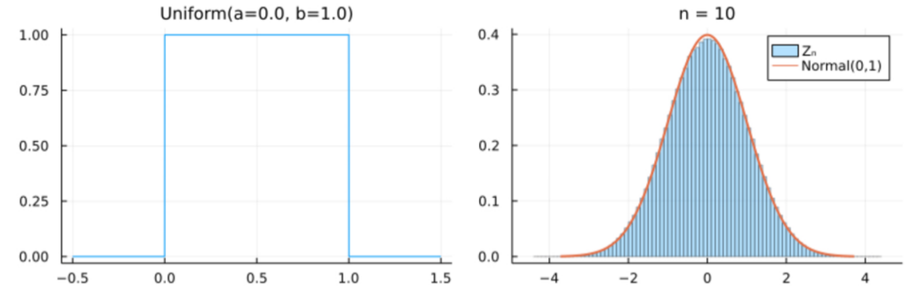
mid-P補正

中心極限定理のデモ

各X_iがしたがう分布によって、$Z_n = \dfrac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ が標準正規分布でよく近似されるようになるnの大きさはかなりちがう。

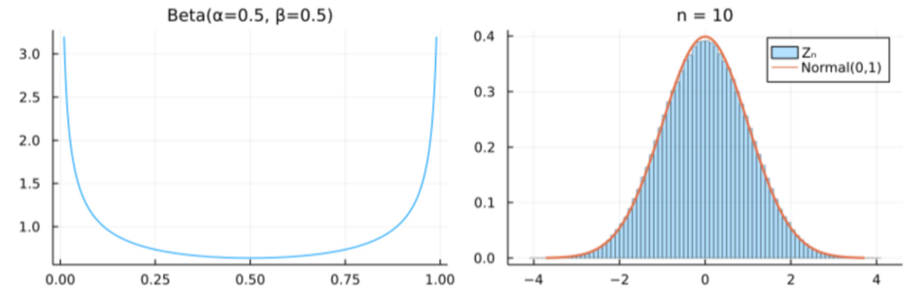
```
# 中心極限定理による収束が速い場合1の再現
plot_central_limit_theorem(Uniform(0, 1), 10; a=-0.5, b=1.5)
```

```
skewness(dist) = 0.0
kurtosis(dist) = -1.2
```



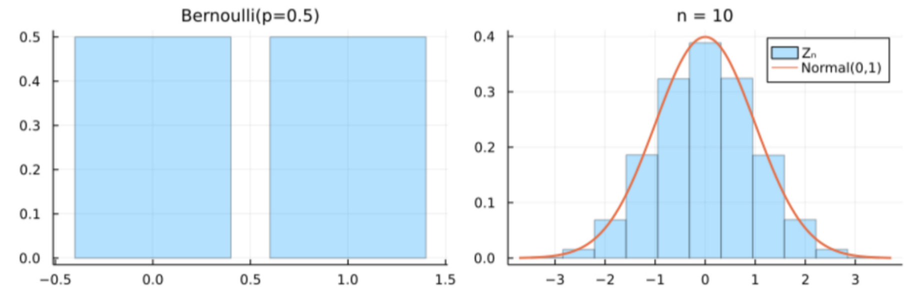

```
# 中心極限定理による収束が速い場合2
plot_central_limit_theorem(Beta(0.5, 0.5), 10; a=0.01, b=0.99)
```

```
skewness(dist) = 0.0
kurtosis(dist) = -1.5
```



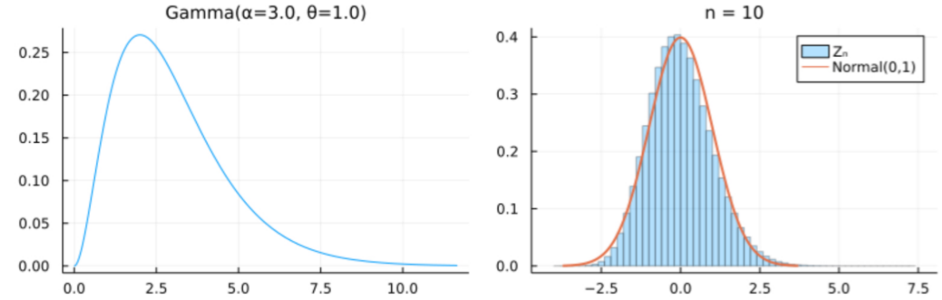

```
# 中心極限定理による収束が速い場合3
plot_central_limit_theorem(Bernoulli(0.5), 10)
```

```
skewness(dist) = 0.0
kurtosis(dist) = -2.0
```



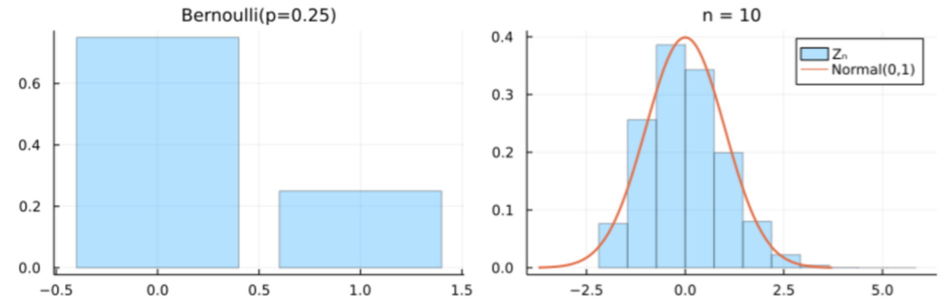

```
# 左右非対称な分布の場合1
plot_central_limit_theorem(Gamma(3, 1), 10)
```

```
skewness(dist) = 1.1547005383792517
kurtosis(dist) = 2.0
```



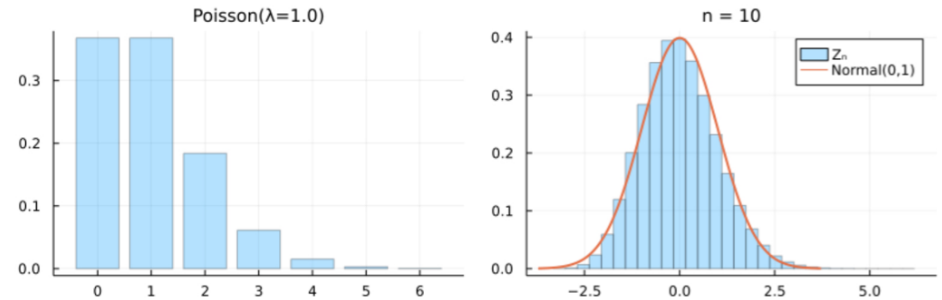

```
# 左右非対称な分布の場合2
plot_central_limit_theorem(Bernoulli(0.25), 10)
```

```
skewness(dist) = 1.1547005383792517
kurtosis(dist) = -0.6666666666666667
```




```
# 左右非対称な分布の場合3
plot_central_limit_theorem(Poisson(1), 10)
```

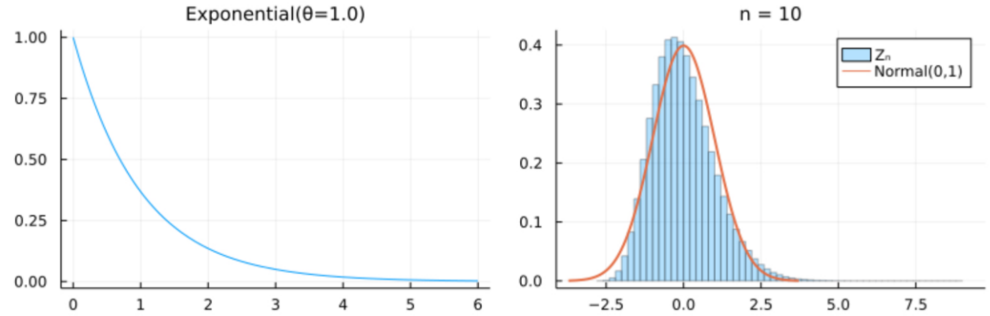
```
skewness(dist) = 1.0
kurtosis(dist) = 1.0
```



大きめのnが必要な場合

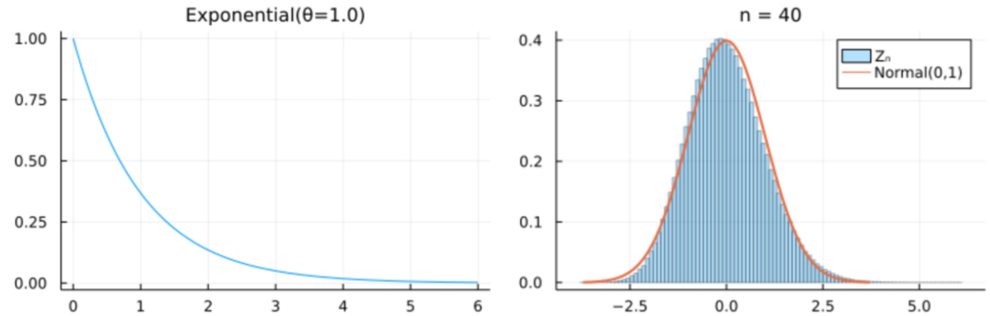
```
# 中心極限定理による収束が遅い場合1
# 左右の非対称性が大きな分布を試してみる
# 指数分布は左右の非対称性が大きな分布になっている
plot_central_limit_theorem(Exponential(), 10)
```

```
skewness(dist) = 2.0
kurtosis(dist) = 6.0
```



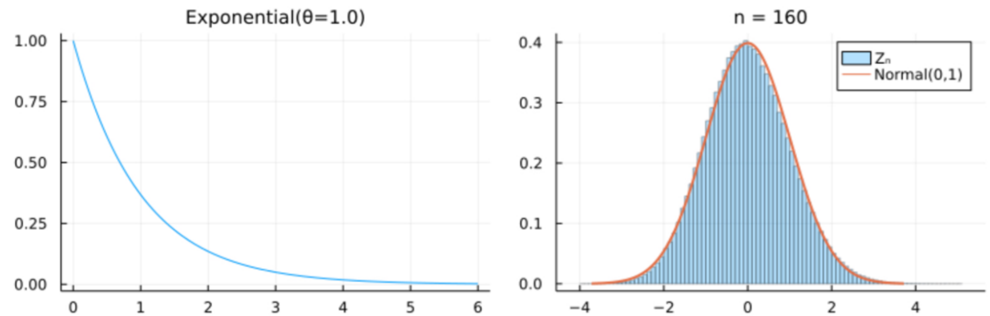

```
plot_central_limit_theorem(Exponential(), 40)
```

```
skewness(dist) = 2.0
kurtosis(dist) = 6.0
```



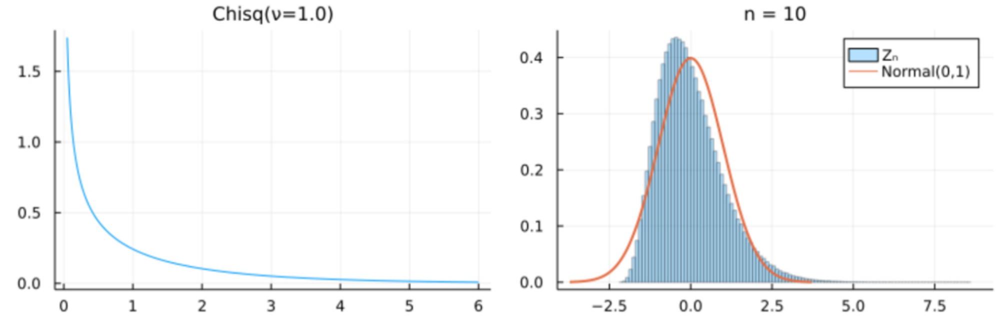

```
plot_central_limit_theorem(Exponential(), 160)
```

```
skewness(dist) = 2.0
kurtosis(dist) = 6.0
```



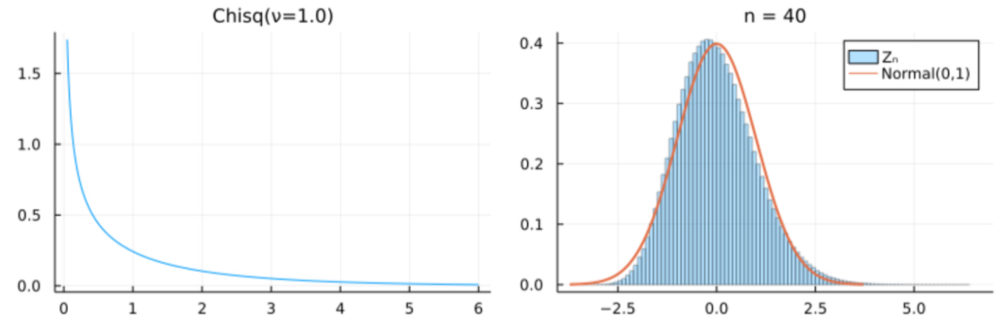

```
# 中心極限定理による収束が遅い場合2
# 自由度1のχ²分は左右の非対称性が指数分布よりも大きな分布になっている
plot_central_limit_theorem(Chisq(1), 10; a=0.05, b=6)
```

```
skewness(dist) = 2.8284271247461903
kurtosis(dist) = 12.0
```



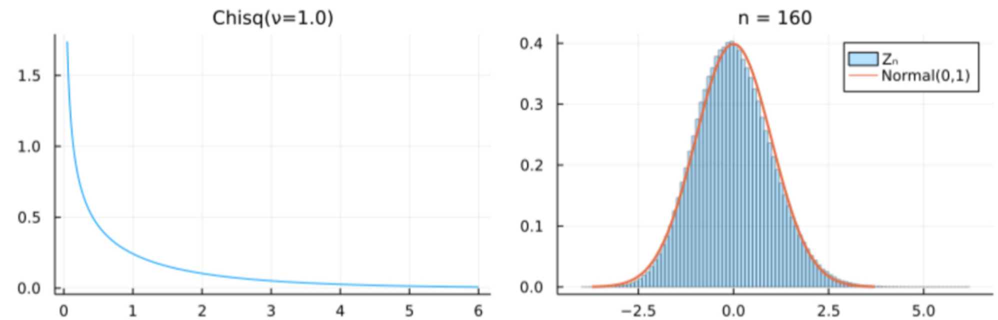

```
plot_central_limit_theorem(Chisq(1), 40; a=0.05, b=6)
```

```
skewness(dist) = 2.8284271247461903
kurtosis(dist) = 12.0
```




```
plot_central_limit_theorem(Chisq(1), 160; a=0.05, b=6)
```

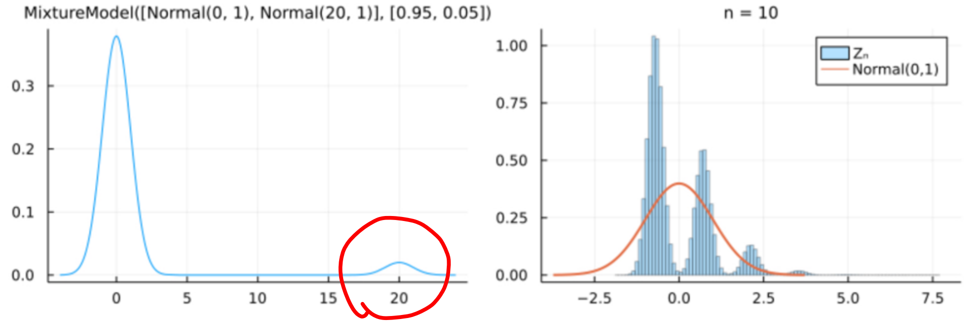
```
skewness(dist) = 2.8284271247461903
kurtosis(dist) = 12.0
```



大きめのnが必要な場合

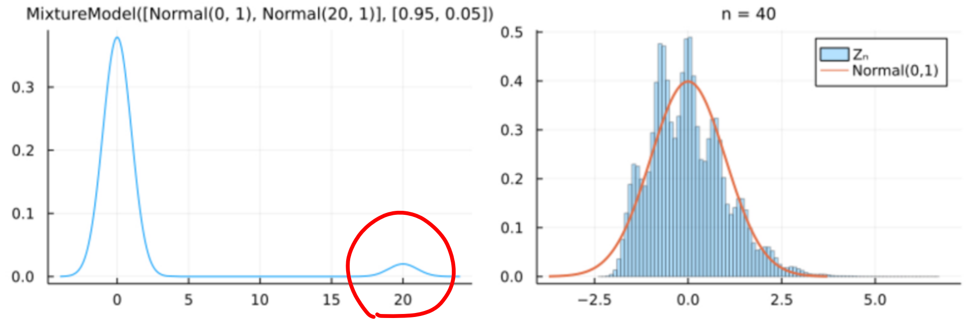
```
# 中心極限定理による収束が遅い場合3
# 以下のような分布distも左右の非対称性が大きな分布に分類される
dist = MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])
disttitle = "MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])"
titlefontsize = 9
a, b = -4, 24
plot_central_limit_theorem(dist, 10; a, b, disttitle, titlefontsize)
```

```
skewness(dist) = 3.8236762415246486
kurtosis(dist) = 13.584999999990963
```



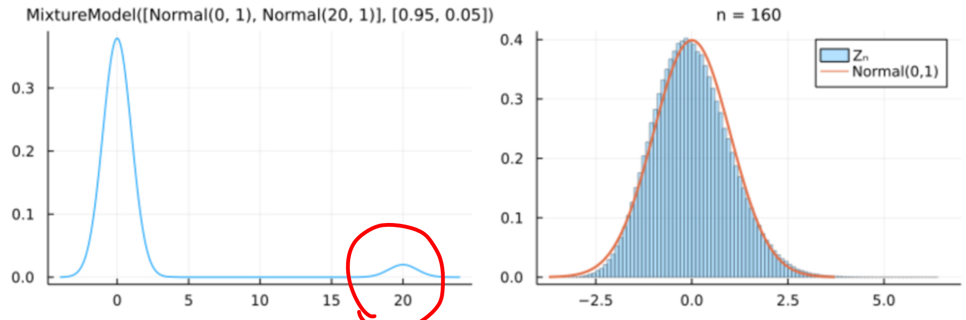

```
plot_central_limit_theorem(dist, 40; a, b, disttitle, titlefontsize)
```

```
skewness(dist) = 3.8236762415246486
kurtosis(dist) = 13.584999999990963
```



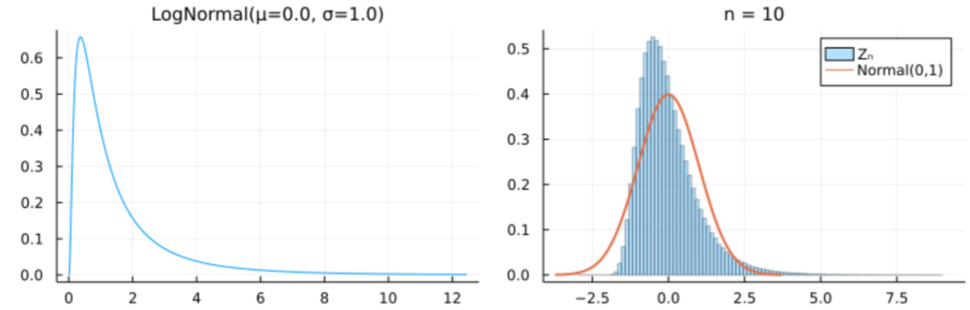

```
plot_central_limit_theorem(dist, 160; a, b, disttitle, titlefontsize)
```

```
skewness(dist) = 3.8236762415246486
kurtosis(dist) = 13.584999999990963
```



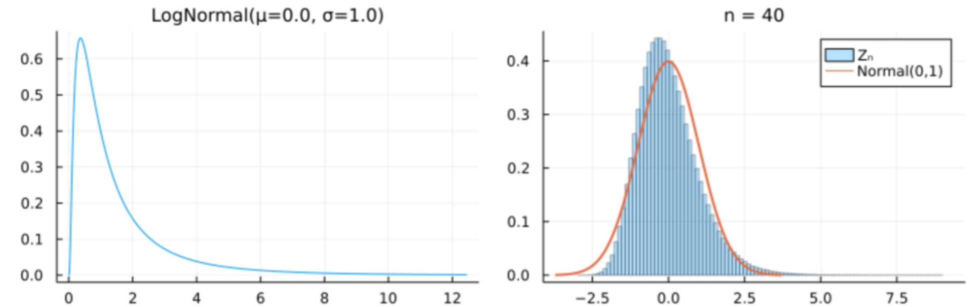

```
# 中心極限定理による収束が遅い場合4
# 対数正規分布は左右の非対称性が非常に大きな分布である
# 右側の裾が太く，外れ値が出やすい
plot_central_limit_theorem(LogNormal(), 10; bin=range(-3, 9, 100))
```

```
skewness(dist) = 6.184877138632554
kurtosis(dist) = 110.9363921763115
```



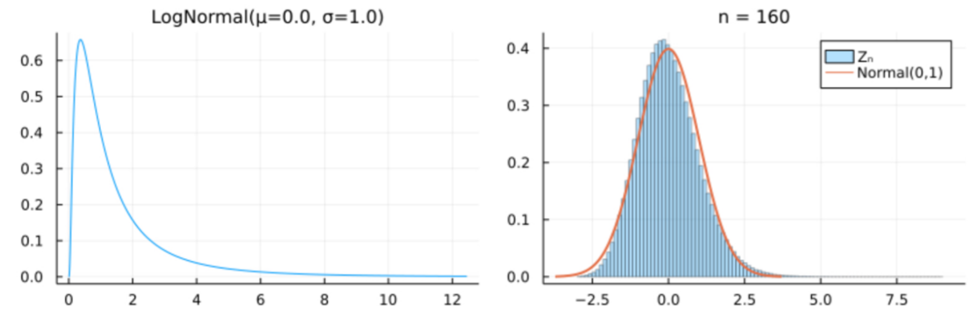

```
plot_central_limit_theorem(LogNormal(), 40; bin=range(-3, 9, 100))
```

```
skewness(dist) = 6.184877138632554
kurtosis(dist) = 110.9363921763115
```




```
plot_central_limit_theorem(LogNormal(), 160; bin=range(-3, 9, 100))
```

```
skewness(dist) = 6.184877138632554
kurtosis(dist) = 110.9363921763115
```

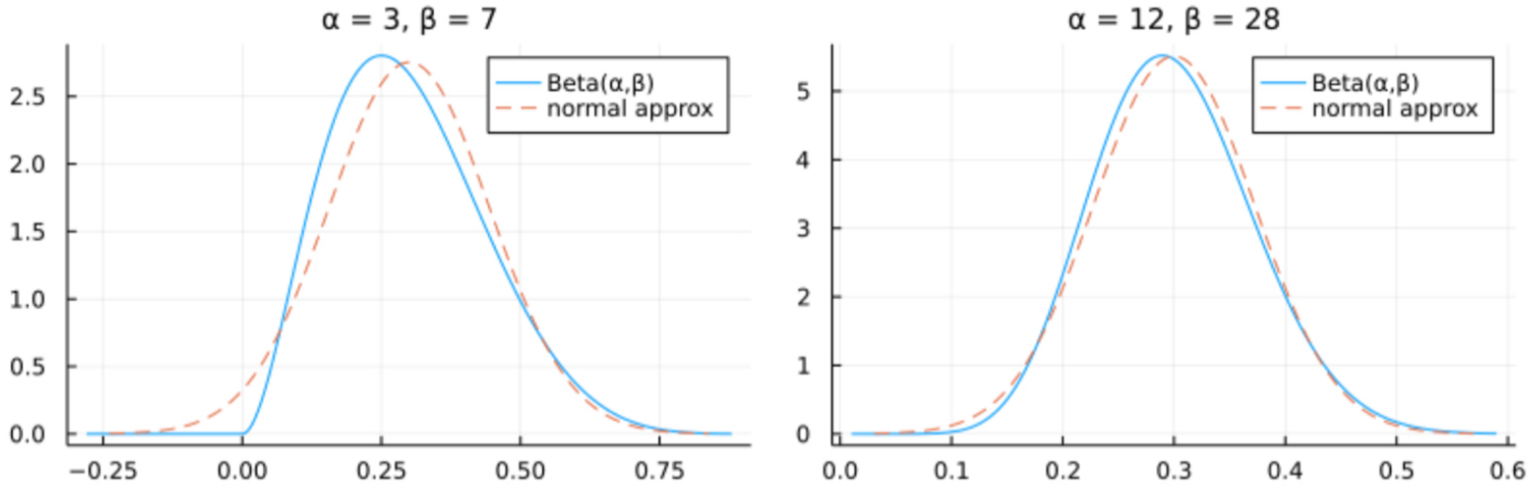


中心極限定理の直接的帰結でなくても「デルタ法」によって

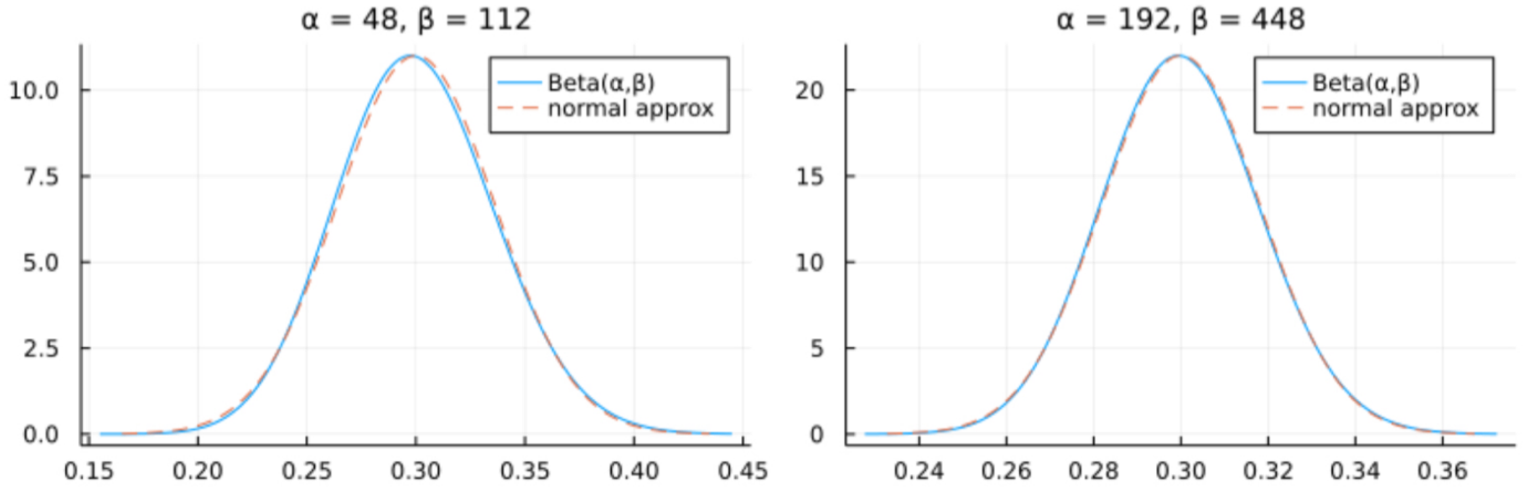
GitHub資料を見よ.

中心極限定理から間接的に正規分布近似できることがわかる場合がたくさんある.

```
plot(plot_betanormal(3, 7), plot_betanormal(12, 28); size=(800, 250))
```



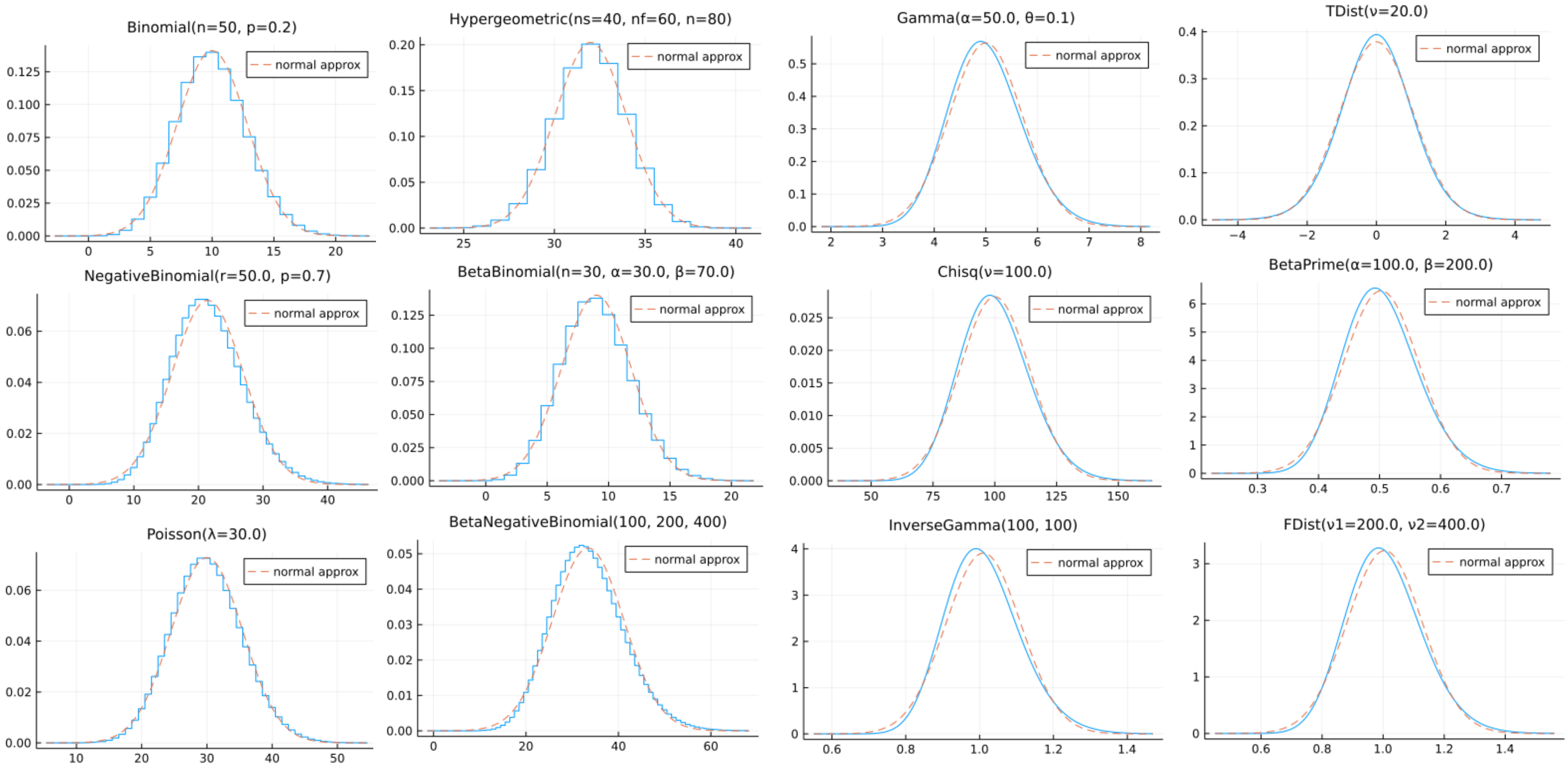

```
plot(plot_betanormal(48, 112), plot_betanormal(192, 448); size=(800, 250))
```



ベータ分布の正規分布近似

正規分布は普偏的!

多くの分布はパラメータに適当な条件を仮定すると正規分布で近似される.



中心極限定理の証明 (の考え方) ← 詳細はGitHubの資料を参照
理解してほしいが理解できなくても単位は出したい!!

どこに神経を集中するべきか

単に正規分布に収束することだけではなく,

(*)　どのようなときに正規分布への収束がおそくなりそうであるか

がわかるようにするべきである. その部分がしっかりしていれば, 論理的に少々ギャップを残してもかまわない.

$\mu = E[X_i]$, $\sigma^2 = Var(X_i)$ とおく.

(*)の答え　$X_1, X_2, \ldots$ は各々が分布Dにしたがう独立同分布確率変数列であるとする.

$\mu = E[X_i]$, $\sigma^2 = Var(X_i)$, $Z_n = \dfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \dfrac{X_i - \mu}{\sigma}$ とおく.

Z_n の分布の正規分布への収束の速さについて,

- 歪度(わいど) $\kappa_3 = E\left[\left(\dfrac{X_i-\mu}{\sigma}\right)^3\right]$ の絶対値が大きいと正規分布への収束がおそくなる.
- $\kappa_3 = 0$ のときは尖度(せんど) $\kappa_4 = E\left[\left(\dfrac{X_i-\mu}{\sigma}\right)^4\right] - 3$ の絶対値の大きさがきいてくる.

特性関数, モーメント母関数, キュムラント母関数

X は確率変数とする.

特性関数 $\varphi_X(t) = E[e^{itX}] = M_X(it) = e^{K_X(it)}$ 常に定義される

モーメント母関数 $M_X(t) = E[e^{tX}] = \varphi_X(-it) = e^{K_X(t)}$

キュムラント母関数 $K_X(t) = \log E[e^{tX}] = \log M_X(t) = \log \varphi_X(-it)$

おおざっぱには本質的に同じ情報を持っていると考えてよい.

特性関数で期待値を書ける

← これより, $\varphi_X(t)$, $M_X(t)$, $K_X(t)$ が等しい X たちは同じ分布を持つことがわかる.

$$\hat{f}(t) = \int_{-\infty}^{\infty} f(x)\, e^{-itx}\, dx \quad (\text{Fourier変換})$$

とおくと,

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \hat{f}(t)\, e^{itx}\, dt \quad (\text{逆 Fourier変換}).$$

ゆえに,

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

したがって, 確率変数 X と Y について適切な意味で $\varphi_X(t)$ と $\varphi_Y(t)$ が近ければ

$$E[f(X)] \approx E[f(Y)] \quad (E[f(X)] \text{ と } E[f(Y)] \text{ も近くなる}).$$

例 $Z \sim \mathrm{Normal}(0,1)$ のとき, $M_Z(t) = E[e^{tZ}] = e^{t^2/2}$

証明 $E[e^{tZ}] = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\infty} e^{tz}\, e^{-z^2/2}\,dz = \dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-t)^2+\frac{1}{2}t^2}\,dz$

$z = y + t$ とおく.

$= \dfrac{e^{t^2/2}}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{\infty} e^{-\frac{1}{2}y^2}\,dy = \dfrac{e^{t^2/2}}{\sqrt{2\pi}}\sqrt{2\pi} = e^{t^2/2}.$ **q.e.d.**

モーメント $\mu_m = \mu_m(X) = E[X^m]$ とおき, X の m 次のモーメントと呼ぶ.

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

← $M_X(t)$ はモーメント達の母函数

キュムラント 次の展開で X のキュムラント (cumulant) $\kappa_m = \kappa_m(X)$ を定める:

$$K_X(t) = \log M_X(t) = \sum_{m=1}^{\infty} \kappa_m \frac{t^m}{m!} = \underbrace{\mu t + \sigma^2 \frac{t^2}{2} + \mu_3 \frac{t^3}{3!} + (\mu_4 - 3)\frac{t^4}{4!} + \cdots}_{\text{すぐ使う.}}$$

例 $\underbrace{Z \sim \mathrm{Normal}(0,1)}$ のとき, $\underbrace{K_Z(t) = \log M_Z(t) = \frac{t^2}{2}}$ より,

$\kappa_1(Z) = 0,\quad \kappa_2(Z) = 1,\quad \kappa_3(Z) = 0,\quad \kappa_4(Z) = 0,\quad \kappa_5(Z) = 0,\quad \ldots$ □

標準化キュムラント X, $\mu = E[X]$, $\sigma^2 = \mathrm{var}(X)$ に対して, $\dfrac{X-\mu}{\sigma}$ を X の標準化と呼び $\bar\kappa_m(X) = \kappa_m\left(\dfrac{X-\mu}{\sigma}\right)$ を X の標準化キュムラントと呼ぶ.

$\bar\kappa_3(X) = E\left[\left(\dfrac{X-\mu}{\sigma}\right)^3\right]$ を歪度(わいど)と, $\bar\kappa_4(X) = E\left[\left(\dfrac{X-\mu}{\sigma}\right)^4\right] - 3$ を尖度(せんど)と呼ぶ.

キュムラント母函数を用いた中心極限定理の証明

$Y_i = \dfrac{X_i - \mu}{\sigma}$ とおくと, $E[Y_i] = 0$, $\operatorname{var}(Y_i) = E[Y_i^2] = \dfrac{1}{\sigma^2}\overbrace{E[(X_i-\mu)^2]}^{=\sigma^2} = 1$ より,

$$\log E[e^{tY_i}] = \frac{t^2}{2} + \bar{\kappa}_3 \frac{t^3}{3!} + \bar{\kappa}_4 \frac{t^4}{4!} + \cdots$$

$\leftarrow$ $\bar{\kappa}_m = \bar{\kappa}_m(X_i) = \bar{\kappa}_m(Y_i)$

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} = \frac{1}{\sqrt{n}}(Y_1 + \cdots + Y_n)$$ について,

$Y_1, \ldots, Y_n$ たちの独立性より

$$\log E[e^{tZ_n}] = \log E\left[e^{\frac{t}{\sqrt{n}}Y_1} \cdots e^{\frac{t}{\sqrt{n}}Y_n}\right] = \log \prod_{i=1}^n E\left[e^{\frac{t}{\sqrt{n}}Y_i}\right] = \sum_{i=1}^n \log E\left[e^{\frac{t}{\sqrt{n}}Y_i}\right]$$

上の式の t に $\frac{t}{\sqrt{n}}$ を代入する

$$= \sum_{i=1}^n \left(\frac{1}{2}\frac{t^2}{n} + \frac{\bar{\kappa}_3}{3!}\frac{t^3}{n\sqrt{n}} + \frac{\bar{\kappa}_4}{4!}\frac{t^4}{n^2} + \cdots\right)$$

$$= \frac{t^2}{2} + \frac{\bar{\kappa}_3}{3!}\frac{t^3}{\sqrt{n}} + \frac{\bar{\kappa}_4}{4!}\frac{t^4}{n} + \cdots$$

ここを見れば
歪度(わいど) $\bar{\kappa}_3$ と尖度(せんど) $\bar{\kappa}_4$ が
収束を遅くすることがわかる

なので, $Z \sim \mathrm{Normal}(0,1)$ ならば, $n \to \infty$ のとき,

$$\log E[e^{tZ_n}] \to \frac{t^2}{2} = \log E[e^{tZ}]$$

$$\therefore E[e^{itZ_n}] \to E[e^{itZ}]$$

$$\therefore E[f(Z_n)] \to E[f(Z)] = \int_{-\infty}^{\infty} f(z) \frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz.$$

雑な説明だが
正当化できる.

$\boxed{\text{q.e.d.}}$

$\boxed{\text{Fourier解析を使わずに } f(x) \text{ のTaylor展開のみを使う中心極限定理の証明もある.}}$

$X_1, X_2, \dots$ は独立同分布, $E[X_i]=0$, $E[X_i^2]=1$
$Y_1, Y_2, \dots$ は独立同分布, $Y_k \sim \mathrm{Normal}(0,1)$
で, $X_1, X_2, \dots$ と $Y_1, Y_2, \dots$ の全体も独立のとき

$$A_n^{(k)} = \frac{X_1 + \cdots + X_{k-1} + Y_{k+1} + \cdots + Y_n}{\sqrt{n}}$$ とおき,

$$f(a+h) = f(a) + f'(a)h + \frac{f''(a)}{2}h^2 + \frac{f'''(a)}{3!}h^3 + \frac{f^{(4)}(a)}{4!}h^4 + O(h^5)$$ を使うと,

$E[Y_k]=0,\ E[Y_k^2]=1,$
$E[Y_k^3]=0,\ E[Y_k^4]=3, \dots$

$$E\left[f\left(\frac{X_1+\cdots+X_k+Y_{k+1}+\cdots+Y_n}{\sqrt{n}}\right)\right] - E\left[f\left(\frac{X_1+\cdots+X_{k-1}+Y_k+\cdots+Y_n}{\sqrt{n}}\right)\right]$$
$$= \frac{E[f'''(A_n^{(k)})]}{3!\,n\sqrt{n}}\underbrace{E[X_k^3]}_{=\kappa_3} + \frac{E[f^{(4)}(A_n^{(k)})]}{4!\,n^2}\underbrace{(E[X_k^4]-3)}_{=\kappa_4} + O\left(\frac{1}{n^2\sqrt{n}}\right)$$ を示せる.

$k=1,2,\dots,n$ に対する n 個のこれらを足し上げることによって, ある $C_3>0$, $C_4>0$ が存在して,

$$\left|E\left[f\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right] - \underbrace{E\left[f\left(\frac{Y_1+\cdots+Y_n}{\sqrt{n}}\right)\right]}_{=E[f(Z)] \text{ for } Z\sim\mathrm{Normal}(0,1)}\right| \leqq \frac{C_3}{\sqrt{n}}|\kappa_3| + \frac{C_4}{n}|\kappa_4| + O\left(\frac{1}{n\sqrt{n}}\right)$$

の形の評価が得られる.

したがって, $n\to\infty$ のとき, $E\left[f\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right] \to E[f(Z)]$ for $Z\sim\mathrm{Normal}(0,1)$.

$\boxed{\text{q.e.d.}}$

条件付き確率分布、尤度、推定、要約

Part 1　Part 2

現実世界のデータ ⟷ 数学の世界での統計モデル

比較

↓

様々な統計分析

使われる基本的な道具達

今回

- 統計モデル内の確率分布を現実から得たデータで条件付けて条件付き確率分布を作ること
- 統計モデルの現実から得たデータの数値への適合度（フィッティング）の指標としての尤度（ゆうど, likelihood）

および P値（ピー値, P-value） ← 別の回でやる

や 事後分布や予測分布 ← この講義よりすすんだ話題

条件付き確率分布

離散分布での条件付き確率分布

$x=(x_1,\dots,x_n)$ に関する確率質量函数 $P(x)$ を考える.

函数 $y=f(x)$ に対応する確率変数 $Y=f(X)$ を考える:

$$E[g(Y)] = \sum_x g(f(x))\,P(x).$$

(その函数の期待値が定義されている変数を確率変数とよぶのであった.)

値 y について, $P(Y=y) = (Y=y\text{ となる確率}) = \sum_{f(x)=y} P(x) > 0$ と仮定する.

条件 $Y=y$ で定義された x の条件付き確率分布を

$P(x|y)$ ← パラメータ

次のパラメータ y を持つ x の確率質量函数 $P(x|y)$ ($f(x)=y$) によって定める:

$$P(x|y) = \frac{P(x)}{P(Y=y)} \quad (x\text{ は } f(x)=y \text{ をみたすものに制限}).$$

{ $f(x)=y$ をみたす x に制限して, $f(x)=y$ をみたす x のどれかが生じる確率を1としたときの x の確率

確率の総和が1になることの確認:

$$\sum_{f(x)=y} \frac{P(x)}{P(Y=y)} = \frac{1}{P(Y=y)} \overbrace{\sum_{f(x)=y} P(x)}^{=P(Y=y)} = 1.$$

$\boxed{例}$ 正二十面体のサイコロ：

$$\begin{cases} 6つの面 \longrightarrow 1 & P(1) = 6/20 \\ 5つの面 \longrightarrow 2 & P(2) = 5/20 \\ 4つの面 \longrightarrow 3 & P(3) = 4/20 \\ 3つの面 \longrightarrow 4 & P(4) = 3/20 \\ 2つの面 \longrightarrow 5 & P(5) = 2/20 \end{cases}$$

$x \in \{1, 2, 3, 4, 5\}$ ← $f(x)$の定義

$$y = f(x) = \begin{cases} 0 & (x\text{は偶数}) \\ 1 & (x\text{は奇数}) \end{cases}, \quad Y = f(X).$$

$$P(Y=0) = P(2) + P(4) = 8/20$$

$$P(Y=1) = P(1) + P(3) + P(5) = 12/20$$

条件 $Y=0$ (xは偶数)で定義された x の条件付き確率分布：

$$x \in \{2, 4\}, \quad P(2|0) = \frac{P(2)}{P(Y=0)} = \frac{5/20}{8/20} = \frac{5}{8}, \quad P(4|0) = \frac{P(4)}{P(Y=0)} = \frac{3/20}{8/20} = \frac{3}{8}.$$

$$\underbrace{2\ 2\ 2\ 2\ 2}_{5/8}\ \underbrace{4\ 4\ 4}_{3/8} \longleftarrow P(x|0) \quad (x\text{は偶数})$$

条件 $Y=1$ (xは奇数)で定義された x の条件付き確率分布：

$$x \in \{1, 3, 5\}, \quad P(1|1) = \frac{P(1)}{P(Y=1)} = \frac{6}{12}, \quad P(3|1) = \frac{P(3)}{P(Y=1)} = \frac{4}{12}, \quad P(5|1) = \frac{P(5)}{P(Y=1)} = \frac{2}{12},$$

$$\underbrace{1\ 1\ 1\ 1\ 1\ 1}_{6/12}\ \underbrace{3\ 3\ 3\ 3}_{4/12}\ \underbrace{5\ 5}_{2/12} \longleftarrow P(x|1) \quad (x\text{は奇数})$$

離散分布でのBayesの定理(ベイズの定理)

← 有名だが自明でつまらない結果.
知らなくてもこまらない.

(x,y) に関する確率質量函数 $P(x,y)$ について考える. $x \leftrightarrow X$, $y \leftrightarrow Y$

$$P(x) = P(X=x) = \sum_y P(x,y), \quad P(y) = P(Y=y) = \sum_x P(x,y) \text{ と書く.} \quad \longleftarrow (1)$$

条件付き確率分布の定義: $P(x|y) = \dfrac{P(x,y)}{P(Y=y)} = \dfrac{P(x,y)}{P(y)}$, $P(y|x) = \dfrac{P(x,y)}{P(X=x)} = \dfrac{P(x,y)}{P(x)}$. ← (2)

すなわち, $P(x,y) = P(x|y)P(y) = P(y|x)P(x)$. ← (2)'

ゆえに, $P(x) \overset{(1)}{\underset{(2)'}{=}} \sum_y P(x|y)P(y)$, $P(y) \overset{(1)}{\underset{(2)'}{=}} \sum_x P(y|x)P(x)$. ← (3)

したがって,

$$P(y|x) \overset{(2)'}{=} \frac{P(x|y)P(y)}{P(x)} \overset{(3)}{=} \frac{P(x|y)P(y)}{\sum_y P(x|y)P(y)},$$

$$P(x|y) \overset{(2)'}{=} \frac{P(y|x)P(x)}{P(y)} \overset{(3)}{=} \frac{P(y|x)P(x)}{\sum_x P(y|x)P(x)}.$$

(1), (2)', (3) から自明

Bayesの定理

$P(x,y) = P(x|y)P(y)$ (by (2)')

$P(y) = \sum_x P(x,y)$ (by (1))

$P(y|x) = \dfrac{P(x,y)}{P(y)}$ (by (2))

これでOK

Bayesの定理は $P(x|y)$ と $P(y)$ から $P(y|x)$ を作ったり,
$P(y|x)$ と $P(x)$ から $P(x|y)$ を作ったりするために使われる.

2×2の分割表での条件付き確率分布（偽陽性，偽陰性の確率について） ← 重要

病気Dにかかっているかどうかの検査をする状況を考える．

検査する人が実際に病気にかかっている確率（有病率）は p であるとする．

その検査の性能は以下の通りとする：

- 病気Dにかかっている人は75%の確率で陽性になる． ← 感度 75%
- 病気Dにかかっていない人は95%の確率で陰性になる． ← 特異度 95%

	病気有	病気無	
陽性	$0.75p$	$0.05(1-p)$	$0.05+0.70p$
陰性	$0.25p$	$0.95(1-p)$	$0.95-0.70p$
	p	$1-p$	1

← 常にこの表を書くようにすると
まちがわずにすむ．

式で書くと，

$$P(\text{陽性},\text{病気有}) = 0.75p, \qquad P(\text{陽性},\text{病気無}) = 0.05(1-p),$$
$$P(\text{陰性},\text{病気有}) = 0.25p, \qquad P(\text{陰性},\text{病気無}) = 0.95(1-p).$$

知りたいこと
- 陽性のときの，病気無の条件付き確率（偽陽性率）
- 陰性のときの，病気有の条件付き確率（偽陰性率）

つづく

	病気有	病気無	
陽性	$0.75p$	$0.05(1-p)$	$0.05+0.70p$
陰性	$0.25p$	$0.95(1-p)$	$0.95-0.70p$
	p	$1-p$	1

$$P(\text{陽性}) = P(\text{陽性}, \text{病気有}) + P(\text{陽性}, \text{病気無}) = 0.05 + 0.70p,$$
$$P(\text{陰性}) = P(\text{陰性}, \text{病気有}) + P(\text{陰性}, \text{病気無}) = 0.95 - 0.70p.$$

$$(\text{偽陽性率}) := P(\text{病気無}|\text{陽性}) = \frac{P(\text{陽性}, \text{病気無})}{P(\text{陽性})} = \frac{0.05(1-p)}{0.05+0.70p},$$
$$(\text{偽陰性率}) := P(\text{病気有}|\text{陰性}) = \frac{P(\text{陰性}, \text{病気有})}{P(\text{陰性})} = \frac{0.25p}{0.95-0.70p},$$

このように2×2の表を書けば易しい!

自習問題

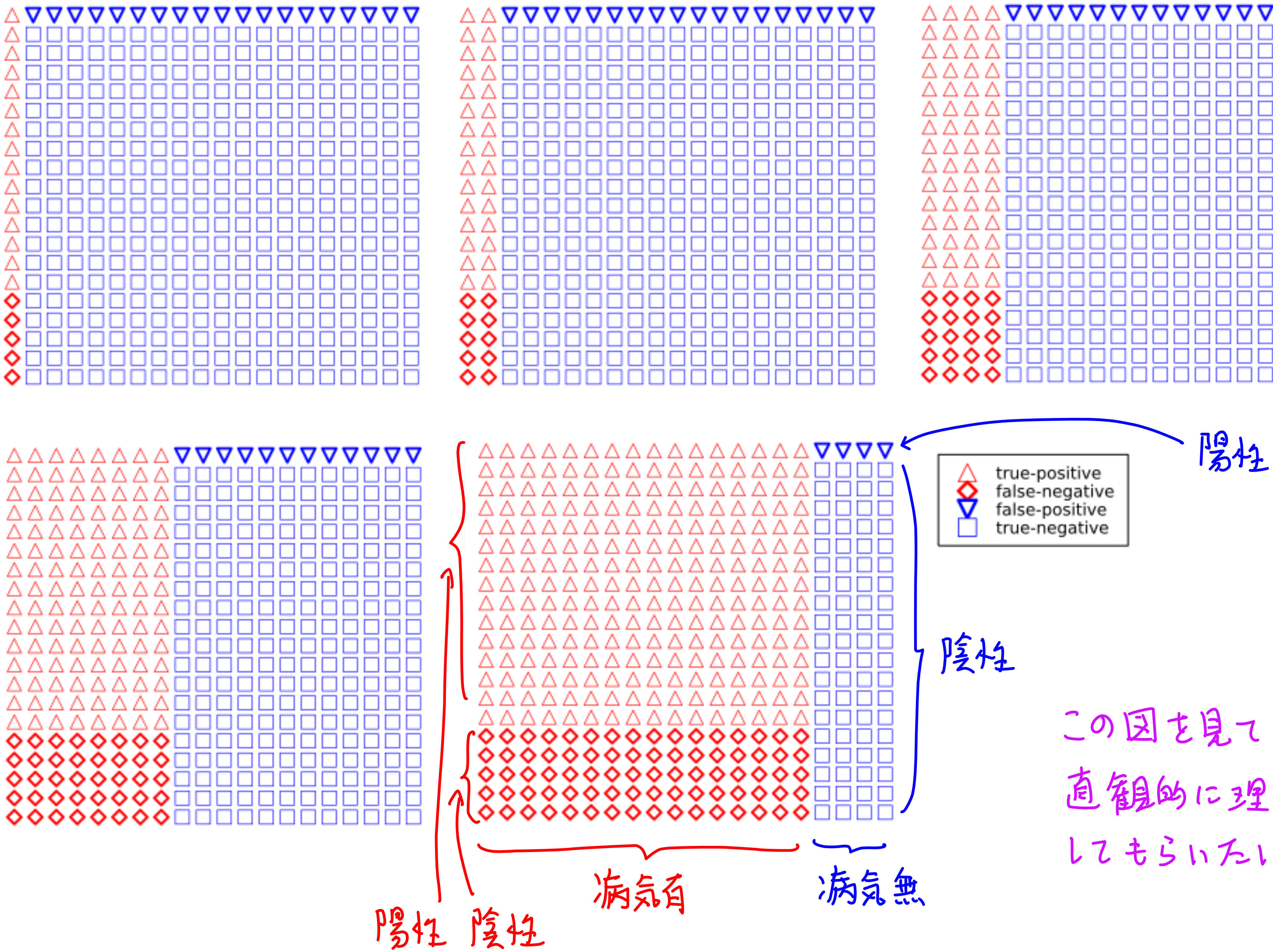
必修の易しい計算問題: 有病率によって偽陽性率と偽陰性率がどのように変化するか

上の状況において, 有病率 p が 5%, 10%, 20%, 40%, 80% の場合の偽陽性率と偽陰性率を求めよ.

次ページにヒントあり!

前ページの自習問題のヒント

赤=病気有，青=病気無，△と▽は陽性



この図を見て
直観的に理解
してもらいたい，

おまけの問題: モンティ・ホール問題 (Monty Hall problem)

以下のようなゲームを考える:

(1) 親は $1, 2, 3$ の中から番号 X を無作為に選んで秘密にしておく.

(2) あなたは, $1, 2, 3$ の中から番号 1 を選んでそれを親に告げる.

(3) 親は $1, 2, 3$ の中から 1 と X 以外の番号 Y を無作為に選んで, あなたに告げる.

(4) あなたは $1, 2, 3$ の中から 1, Y を除いて残った1つの番号 Z を選ぶ.

あなたが最後に選んだ番号 Z が秘密の番号 X に一致する確率を求めよ.

注意: (2)で 1 を選ばずに $1, 2, 3$ のどれかを無作為に選んでも結果は同じになる.

代表的誤答例: $Z = X$ となる確率は $1/2$ である. **誤答終**

例 親 $X=2$ → あなた 1 → 親 $Y=3$ → あなた 1と3以外の $Z=2$

親 $X=1$ → あなた 1 → 親 $Y=2$ → あなた $Z=3$

→ 親 $Y=3$ → あなた $Z=2$

連続分布の場合の条件付き確率分布とBayesの定理

確率質量関数と和 $\longrightarrow$ 確率密度関数と積分　とするだけで同様

(x,y)の確率密度関数 $p(x,y)$ を考える.

$$p(x) = p(X=x) = \int p(x,y)\,dy, \quad p(y) = p(Y=y) = \int p(x,y)\,dx. \quad \longleftarrow (1)$$

$$p(x|y) = \frac{p(x,y)}{p(y)}, \quad p(y|x) = \frac{p(x,y)}{p(x)} \quad \longleftarrow (2)$$

$$p(x,y) = p(x|y)\,p(y) = p(y|x)\,p(x) \quad \longleftarrow (2)'$$

$\longleftarrow$ これがポイント!

$$p(x) \underset{(2)'}{\overset{(1)}{=}} \int p(x|y)\,p(y)\,dy, \quad p(y) \underset{(2)'}{\overset{(1)}{=}} \int p(y|x)\,p(x)\,dx \quad \longleftarrow (3)$$

$$\left.\begin{aligned} p(y|x) &\overset{(2)'}{=} \frac{p(x|y)\,p(y)}{p(x)} \overset{(3)}{=} \frac{p(x|y)\,p(y)}{\int p(x|y)\,p(y)\,dy} \\ p(x|y) &\overset{(2)'}{=} \frac{p(y|x)\,p(x)}{p(y)} \overset{(3)}{=} \frac{p(y|x)\,p(x)}{\int p(y|x)\,p(x)\,dx} \end{aligned}\right\} \text{Bayesの定理}$$

数学に強くなりたい人のためのBayes統計の自習問題

問題: 連続分布の条件付き確率分布の例 (分散が固定された正規分布モデルのBayes統計)

無理して解かなくても良い。
ゆえに文字が小さい。

しかし，目を通しておいて，
必要になったら思い出すとよい。

答えはGitHubにある。

この節の式には細かい誤りが多数含まれている可能性がある. 誤りを見付けたら自分で訂正すること. 誤りがなかったら, 拍手喝采して欲しい.

$n, \rho > 0$ を任意に取って固定する. $x, \mu \in \mathbb{R}$ の同時確率密度函数

(prior)
事前分布

$$p(x, \mu | n, \rho) = \frac{e^{-n(x-\mu)^2/2}}{\sqrt{2\pi/n}} \frac{e^{-\mu^2/(2\rho^2)}}{\sqrt{2\pi\rho^2}}.$$

によって定義される $(x, \mu) \in \mathbb{R}^2$ の確率分布を考える. その分布において, 変数 x に対応する確率変数 X に関する条件 $X = x$ で定義された μ に関する条件付き確率分布の密度函数が

$$p(\mu | x, n, \rho) = \frac{1}{\sqrt{2\pi\rho^2/(1+n\rho^2)}} \exp\left(-\frac{1+n\rho^2}{2\rho^2}\left(\mu - \frac{n\rho^2 x}{1+n\rho^2}\right)^2\right).$$

事後分布 (posterior)

になることを示せ. これは平均と分散がそれぞれ

$$\frac{n\rho^2}{1+n\rho^2}x, \qquad \frac{\rho^2}{1+n\rho^2}$$

の正規分布である.

注意: 特に, $n \to \infty$ とすると, 条件 $X = x$ で定義された μ の条件付き確率分布の期待値は x に収束し, 分散は 0 に収束するので, μ の条件付き確率分布は x の近くに集中するようになる.

注意: この問題の例は正規分布モデル $p(x|n, \mu) = e^{-n(x-\mu)^2/2}/\sqrt{2\pi/n}$ と事前分布 $p(\mu|\rho) = e^{-\mu^2/(2\rho^2)}/\sqrt{2\pi\rho^2}$ に関するBayes統計におけるデータの数値 x が定める事後分布 $p(\mu|x, n, \rho)$ を求める計算になっている. x は標本平均に, n はサンプルサイズに対応しており, $p(x|n, \mu)$ の形は分散 1 の正規分布のサイズ n 標本分布のおいて標本平均の分散が $1/n$ になることを表している. だから, この場合には固定された分散 1 を持つ正規分布の標本分布を統計モデルとして採用している場合になっていると考えて欲しい. $n = 0$ のとき事後分布が事前分布に一致することにも注意せよ.

注意: 条件付き確率分布の概念はBayes統計に限らず統計学における最も基本的かつ重要な確率論の道具であるが, 特にBayes統計では条件付き確率分布を使った推論が系統的に利用される.

分散が固定された正規分布モデルのBayes統計の事後分布の視覚化

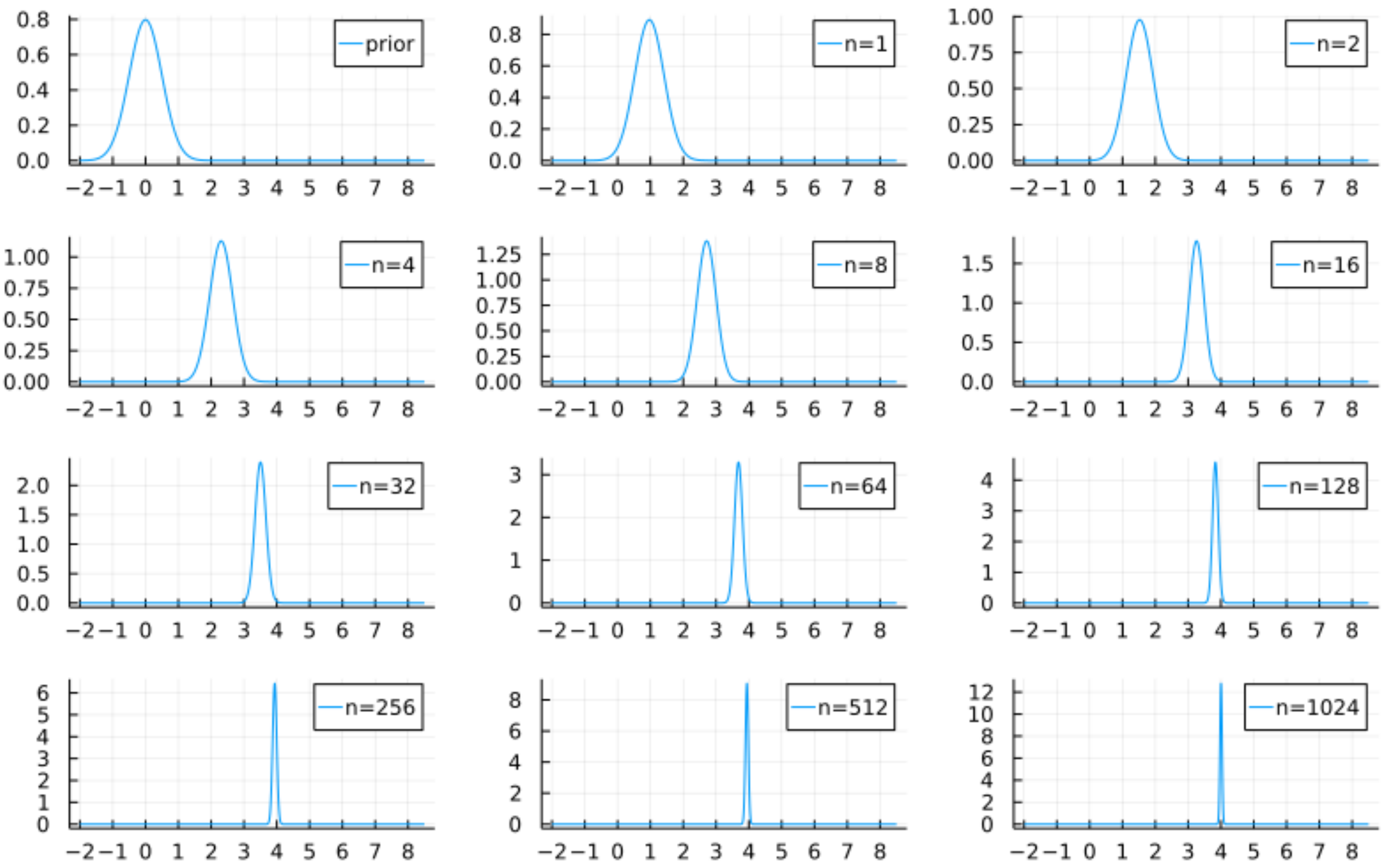
前ページの問題のつづきの解説

$\mu_0 = 4, \rho = 1/2$ とおく.

平均 $\mu_0 = 4$, 分散 1 の正規分布のサイズ $N = 2^{10}$ の標本 $X_1, X_2, \dots, X_N$ をランダムに生成し, $n = 0, 1, 2, 4, 8, \dots, 1024 = 2^{10}$ について標本平均 $x = \bar{X}_n = (X_1 + \cdots + X_n)/n$ を求め, 前節で求めた μ の事後分布 $p(\mu|x, n, \rho)$ をプロットしてみよう.

n を大きくすると μ の事後分布はサンプルを生成した分布の平均値 μ_0 に集中して行く.

これはBayes統計の漸近論的な基礎付けの最も簡単な場合になっている.



データは平均4分散1の正規分布のn個の乱数として生成されている.

n→大で事後分布とよばれる条件付き確率分布が4に集中して行く!

これも数学に強くなりたい人向けの問題

これも無理して解かなくてもよい.

問題: 2変量の正規分布とその条件付き確率分布の例

$$\Sigma = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$$

は固有値がすべて正の実対称行列であるとし, $\mu = (\mu_x, \mu_y) \in \mathbb{R}^2$ であるとする. このとき,

$$\Sigma^{-1} = \frac{1}{ac - b^2} \begin{bmatrix} c & -b \\ -b & a \end{bmatrix}$$

であり, $(x, y) \in \mathbb{R}^2$ の確率密度函数を

$$\begin{aligned} p(x, y) &= \frac{1}{\det(2\pi\Sigma)^{1/2}} \exp\left(-\frac{1}{2}\left[x - \mu_x,\ \ y - \mu_y\right]\Sigma^{-1}\begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix}\right) \\ &= \frac{1}{\sqrt{(2\pi)^2(ac - b^2)}} \exp\left(-\frac{a(y - \mu_y)^2 - 2b(x - \mu_x)(y - \mu_y) + c(x - \mu_x)^2}{2(ac - b^2)}\right) \end{aligned}$$

と定めることができる. これが $\iint_{\mathbb{R}^2} p(x, y)\,dx\,dy = 1$ を満たすことは, Σ を直交行列で対角すれば示せる. (詳細は略す. 以下では認めて使ってよい.) この確率密度函数が定める確率分布を **2変量正規分布** と呼び,

$$\mathrm{MvNormal}(\mu, \Sigma)$$

と表すことにする. μ を平均(もしくは期待値)と呼び, Σ を **分散共分散行列** (variance-covariance matrix)と呼ぶ. (この記号法は, 2変量の場合に限定せずに, 任意の **多変量正規分布** (multivariate normal distribution)にも使用することにする.)

この場合には条件 $X = x$ が定める y に関する条件付き確率分布が, 平均と分散がそれぞれ

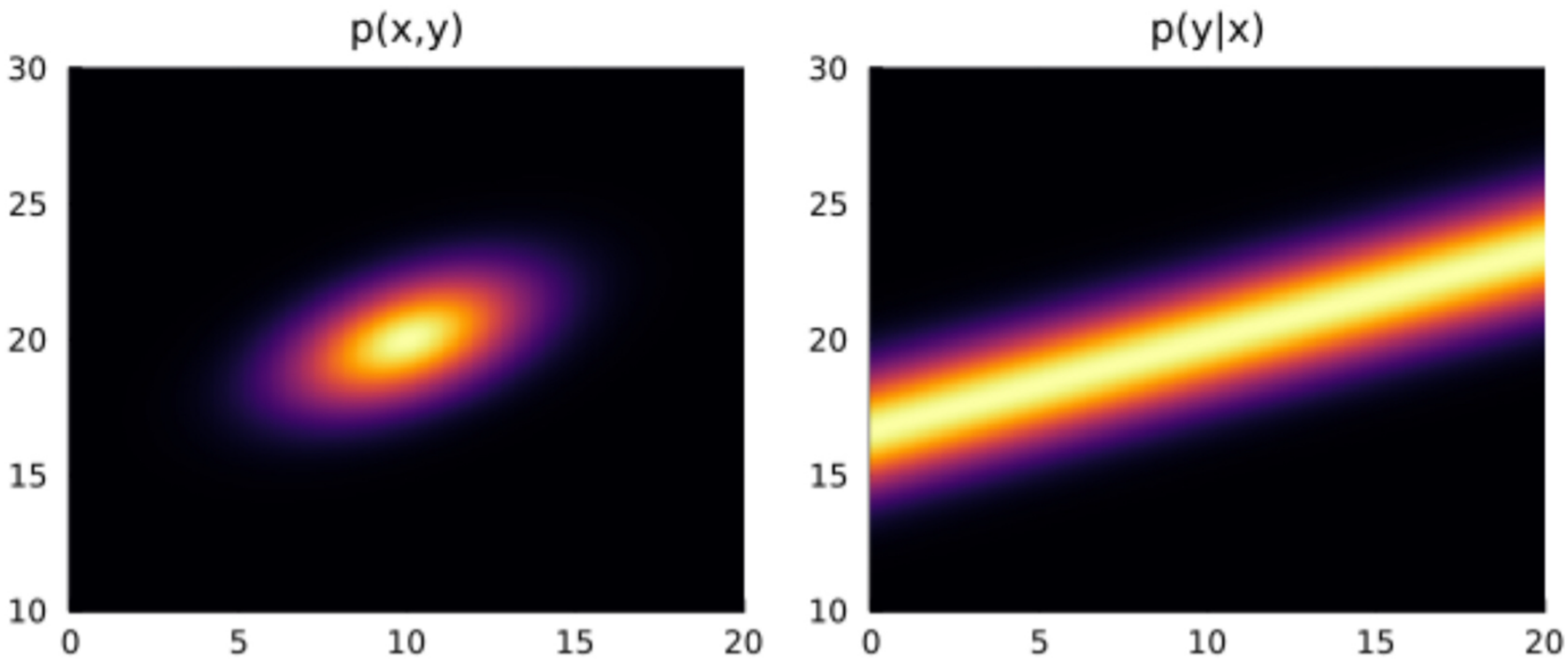
$$\mu = (b/a)(x - \mu_x) + \mu_y, \quad \sigma^2 = (ac - b^2)/a$$

の正規分布になることを示せ.

前ページのつづき

$\mu = (10, 20)$, $\Sigma = \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix}$ の場合の $p(x,y)$ と $p(y|x) = \dfrac{p(x,y)}{p(x)}$ の視覚化

$p(x) = \int p(x,y)\,dy$



$p(x,y)$ は xy 平面上の確率密度函数になっている.

$p(y|x)$ は各 x ごとに y 軸に平行な直線上の確率密度函数になっている.

その様子の違いの典型例が上のグラフからわかる.

$p(y|x)$ のヒートマップを見れば, パラメータ x を持つ y の確率密度函数は x を y に対応させる函数の一般化(x に対応する y の値が確率的に揺らぐ)になっていることもわかる.

$\boxed{\text{尤度と推定}}$

尤度は「ゆうど」と読む．犬度(いぬど)ではない．

$\boxed{\text{統計モデルの定義}}$ パラメータ $\theta = (\theta_1, \ldots, \theta_d)$ を持つ $x = (x_1, \ldots, x_n)$ に関する確率質量函数 $P(x|\theta)$ または確率密度函数 $p(x|\theta)$ で定まる確率分布を現実世界におけるデータ x の生成のされ方の数学的モデルとみなすとき，その確率分布を<u>統計モデル</u>と呼ぶ．(単に<u>モデル</u>ということも多い．)

$\boxed{\text{尤度の定義}}$ データの数値 x が得られたとき，パラメータ θ を統計モデル内で数値 x が生成される確率 $P(x|\theta)$ または確率の密度 $p(x|\theta)$ に対応させる函数

$$\theta \longmapsto P(x|\theta) \quad \text{または} \quad \theta \longmapsto p(x|\theta)$$

をデータの数値 x に関する統計モデルの<u>尤度函数</u> (likelihood function) と呼ぶ．

さらにパラメータの値 θ を与えたときの尤度函数の値 $P(x|\theta)$ または $p(x|\theta)$ をデータの数値 x とパラメータの数値 θ に関する統計モデルの<u>尤度</u> (likelihood) と呼ぶ．

尤度の使い方.

尤度＝モデル内でデータと同じ数値が生成される確率または確率密度.

尤度はデータの数値 x へのパラメータ値が θ の統計モデルの適合度（フィッティング）の指標の1つとして基本的かつ重要である.

尤度＝データの数値 x へのモデルの適合度（フィッティッング）の指標のひとつ

最尤法 (maximum likelihood method, 最尤推定法, maximum likelihood estimation)

データの数値 x にパラメータ θ を持つ統計モデルを最もよくフィットさせるために尤度が最大になるようなパラメータ値 $\hat{\theta}$ をみつけて,

パラメータを $\theta = \hat{\theta}$ と設定することを最尤法という.

最尤法の解 $\hat{\theta}$ を最尤推定値 (maximum likelihood estimate) とよぶ.

注意 尤度は「もっともらしさ」の指標としては使えない. ← オーバーフィッティングの問題がある

例 Bernoulli試行モデルの最尤法　$x_i = 1, 0$

あたりが未知の確率で出るルーレットを回した結果を，あたりを1，はずれを0として記録したデータの数値を $x_1, x_2, \dots, x_n$ と書く。

$$P(x_1, \dots, x_n \mid p) = p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)}$$ ← Bernoulli試行の統計モデル

尤度函数　$L(p) = p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)}$　($x_1, \dots, x_n$ はデータの数値)

データの要約　$k = x_1 + \cdots + x_n$ とおく。「n 回中 k 回あたりが出た。」

対数尤度函数　$\ell(p) = \log L(p) = k \log p + (n-k)\log(1-p)$.

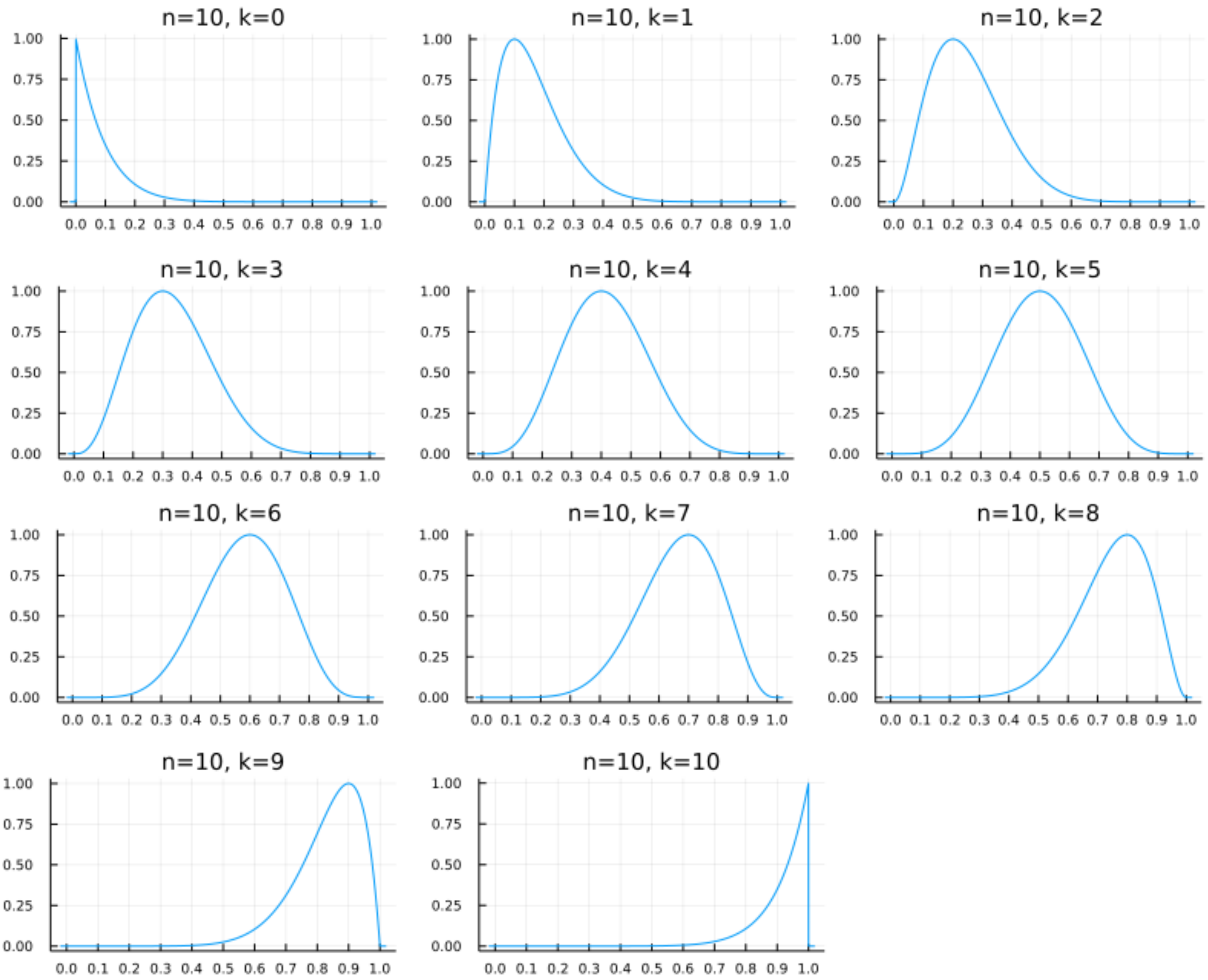
最尤法の計算　$\ell'(p) = \dfrac{k}{p} - \dfrac{n-k}{1-p} = \dfrac{k-np}{p(1-p)} = 0$ の解 $\hat{p} = \dfrac{k}{n}$.

解釈　n 回中 k 回あたりが出たというデータに最尤法の意味で最もフィットするBernoulli試行モデルのパラメータ値は $\hat{p} = \dfrac{k}{n}$ である。

注意　データが運悪くかたよっていなければ，ルーレットであたりが出る未知の確率は $\hat{p} = \dfrac{k}{n}$ に近い値になっているだろうが，運が悪ければダメ！
その「運」を確率論の言葉で見つもることが重要である。

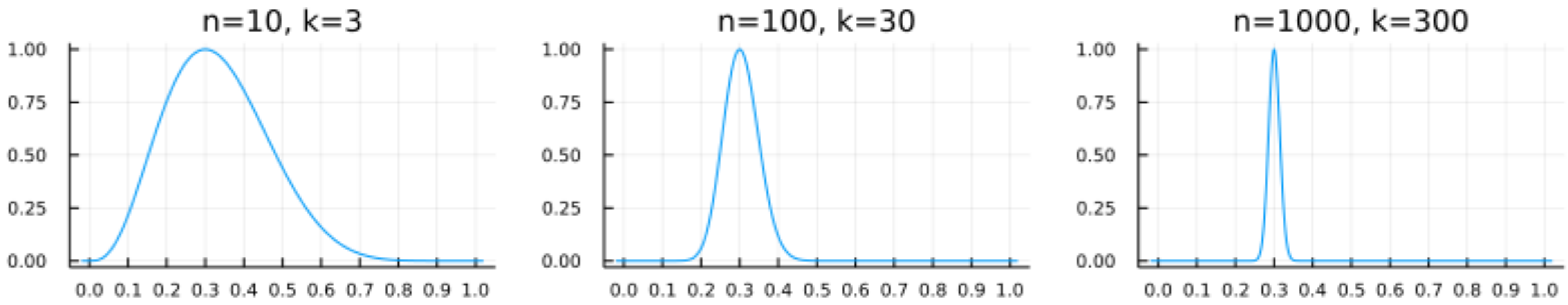
重要.

Bernoulli試行モデルの尤度函数（の定数倍）のグラフ (1)



Bernoulli試行モデルの尤度函数の定数倍のグラフ(2)

k/n を固定して n を大きくした場合には, 尤度函数 $L(p)$ の台は k/n に集中して行く.



このグラフから, データに基づくパラメータの推定において, パラメータの推定値としてどれだけの幅を持たせるかについて, 尤度函数がなにがしかの情報を持っていそうなこともわかる.

正則モデルと呼ばれるクラスに属する統計モデルでも同様のことが起こる.

重要自習問題

答えは
GitHubの
資料にある

データ中のi番目が
• 当たり ⇒ モデル内での
i番目の当たりの確率を
1とする.
• はずれ ⇒ モデル内での
i番目の当たりの確率を
0とする.
こうするとモデル1の
最尤法の解になる.

問題: 尤度が高くても全然もっともらしくない例(オーバーフィッティングの例)

データは当たり(1 と書く)とはずれ(0 と書く)がランダムに出るルーレットを n 回まわして得た 1 と 0 の列 $x_1, x_2, \ldots, x_n$ として得られるとする.

統計モデルとして以下の2つを考える.

モデル0: Bernoulli試行モデル: このモデルのパラメータ $0 \le p \le 1$ を持つ確率質量函数は次のように書けるのであった:

$$P_0(x_1, \ldots, x_n | p) = p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)} \quad (x_i = 1, 0).$$

モデル1: ルーレットをまわすごとに当たりが出る確率が変わってもよいと想定したモデル: このモデルは n 個のパラメータ $0 \le p_1, \ldots, p_n \le 1$ を持つ次の確率質量函数によって与えられる:

$$P_1(x_1, \ldots, x_n | p_1, \ldots, p_n) = \prod_{i=1}^{n} \left(p_i^{x_i}(1-p_i)^{1-x_i}\right) \quad (x_i = 1, 0).$$

データ(1 と 0 の列) $x_1, \ldots, x_n$ について, モデル0の最尤法の解 $p = \hat{p}$ は $\hat{p} = \dfrac{k}{n}$ ($k = x_1 + \cdots + x_n$) になるのであった. 以下の2つを同時に示せ:

(1) モデル1の最尤法の解 $p_i = \hat{p}_i$ は $\hat{p}_i = x_i$ になる.

(2) どのようなデータ $x_1, x_2, \ldots, x_n$ についてもモデル1の最大尤度は離散分布モデルの場合に可能な最大の値 1 になる.

注意: 確率と違って確率密度の値はいくらでも大きくなりえるので, 尤度が確率密度の値になる連続分布モデルにおいては尤度はいくらでも大きな値を取り得る. 離散分布モデルの場合には尤度は確率の値になるのでその最大値は 1 になる.

注意: この例はオーバーフィッティング(過剰適合)のシンプルな例になっている. パラメータの個数が多いモデルではオーバーフィッティングし易くなる.

注意: この例を知っていれば, 「尤度が高ければもっともらしい」と考えることは単純に誤りであることがわかる. 尤度を常識的な意味でのもっともらしさと繋げるためには特別な条件が必要になる.

数学が強くなりたい人向けの自習問題

問題: 正規分布の標本分布モデルの尤度函数と最尤法

データは n 個の実数達 $x_1, \ldots, x_n$ であるとし, 統計モデルとして平均 μ, 分散 σ^2 の正規分布のサイズ n の標本分布を考える. その統計モデルの確率密度函数 $p(x_1, \ldots, x_n | \mu, \sigma^2)$ は,

$$p(x_i | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right)$$

とおくとき次のように表される:

$$p(x_1, \ldots, x_n | \mu, \sigma^2) = \prod_{i=1}^n p(x_i | \mu, \sigma^2).$$

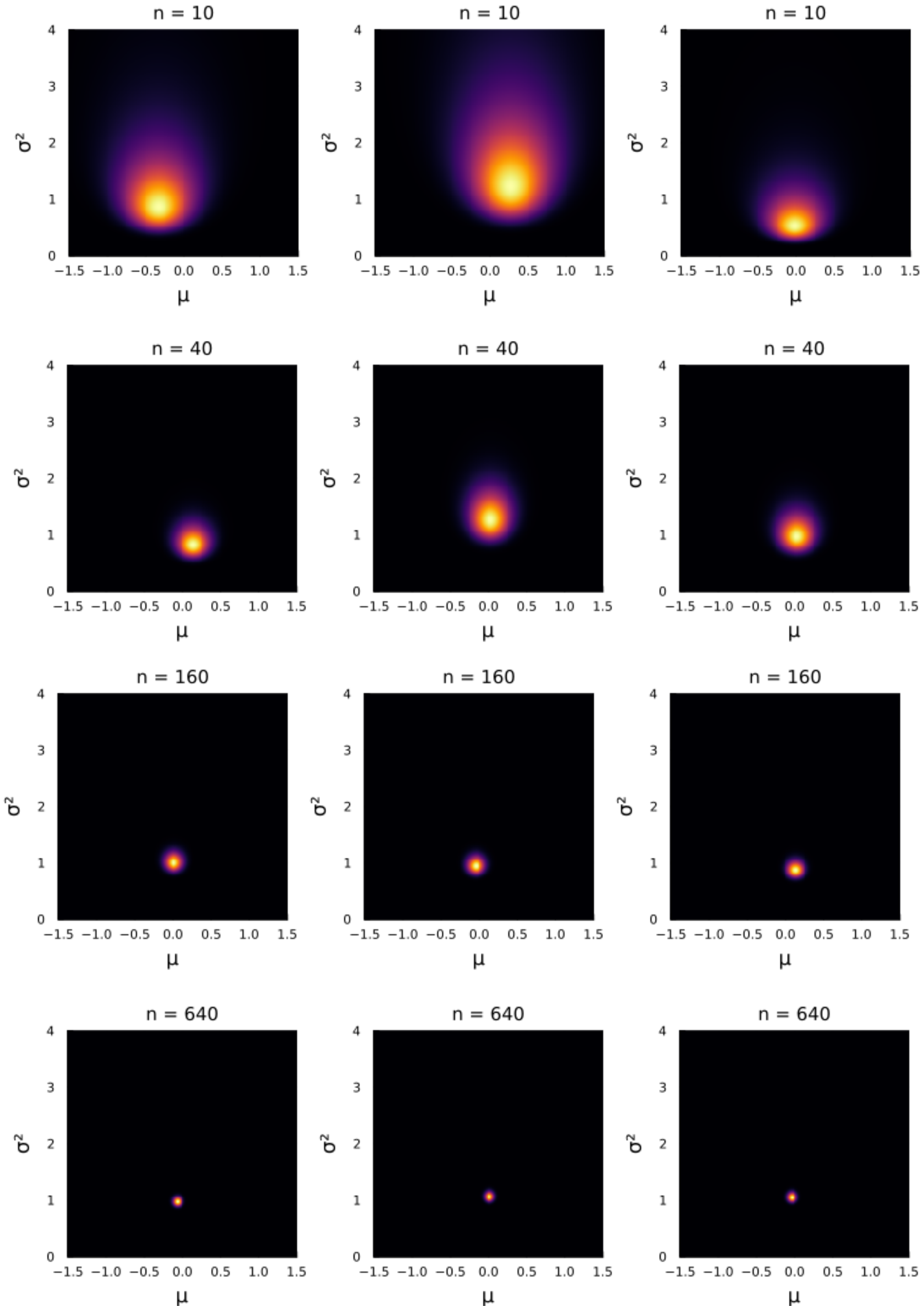
データ $x_1, \ldots, x_n$ に関するこのモデルの尤度函数を最大化する $\mu = \hat{\mu}, \sigma^2 = \hat{\sigma}^2$ はそれぞれ次になることを示せ:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \hat{\mu})^2.$$

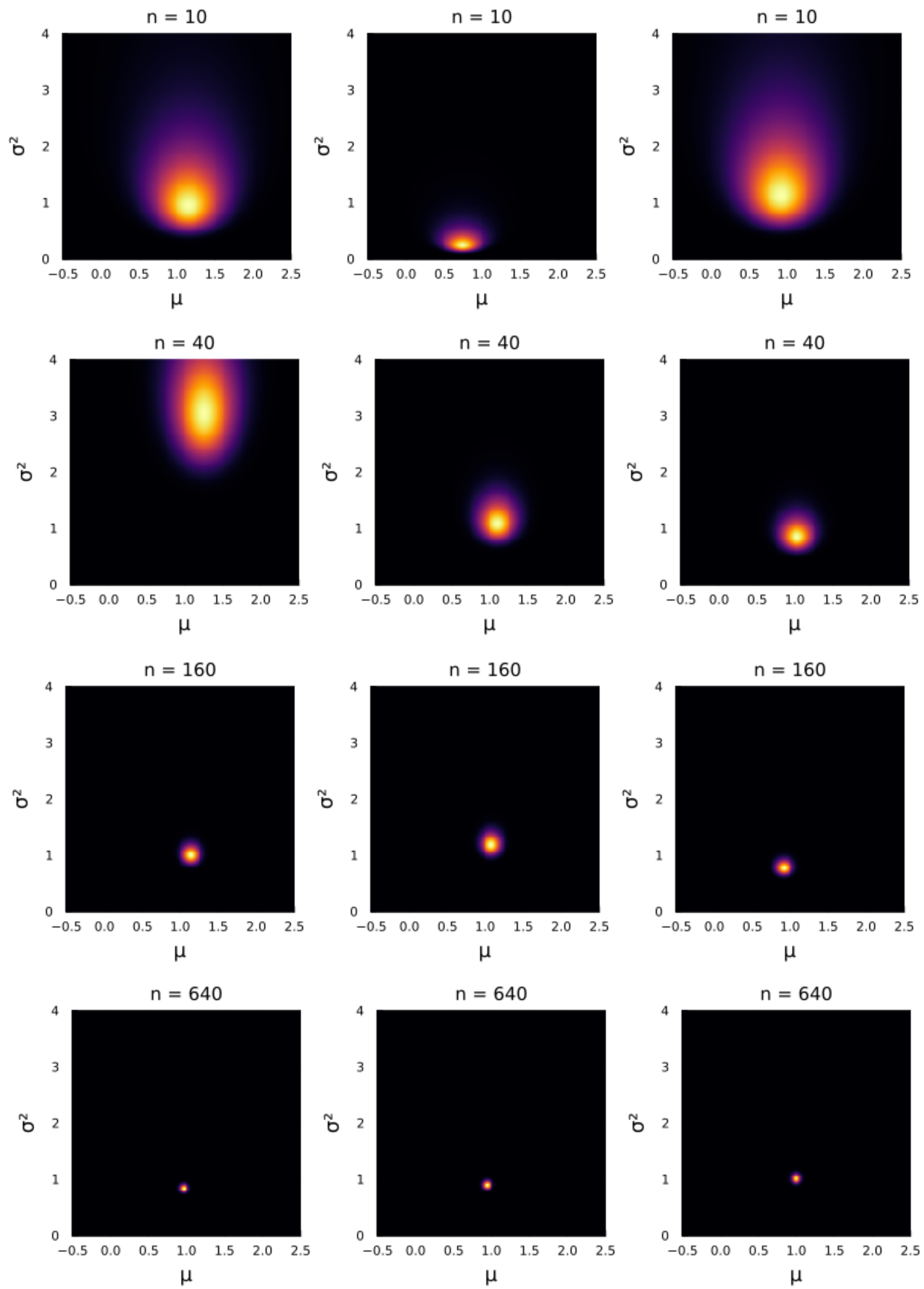
これは, 正規分布の標本分布モデルの最尤法の解が標本平均と(不偏にするための補正無しの)標本分散になることを意味している.

注意: この結果はデータの標本平均と標本分散を求めることは, 正規分布の標本分布モデルがデータに最もよくフィットするような μ と σ^2 を求めることに等しいことがわかる. この点は次の節の要約統計の説明でも触れる.

データをNormal(0,1)で生成　平均0　分散1



データをExponential(1)で生成　平均1　分散1



データの要約の仕方の注意

n 個の 1 と 0 からなるデータ $x_1, \dots, x_n$ の要約

n 個の 1 と 0 からなるデータ $x_1, \dots, x_n$ の要約として, $x_1, \dots, x_n$ の中の 1 の個数 $k = x_1 + \cdots + x_n$ がよく使われる. 報告の形式は「n 個中 k 個が 1 である」になる. その k はBernoulli試行モデルの十分統計量になっているのであった.

そのようなデータの要約の仕方は $x_1, \dots, x_n$ が並んでいる順序が重要な場合には適さない.

「n 個中の k 個が 1 である」と報告するときには n と k の両方の数値を報告することが重要である. なぜならばデータのサイズ n は統計分析の精度を見積もる上で非常に重要な情報だからである.

n の情報を省いて「1 の割合は 30% であった」のように k/n の情報だけを伝えることは好ましくない.

より一般の離散的なデータの要約

n 個の $1, 2, \dots, r$ からなるデータ $x_1, \dots, x_n$ の要約として, $x_1, \dots, x_n$ の中含まれる $a = 1, 2, \dots, r$ の個数 k_a がまとめた $(k_1, \dots, k_r)$ がよく使われる. ($(k_1, \dots, k_r)$ はカテゴリカル分布の標本分布モデルの十分統計量になっている.)

そのようなデータの要約の仕方は $x_1, \dots, x_n$ が並んでいる順序が重要な場合には適さない.

データの要約(summary)を報告する場合には, 必ずデータのサイズがわかるように報告しなければいけない. なぜならばデータのサイズ n は統計分析の精度を見積もる上で非常に重要な情報だからである.

例えば「$1, 2, 3, 4$ の割合がそれぞれ $10\%, 20\%, 30\%, 40\%$ であった」のように k_a/n の情報だけを伝えることは好ましくない. 全体の割合の情報だけではなく, 全体の個数の情報も必ず報告しなければいけない.

n 個の実数からなるデータ $x_1, \ldots, x_n$ の要約

n 個の実数からなるデータ $x_1, \ldots, x_n$ の要約としては以下がよく使われている:

- 標本平均 $\bar{x}$ と不偏分散 s^2
- 中央値
- 四分位数

標本平均と不偏分散の組み合わせは正規分布の標本分布モデルの十分統計量になっており, それだけではなく, その最尤法の解にもなっているのであった.

標本平均だけではなく, 中央値の情報も報告されているとき, それらに大きなずれが存在するならば, 分布に非対称性があることもわかる.

四分位数は外れ値がある場合の分布の散らばり方の要約として頑健である.

以上で挙げた, 標本平均, 不偏分散, 中央値, 四分位数はデータの分布が単峰型の場合にはデータの要約の仕方として適切だが, 山が2以上ある分布の場合には適さない.

この型のデータの様子は, ヒストグラムや経験累積分布函数(x に x 以下の x_i 達の割合を対応させる函数)をプロットして, **データ全体の様子を視覚的に必ず確認した方がよい.**

これが最も重要

n 個の実数の対からなるデータ $(x_1, y_1), \ldots, (x_n, y_n)$ の要約

n 個の実数の対からなるデータ $(x_1, y_1), \ldots, (x_n, y_n)$ の要約としてよく使われているのは,

- 標本平均 $\bar{x}, \bar{y}$ と不偏分散と不偏共分散 s_x^2, s_y^2, s_{xy}

である． この5つの量を合わせたものは二変量正規分布の標本分布モデルの十分統計量になっている.

それらの統計量を使って線形回帰の結果を表示することもできる. 線形回帰もデータの要約の仕方だとみなされる.

このようなデータを取得する場合には x_i と y_i のあいだの関係がどうなっているかに興味がある.

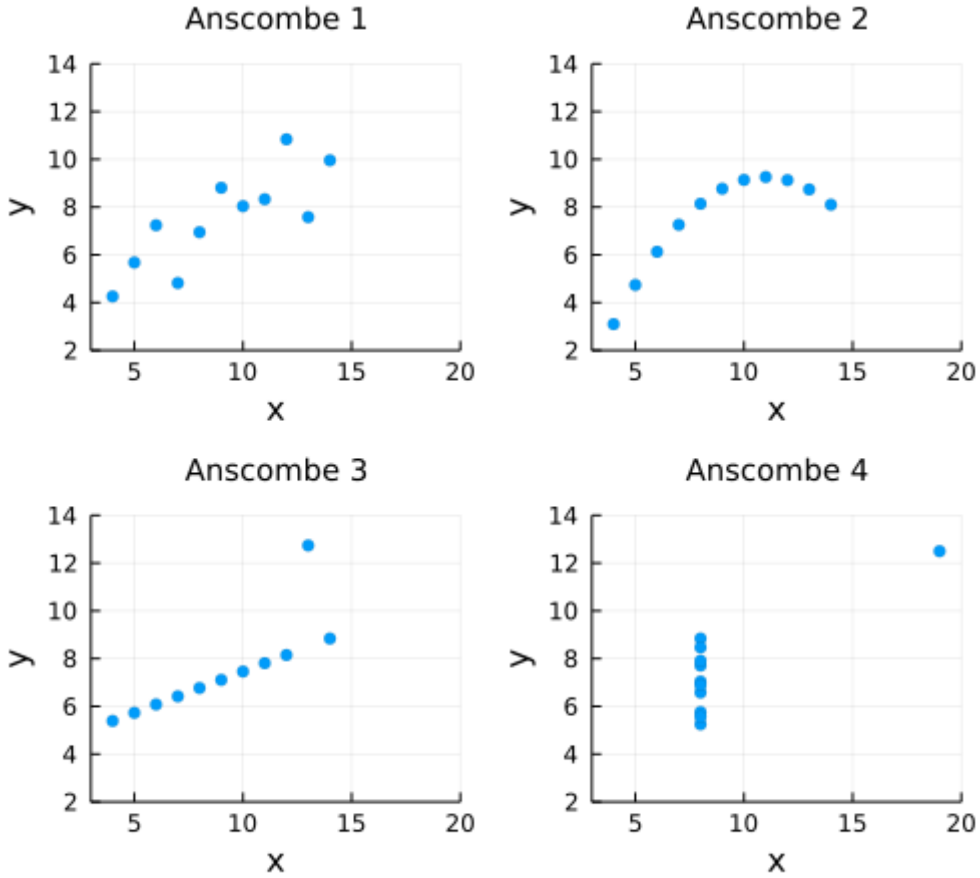
そのときに上の5つの要約統計量(や線形回帰の結果)だけしか見ないと, x_i と y_i のあいだの重要な関係を見逃してしまうことがある.

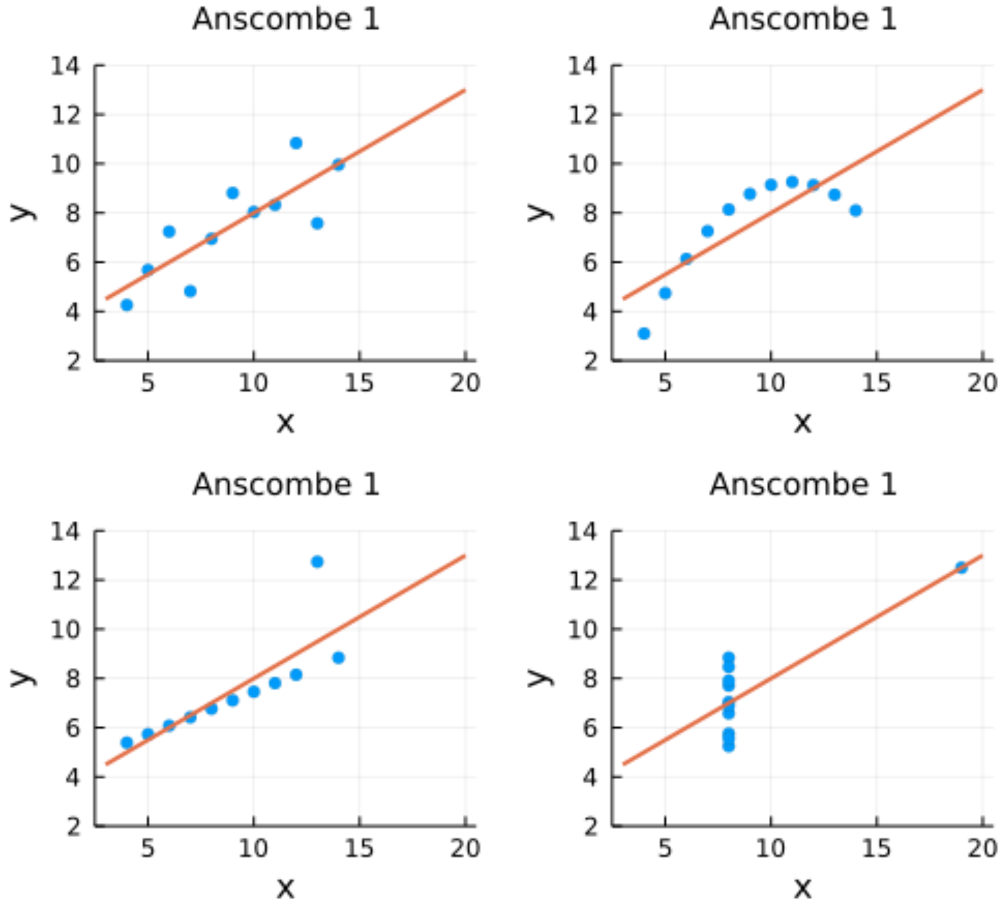
この型のデータを扱う場合には, **散布図** (scatter plot)を書くなどして、**データ全体の様子を視覚的に必ず確認した方がよい.**

これが最重要, 次ページの例を見よ!

Anscombeの例(アンスコムの例, 再)

以下の散布図(scatter plot)を見れば, 要約統計量だけを見てはいけないことがわかる.





散布図 (scatter plot)
非常によく使う!

もしも回帰直線の情報だけしか
見ていないと大変なことになってしまう.

重要必修自習問題

GitHubのサポートページからデータをダウンロードできる

https://github.com/genkuroki/Statistics/tree/master/2022 の第06節からデータをダウンロード可能

問題: DataSaurusの例

詳細な答えがGitHubにある資料にある。

以下のデータは

* http://www.thefunctionalart.com/2016/08/download-datasaurus-never-trust-summary.html

から取得したものである.

* CSVファイル
* x座標のみ, y座標のみ
* x座標のみコンマ付き, y座標のみコンマ付き

以下のデータの左列はx座標のデータであり, 右列はy座標のデータである.

(1) 標本平均 $\bar{x}, \bar{y}$ と不偏分散と不偏共分散 s_x^2, s_y^2, s_{xy} を小数点以下2桁以上求めよ.

(2) 何らかの方法で散布図(scatter plot)を描け.

さらに datasaurus same stats についてインターネットで検索して統計分析で注意するべきことを学べるサイトを見つけて内容を理解するように努力せよ.

ヒント: (1),(2)についてはインターネット上で適切に検索すれば以上の問題を解くことができるサイトを発見できる.