

中心極限定理などについて

- 黒木玄
- 2022-04-11~2022-05-08

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

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

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

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

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

目次

- ▼ [1 大数の法則](#)
 - [1.1 Markovの不等式](#)
 - [1.2 Chebyshevの不等式](#)
 - [1.3 大数の弱法則](#)
 - [1.4 不偏分散の \$n \rightarrow \infty\$ での挙動](#)
 - ▼ [1.5 大数の強法則](#)
 - [1.5.1 Borel-Cantelliの補題](#)
 - [1.5.2 4次以下のモーメントの存在を仮定した場合の大数の強法則](#)
 - ▼ [1.6 大数の法則のデモンストレーション](#)
 - [1.6.1 Bernoulli試行](#)
 - [1.6.2 正規分布のサンプルの場合](#)
 - [1.6.3 ガンマ分布の場合](#)
 - [1.6.4 大数の法則が成立しない場合 \(Cauchy分布\)](#)
- ▼ [2 二項分布の中心極限定理](#)
 - [2.1 二項分布の中心極限定理の内容](#)
 - [2.2 二項分布の中心極限定理の証明の方針](#)
 - [2.3 \(1\) Stirlingの公式を使った二項分布の確率密度関数の近似](#)
 - [2.4 注意: Kullback-Leibler情報量とSanovの定理との関係](#)
 - [2.5 \(2\) 二項分布の確率質量関数から正規分布の密度関数が出て来ること](#)
 - [2.6 問題: 二項分布の中心極限定理を使った簡単な計算問題](#)
- ▼ [3 中心極限定理](#)
 - [3.1 中心極限定理のラフな説明](#)
 - [3.2 中心極限定理の特性関数を使った証明](#)
 - [3.3 中心極限定理の収束の速さと歪度](#)
 - [3.4 中心極限定理のキュムラント母関数を使った証明](#)
 - [3.5 中心極限定理の収束の速さと歪度と尖度](#)
 - ▼ [3.6 中心極限定理のTaylorの定理のみを使う証明](#)
 - [3.6.1 多重積分の書き方](#)
 - [3.6.2 積分剰余項型のTaylorの定理](#)
 - [3.6.3 積分剰余項の別の表示](#)
 - [3.6.4 微分剰余項型のTaylorの定理](#)
 - [3.6.5 Taylorの定理を使った中心極限定理の証明](#)
 - [3.6.6 中心極限定理の収束の速さと歪度と尖度\(再\)](#)
 - [3.7 問題: 中心極限定理の収束の様子のグラフ](#)
 - [3.8 問題: デルタ法 \(実は単なる一次近似\)](#)

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

```
In [2]: 1 # Override the Base.show definition of SymPy.jl:
2 # https://github.com/JuliaPy/SymPy.jl/blob/29c5bfd1d10ac53014fa7fef468bc8deccadc2fc/src/types.
3
4 @eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::SymbolicObject)
5     print(io, as_markdown("\displaystyle " * sympy.latex(x, mode="plain", fold_short_frac=false)))
6 end
7 @eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::AbstractArray{Sym})
8     function toeqnarray(x::Vector{Sym})
9         a = join(["\displaystyle " * sympy.latex(x[i]) for i in 1:length(x)], "\\")
10        "\left[ \begin{array}{r}$a\end{array} \right]"
11    end
12    function toeqnarray(x::AbstractArray{Sym,2})
13        sz = size(x)
14        a = join([join(["\displaystyle " * map(sympy.latex, x[i,:]), "&") for i in 1:sz[1]],
15        "\left[ \begin{array}{r}" * repeat("r",sz[2]) * "}" * a * "\end{array}\right]"
16    end
17    print(io, as_markdown(toeqnarray(x)))
18 end
```

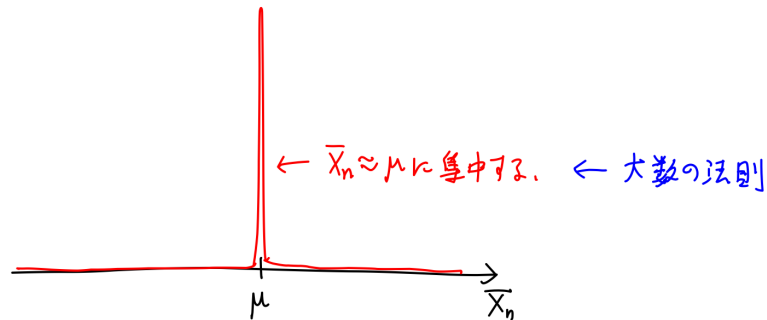
大数の法則と中心極限定理のイメージ

← $\mu = E[X_i], \sigma = \sqrt{E[(X_i - \mu)^2]}$ とおく.

X_1, X_2, X_3, \dots は独立同分布確率変数列 であるとし, $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ とおく.

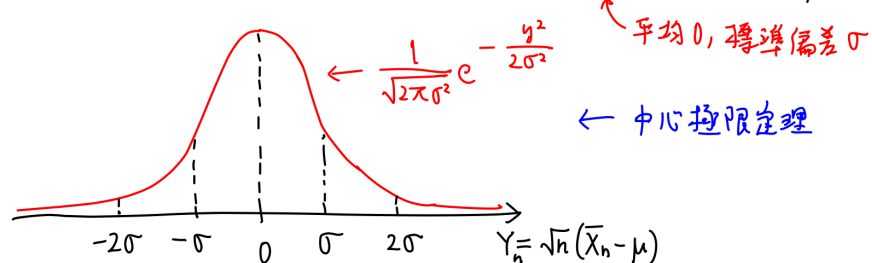
サイズ n の標本の平均 \bar{X}_n の分布の $n \rightarrow \infty$ での様子:

← 平均 μ , 標準偏差 $\frac{\sigma}{\sqrt{n}} \rightarrow 0$



$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ の分布は $n \rightarrow \infty$ で $\mu = E[X_i]$ に集中する.

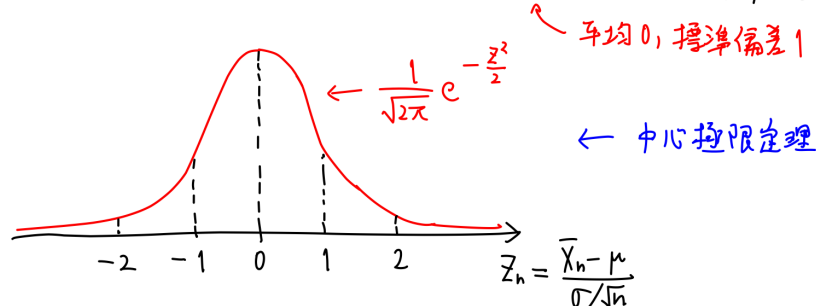
\bar{X}_n と μ の差を \sqrt{n} 倍して拡大して見るとこうなる: $Y_n = \sqrt{n}(\bar{X}_n - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu)$.



$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ と μ の差の大きさはだんだん $\frac{\sigma}{\sqrt{n}}$ の大きさであり,

差を \sqrt{n} 倍拡大して分布を見ると標準偏差 σ の正規分布になっている.

さらに σ 分の 1 倍するとこうなる: $Z_n = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}$



以上のように, 大数の法則により 1 点に集中する標本平均の分布を \sqrt{n} 倍 (もしくは σ/\sqrt{n} 分の 1 倍) することにより精密に見ると中心極限定理が得られる.

1 大数の法則

大数の法則 (law of large numbers, LLN) とは, 大雑把に言えば, 独立同分布確率変数列 X_1, X_2, \dots の最初の n 個の標本平均を $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ と書くと, \bar{X}_n が X_i 達の共通の期待値に収束するという結果のことである. (より正確に言うと, その「収束」の仕方に強弱があって, それぞれについて「大数の強法則」と呼んだり, 「大数の弱法則」と呼んだりする. しかし, 応用上, 多くの場合にその違いに神経質になる必要はない.)

以下では大数の法則について解説する.

1.1 Markovの不等式

確率変数 X と任意の $a > 0$ について

$$P(|X| \geq a) \leq \frac{1}{a} E[|X|].$$

証明: 関数 $1_{|x| \geq a}(x)$ を

$$1_{|x| \geq a}(x) = \begin{cases} 1 & (|x| \geq a) \\ 0 & (|x| < a) \end{cases}$$

とおくと, $|x| \geq a$ と $1 \leq |x|/a$ は同値なので常に $1_{|x| \geq a}(x) \leq |x|/a$ となる. ゆえに期待値を取る操作の単調性と線形性より,

$$P(|X| \geq a) = E[1_{|x| \geq a}(x)] \leq E\left[\frac{|X|}{a}\right] = \frac{1}{a} E[|X|].$$

証明終

注意: Markovの不等式, Chebyshevの不等式, Jensenの不等式, Gibbsの情報不等式などについては「Bernoulli試行と関連確率分布」のノートですでに説明していたのであった.

1.2 Chebyshevの不等式

Markovの不等式で X を $(X - \mu)^2$ ($\mu = E[X]$) で置き換えた場合をChebyshevの不等式と呼ぶ.

確率変数 X について $E[|X|] < \infty$ が成立しているならば, X の期待値 $\mu = E[X]$ がうまく定義されている. そのとき, X の分散を $\sigma^2 = E[(X - \mu)^2]$ と書くと, 任意の $\varepsilon > 0$ について

$$P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

証明: 前節のMarkovの不等式を X , a がそれぞれ $(X - \mu)^2$, ε^2 の場合に適用すると,

$$P(|X - \mu| \geq \varepsilon) = P((X - \mu)^2 \geq \varepsilon^2) \leq \frac{1}{\varepsilon^2} E[(X - \mu)^2] = \frac{\sigma^2}{\varepsilon^2}.$$

証明終

1.3 大数の弱法則

X_1, X_2, X_3, \dots は独立同分布確率変数の列であるとし, 共通の期待値 $\mu = E[X_i]$ が定義されており, 共通の分散 $\sigma^2 = E[(X_i - \mu)^2]$ は有限の値であると仮定する. このとき, 最初の n 個の標本平均を

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

と書くと, 任意の $\varepsilon > 0$ について,

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0.$$

注意: この結論は

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1$$

と同値なので標本平均 \bar{X}_n の分布が $n \rightarrow \infty$ で μ の近く(任意に小さな $\varepsilon > 0$ に関する μ から距離 ε 未満の範囲)に集中して行くことを意味している. このとき \bar{X}_n は μ に **確率収束** するというが, 標本平均 \bar{X}_n の分布が $n \rightarrow \infty$ で μ の近くに集中して行くことを理解していれば, 確率変数の収束に関する細かいことは応用上あまり神経質になる必要はない.

証明: 標本平均 \bar{X}_n の期待値と分散は以下になるのであった:

$$\begin{aligned}
E[\bar{X}_n] &= \frac{1}{n} \sum_{i=1}^n E[X_i] = \frac{1}{n} \sum_{i=1}^n \mu = \mu, \\
(\bar{X}_n - \mu)^2 &= \left(\frac{1}{n} \sum_{i=1}^n (X_i - \mu) \right)^2 = \frac{1}{n^2} \sum_{i,j=1}^n (X_i - \mu)(X_j - \mu), \\
\text{var}(\bar{X}_n) &= E[(\bar{X}_n - \mu)^2] = \frac{1}{n^2} \sum_{i,j=1}^n E[(X_i - \mu)(X_j - \mu)] = \frac{1}{n^2} \sum_{i,j=1}^n \delta_{ij} \sigma^2 = \frac{\sigma^2}{n}.
\end{aligned}$$

最後から2つめの等号で X_i 達が独立(特に無相関)であることを使った。ゆえに, Chebyshevの不等式を X, μ, σ^2 がそれぞれ $\bar{X}_n, \mu, \text{var}(\bar{X}_n) = \sigma^2/n$ の場合に適用すると, 任意の $\varepsilon > 0$ について,

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2/n}{\varepsilon^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

証明終

1.4 不偏分散の $n \rightarrow \infty$ 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, $\sigma^2 = E[(X_i - \mu)^2] < \infty$, $E[|X_i - \mu|^3] < \infty$, $E[(X_i - \mu)^4] < \infty$ となつていると仮定する。

「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,

$$\begin{aligned}
E[\bar{X}_n] &= \mu, \quad E[S_n^2] = \sigma^2, \\
\text{var}(\bar{X}_n) &= \frac{\sigma^2}{n}, \quad \text{cov}(\bar{X}_n, S_n^2) = \sigma^3 \frac{\bar{\kappa}_3}{n}, \quad \text{var}(S_n^2) = \sigma^4 \left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1} \right).
\end{aligned}$$

ゆえに, Chebyshevの不等式を X, μ, σ^2 がそれぞれ $S_n^2, E[S_n^2] = \sigma^2, \text{var}(S_n^2) = \sigma^4(\bar{\kappa}_4/n + 2/(n-1))$ の場合に適用すると, 任意の $\varepsilon > 0$ について,

$$P(|S_n^2 - \sigma^2| \geq \varepsilon) \leq \frac{\sigma^4}{\varepsilon^2} \left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1} \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

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 $n \rightarrow \infty$ で σ^2 の近くに集中して行く。

1.5 大数の強法則

この節の内容は無理して読まなくてもよい。定義していない用語や記号法が登場する。説明も少々雑になるだろう。

1.5.1 Borel-Cantelliの補題

確率空間において次が成立する:

$$\sum_{k=1}^{\infty} P(A_k) < \infty \implies P\left(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k\right) = 0.$$

証明: $\sum_{k=1}^{\infty} P(A_k) < \infty$ と仮定し, $B_n = \bigcup_{k \geq n} A_k$ とおく。 $P\left(\bigcap_{n=1}^{\infty} B_n\right) = 0$ を示せばよい。 $\sum_{k=1}^{\infty} P(A_k) < \infty$ より,

$$P(B_n) \leq \sum_{k \geq n} P(A_k) = \sum_{k=1}^{\infty} P(A_k) - \sum_{k=1}^{n-1} P(A_k) \rightarrow 0 \text{ as } n \rightarrow \infty.$$

$B_1 \supset B_2 \supset B_3 \supset \dots$ なので

$$P\left(\bigcap_{n=1}^{\infty} B_n\right) = \lim_{n \rightarrow \infty} P(B_n) = 0.$$

証明終

1.5.2 4次以下のモーメントの存在を仮定した場合の大数の強法則

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 個の標本平均を $\bar{X}_n = (1/n) \sum_{i=1}^n X_i$ と書くことにする。さらに、 $\mu = E[X_i]$ がうまく定義されており、 $\sigma^2 = E[(X_i - \mu)^2] < \infty$, $\mu_4 = E[(X_i - \mu)^4] < \infty$ となっていると仮定する。このとき、

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1.$$

注意: これは標本平均が μ に確率 1 で収束することを意味している。

証明: 大数の弱法則の場合と同様に補事象の側の同値な主張を示す。すなわち、標本平均が μ に収束しない確率が 0 であることを示す。標本平均が μ に収束しないという条件は以下のように書き換えられる：

$\bar{X}_k(\omega)$ は $k \rightarrow \infty$ で μ に収束しない

\Leftrightarrow ある $\varepsilon > 0$ が存在して、 $|\bar{X}_k(\omega) - \mu| \geq \varepsilon$ となる k が無限個ある

\Leftrightarrow ある $\varepsilon > 0$ が存在して、 $\omega \in \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} \{ \omega \mid |\bar{X}_k(\omega) - \mu| \geq \varepsilon \} = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k(\varepsilon)$

$\Leftrightarrow \omega \in \bigcup_{N=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k(1/N).$

ここで $A_k(\varepsilon) = \{ \omega \mid |\bar{X}_k(\omega) - \mu| \geq \varepsilon \}$ とおいた。任意の $\varepsilon > 0$ について $\sum_{k=1}^{\infty} P(A_k(\varepsilon)) < \infty$ を示せば、Borel-Cantelliの補題より $P\left(\bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k(\varepsilon)\right) = 0$ が得られ、それより

$$P\left(\bigcup_{N=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} A_k(1/N)\right) = 0$$

が得られ、 \bar{X}_k が $k \rightarrow \infty$ で μ に収束しない確率が 0 であることが示される。

必要ならば X_i を $X_i - \mu$ で置き換えることによって、 $\mu = 0$ と仮定しても一般性が失われないので、 $\mu = 0$ と仮定する。 $\varepsilon > 0$ を任意に取る。このとき、

$$P(|\bar{X}_n| \geq \varepsilon) = P\left(\frac{\bar{X}_n^4}{\varepsilon^4} \geq 1\right) = E\left[1_{\frac{\bar{X}_n^4}{\varepsilon^4} \geq 1}(\bar{X}_n)\right] \leq E\left[\frac{\bar{X}_n^4}{\varepsilon^4}\right] = \frac{1}{\varepsilon^4} E[\bar{X}_n^4],$$

でかつ

$$\begin{aligned} E[\bar{X}_n^4] &= \frac{1}{n^4} \sum_{i,j,k,l=1}^n E[X_i X_j X_k X_l] = \frac{1}{n^4} \left(n E[X_1^4] + \underbrace{\binom{4}{2} \binom{n}{2} E[X_1^2]^2}_{=3n(n-1)} \right) \\ &= \frac{\mu_4}{n^3} + \frac{3(n-1)\sigma^4}{n^3} \leq \frac{\mu_4 + 3\sigma^4}{n^2} \end{aligned}$$

なので、

$$\sum_{k=1}^{\infty} P(A_k) = \sum_{n=1}^{\infty} P(|\bar{X}_n| \geq \varepsilon) \leq \frac{\mu_4 + 3\sigma^4}{\varepsilon^4} \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

これで示すべきことが示された。

証明終

注意: $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$ はRiemannのゼータ函数と呼ばれ、 $s = 1$ では調和級数になって無限大に発散するが、 $s > 1$ ならば有限の値に収束する。 $\zeta(2) = \sum_{n=1}^{\infty} 1/n^2$ を求めよという問題は **Basel問題** (パーゼル問題) と呼ばれ、18世紀の超大大数学者のEulerが解いた。答えは $\pi^2/6$ になる。その逆数の $6/\pi^2 \approx 0.6079$ は2つの整数の互いに素になる(最大公約数が1になる)「確率」に等しい。

In [3]: 1 6/π^2

Out[3]: 0.6079271018540267

```

In [4]: 1 """
2 `IsPrime = make_isprime(N)` のとき
3 `IsPrime[p]` は `p` が素数なら真に素数でないなら偽になる。
4 """
5 function make_isprime(N)
6     IsPrime = trues(N)
7     # 1は素数ではない
8     IsPrime[1] = false
9     for p in 2:isqrt(N)
10         if IsPrime[p]
11             for i in 2p:p:N
12                 # 素数の非自明な倍数は素数ではない
13                 IsPrime[i] = false
14             end
15         end
16     end
17     IsPrime
18 end
19
20 """N以下の2つの正の整数が互いに素な確率"""
21 function prob_mutually_prime(N)
22     IsPrime = make_isprime(N)
23     C = trues(N, N)
24     for p in 2:N
25         if IsPrime[p]
26             for i in p:p:N
27                 for j in i:p:N
28                     # 共通の素因数 p を持つ i, j を除外する
29                     C[i, j] = false
30                     C[j, i] = false
31                 end
32             end
33         end
34     end
35     c = count(C)
36     c/N^2
37 end

```

Out[4]: prob_mutually_prime

```

In [5]: 1 primes = make_isprime(100) ▷ findall
2 primes'

```

Out[5]: 1×25 adjoint(::Vector{Int64}) with eltype Int64:
 2 3 5 7 11 13 17 19 23 29 31 37 41 43 47 53 59 61 67 71 73 79 83 89 97

```

In [6]: 1 @time prob_mutually_prime(10000)

0.166374 seconds (5 allocations: 11.922 MiB, 22.67% gc time)

```

Out[6]: 0.60794971

```

In [7]: 1 @time prob_mutually_prime(100000)

17.651889 seconds (5 allocations: 1.164 GiB, 0.08% gc time)

```

Out[7]: 0.6079301507

1.6 大数の法則のデモンストレーション

```

In [8]: 1 function plot_law_of_large_numbers(dist, N;  $\mu$  = mean(dist), kwargs...)
2         X = rand(dist, N)
3          $\bar{X}$  = cumsum(X) ./ (1:N)
4         plot(1:N,  $\bar{X}$ ; label="")
5         hline!([ $\mu$ ]; label="")
6         plot!(); kwargs...)
7     end
8
9     function plot_law_of_large_numbers_6x3(dist, N;
10          $\mu$  = mean(dist), seed=4649363, title="", kwargs...)
11         Random.seed!(seed)
12         PP = []
13         for _ in 1:18
14             P = plot_law_of_large_numbers(dist, N;  $\mu$ , kwargs...)
15             push!(PP, P)
16         end
17         P0 = plot(; title,
18             framestyle=nothing, showaxis=false, ticks=false, margin=0Plots.mm)
19         layout = @layout [a{0.005h}; grid(6, 3)]
20         plot(P0, PP...; size=(800, 800), layout,
21             titlefontsize=12, guidefontsize=8, tickfontsize=6)
22     end

```

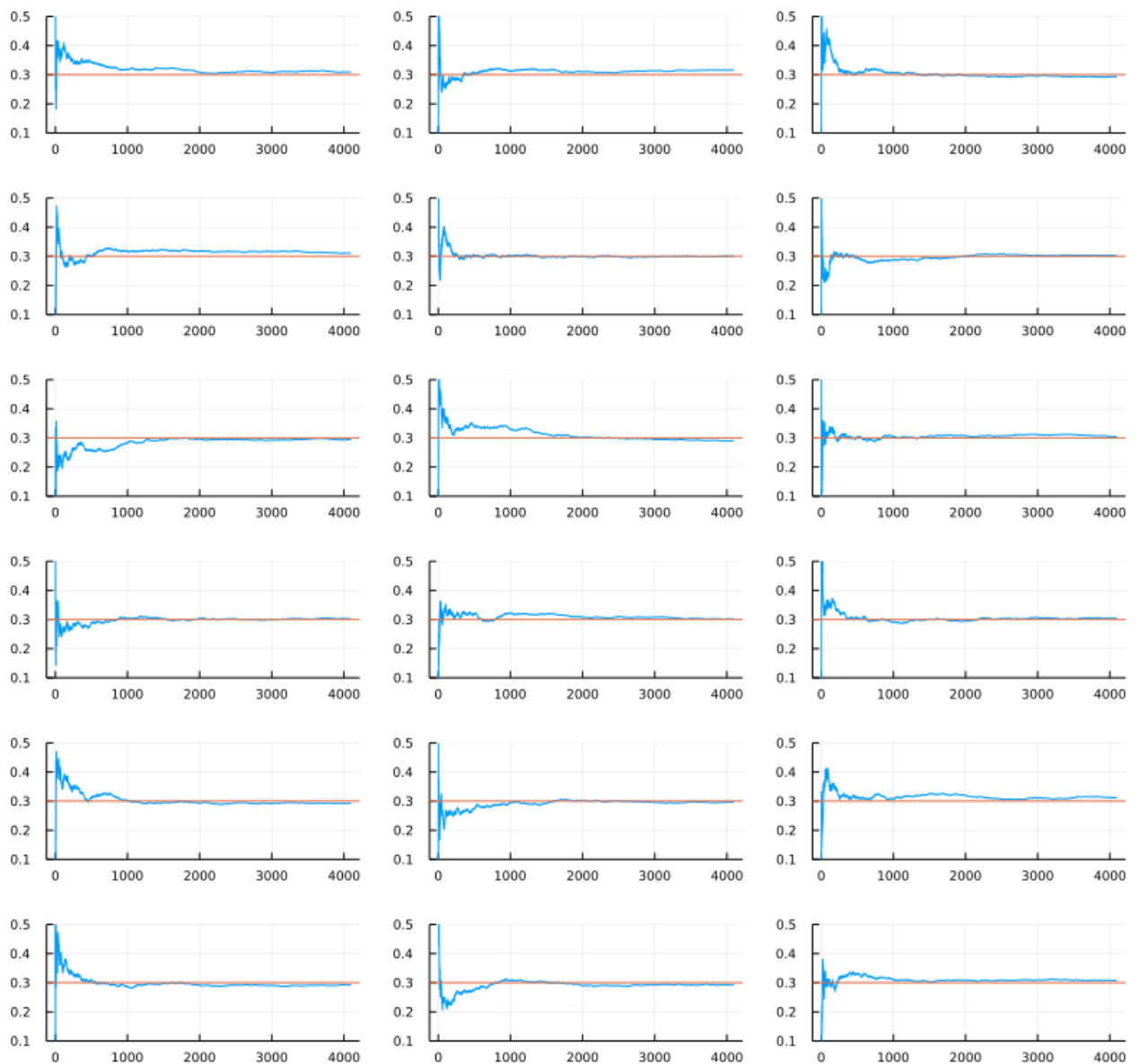
Out[8]: plot_law_of_large_numbers_6x3 (generic function with 1 method)

1.6.1 Bernoulli試行


```
In [9]: 1 plot_law_of_large_numbers_6x3(Bernoulli(0.3), 2^12;  
2      title="Law of large numbers for Bernoulli(0.3)", ylim=(0.1, 0.5))
```

Out[9]:

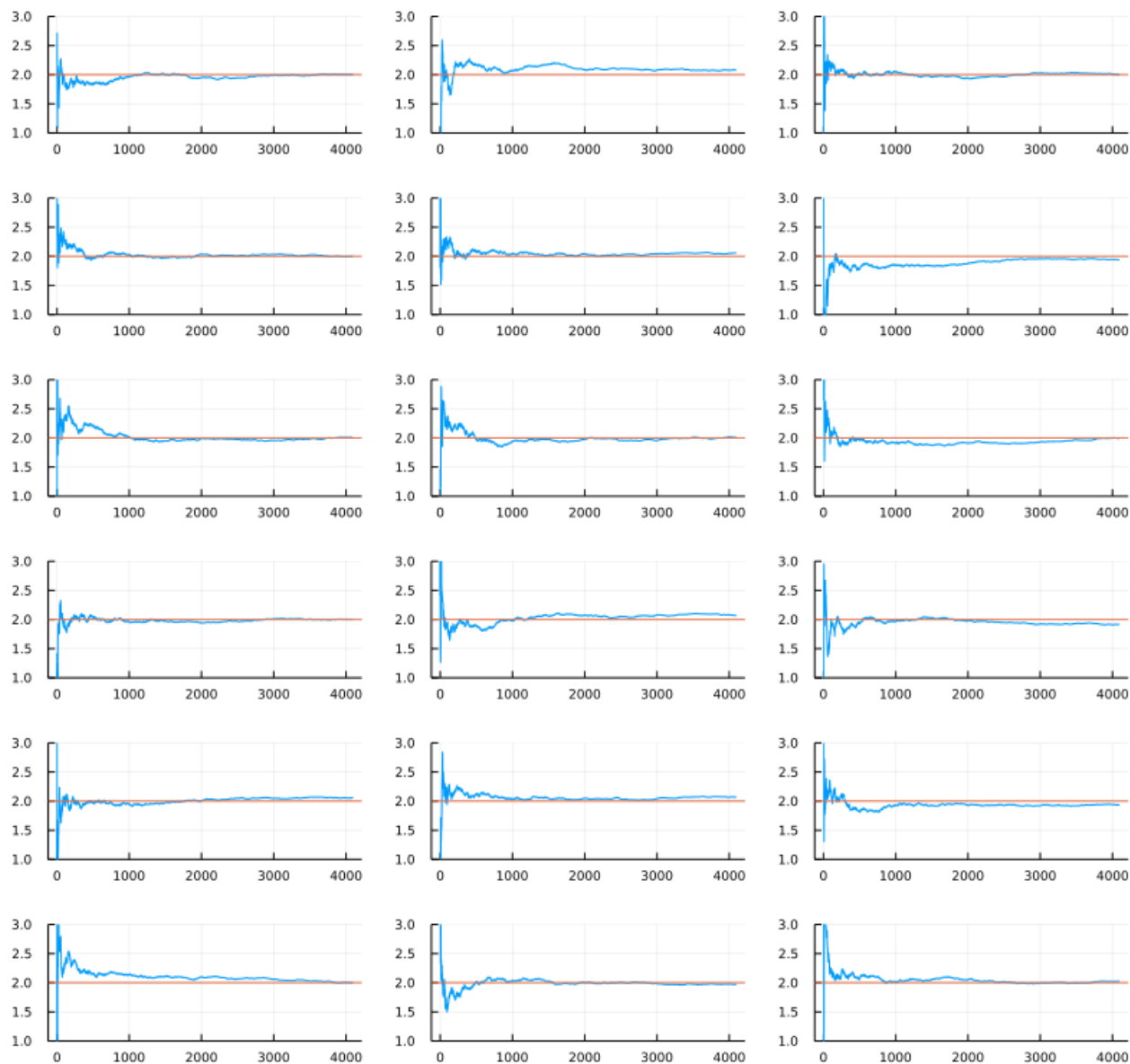
Law of large numbers for Bernoulli(0.3)



1.6.2 正規分布のサンプルの場合

```
In [10]: 1 plot_law_of_large_numbers_6x3(Normal(2, 3), 2^12;  
2      title="Law of large numbers for Normal(2, 3)", ylim=(1, 3))
```

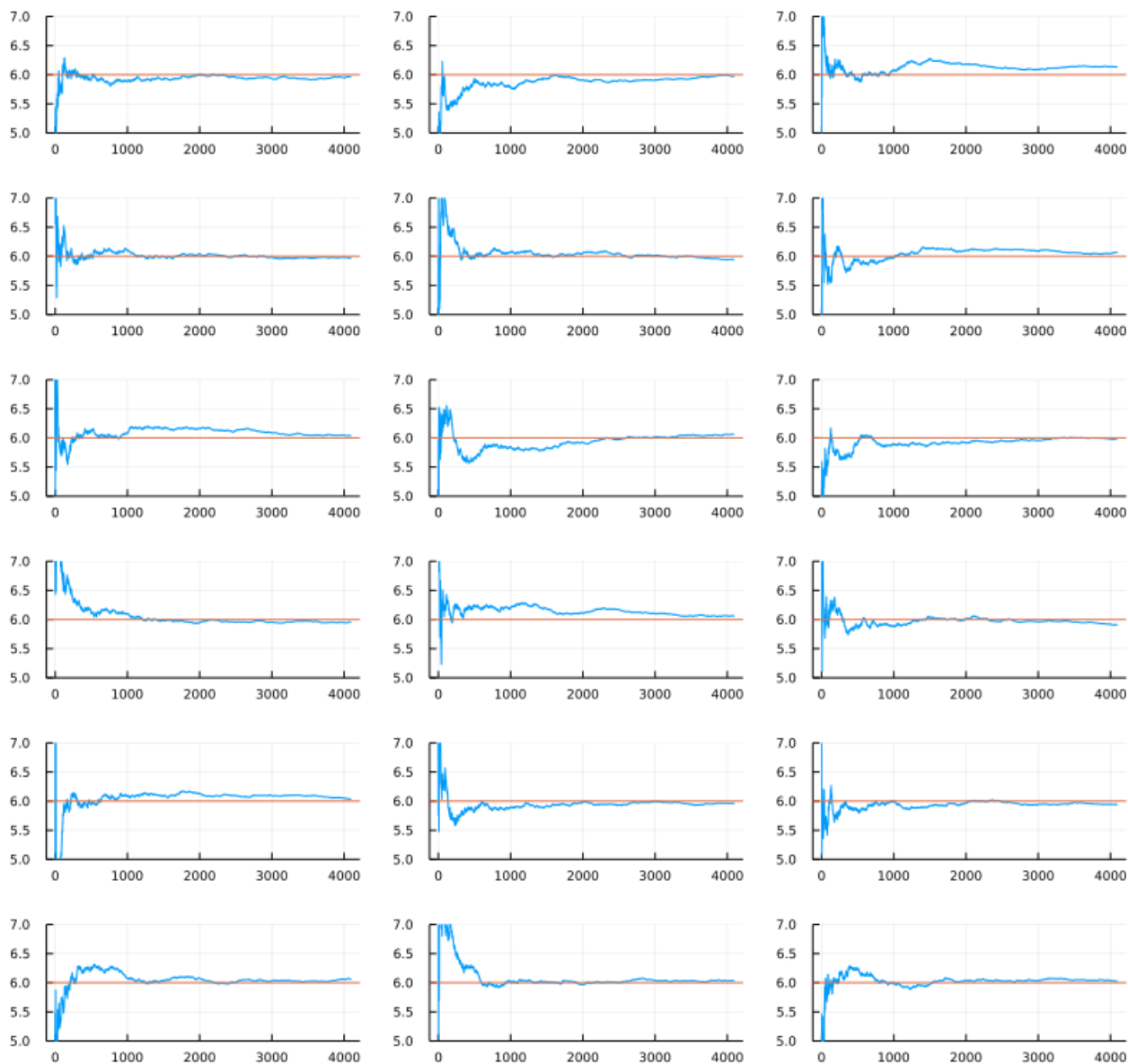
Out[10]: Law of large numbers for Normal(2, 3)



```
In [11]: 1 plot_law_of_large_numbers_6x3(Gamma(2, 3), 2^12;  
2      title="Law of large numbers for Gamma(2, 3)", ylim=(5, 7))
```

Out[11]:

Law of large numbers for Gamma(2, 3)

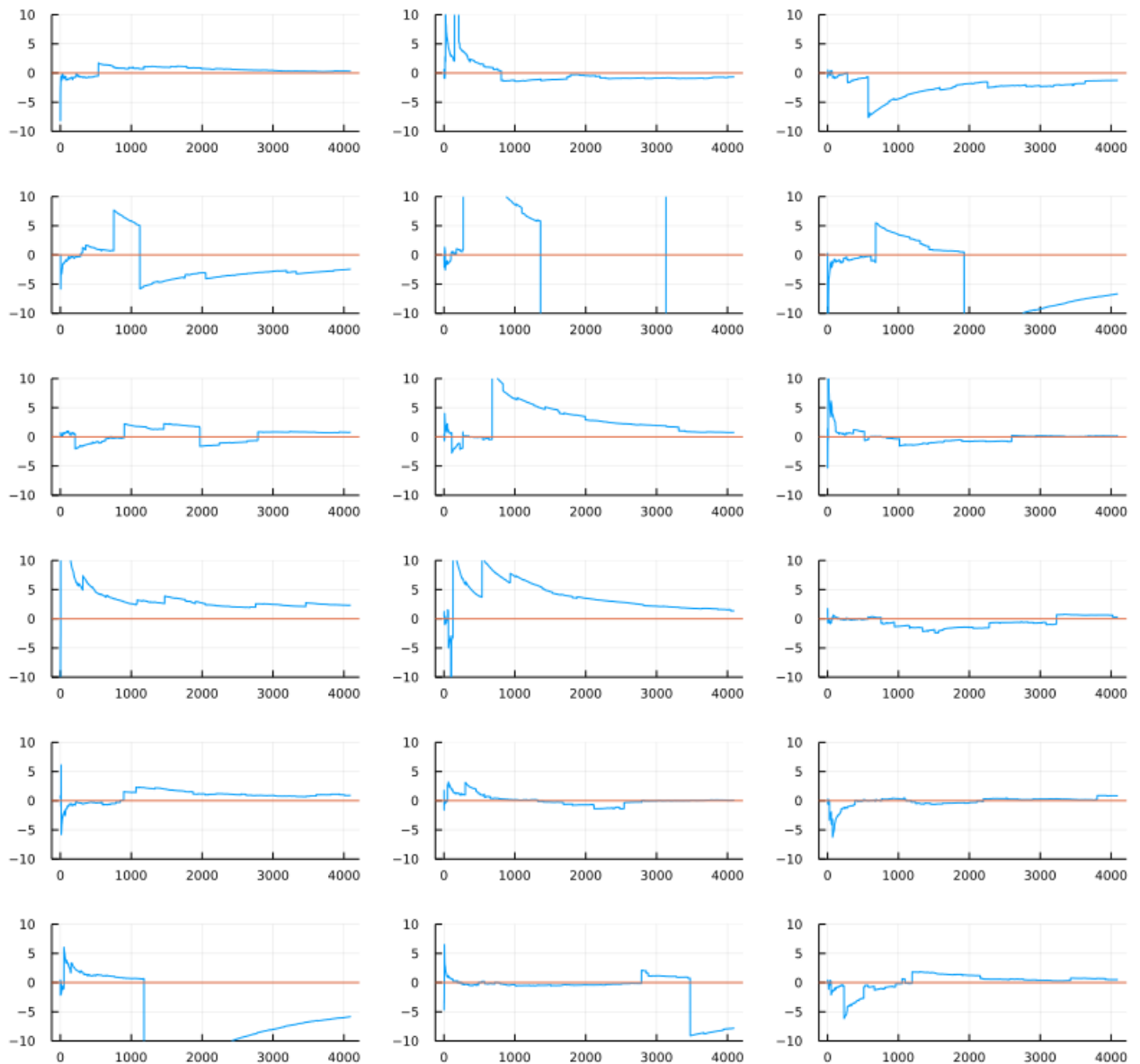


1.6.4 大数の法則が成立しない場合 (Cauchy分布)

```
In [12]: 1 plot_law_of_large_numbers_6x3(Cauchy(), 2^12;  $\mu = 0.0$ ,
2       title="Law of large numbers for Cauchy() does not hold", ylim=(-10, 10))
```

Out[12]:

Law of large numbers for Cauchy() does not hold



2 二項分布の中心極限定理

2.1 二項分布の中心極限定理の内容

二項分布 $\text{Binomial}(n, p)$ が正規分布 $\text{Normal}(np, \sqrt{np(1-p)})$ で近似されることを **二項分布の中心極限定理** と呼ぶことにしよう。(ただしその近似の精度をよくするためには np と $n(1-p)$ をある程度以上に大きくする必要がある。「大きくする」と言っても粗い近似でよければ3から5程度でも十分な場合がある.)

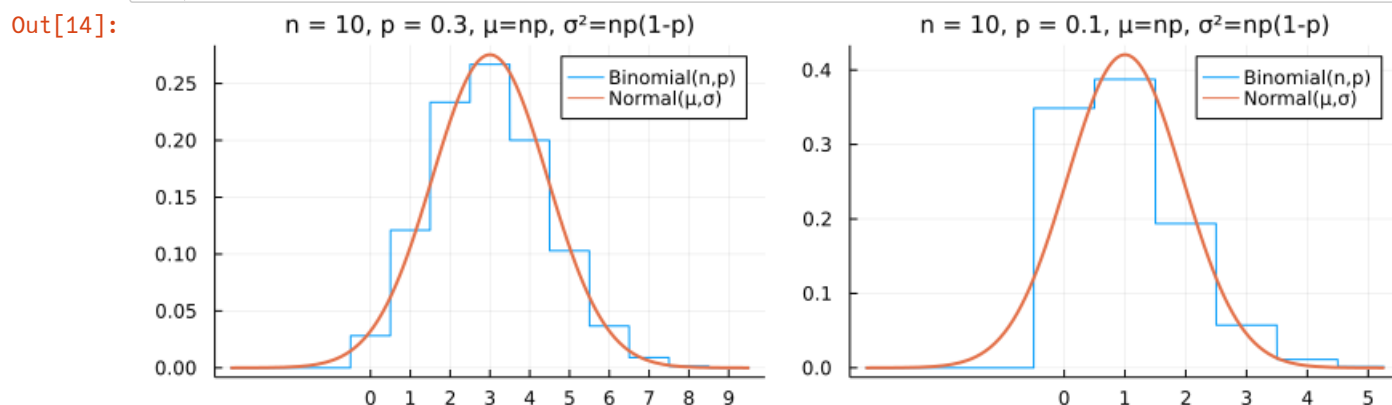
以下は、試行回数 n 、成功確率 p の二項分布(期待値と分散はそれぞれ np , $np(1-p)$ になる)の確率質量函数と平均 np 、分散 $np(1-p)$ を持つ正規分布の確率密度函数の同時プロットである。それらを見れば、証明しなくても、二項分布の中心極限定理が成立していることは明らかだろう。(証明する前にコンピュータでグラフを確認するべき!)

注意: 以下のグラフを見れば、二項分布の正規分布近似の精度を高めるためには、 p が小さいほど n を大きくする必要があることがわかる。 p が 1 に近い場合にも n を大きくする必要がある。

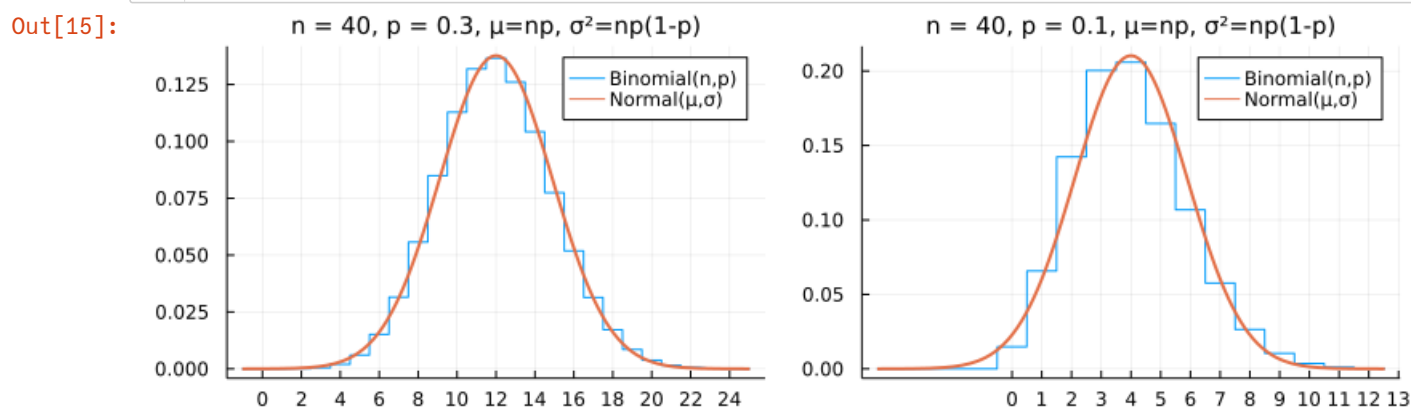
```
In [13]: 1 function plot_binomial_clt(n, p, s = 1; c = 4.5)
2         μ, σ = n*p, √(n*p*(1-p))
3         xlim = (μ-c*σ, μ+c*σ)
4         plot(x → pdf(Binomial(n, p), round(x)), xlim...; label="Binomial(n,p)")
5         plot!(Normal(μ, σ), xlim...; label="Normal(μ,σ)", lw=2)
6         # plot!(x → pdf(Poisson(μ), round(x)), xlim...; label="Poisson(μ)", ls=:dash)
7         title!("n = $n, p = $p, μ=np, σ²=np(1-p)")
8         plot!(); xtick=0:s:n
9     end
```

Out[13]: plot_binomial_clt (generic function with 2 methods)

```
In [14]: 1 plot(plot_binomial_clt(10, 0.3), plot_binomial_clt(10, 0.1); size=(800, 250))
```

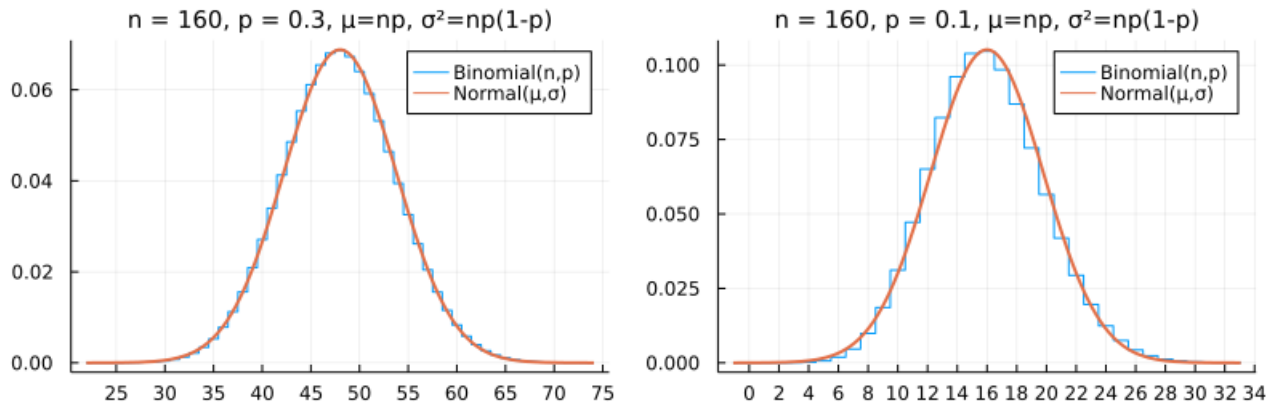


```
In [15]: 1 plot(plot_binomial_clt(40, 0.3, 2), plot_binomial_clt(40, 0.1); size=(800, 250))
```



```
In [16]: 1 plot(plot_binomial_clt(160, 0.3, 5), plot_binomial_clt(160, 0.1, 2); size=(800, 250))
```

Out[16]:



2.2 二項分布の中心極限定理の証明の方針

二項分布の中心極限定理を以下の2段階で示す.

記号法に関する準備: $0 < p < 1$ は固定するが, n, k は動かす. k は n ごとに決まっていると仮定する. 混乱を完全に防ぐには k を k_n のように書いた方がよいが, 記号が煩雑になって書くのが面倒になるので, 単に k と書くことにする. $n \rightarrow \infty$ で 0 に収束する量を $o(1)$ と書く. より一般に a_n で割った結果が $n \rightarrow \infty$ で 0 に収束する量を $o(a_n)$ と書く. 例えば $o(1/\sqrt{n})$ は \sqrt{n} をかけた後であっても $n \rightarrow \infty$ で 0 に収束する量になる. この記号法のもとで, Stirlingの公式は次のように書ける:

$$n! = n^n e^{-n} \sqrt{2\pi n} (1 + o(1)).$$

(1) n ごとに決まっている k は $n \rightarrow \infty$ で

$$\frac{k}{n} = p + o(1)$$

という条件を満たしていると仮定する. このとき, 二項分布の確率質量関数

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

における二項係数の中の $n!, k!, (n-k)!$ に階乗に関するStirlingの公式による近似を適用すると次の近似式が得られる:

$$P(k|n, p) = \frac{1}{\sqrt{2\pi np(1-p)}} \left(\left(\frac{k/n}{p} \right) \left(\frac{1-k/n}{1-p} \right) \right)^{-n} (1 + o(1)).$$

途中の段階としてこの形に整理しておくこと計算の見通しが良くなる. (しかもこの形には重要な意味がある. その点については後述する.)

(2) 固定された実数 x について, 上の(1)での近似式を示すために使った仮定 $k/n = p + o(1)$ よりも強い条件

$$\frac{k}{n} = p + \frac{x}{\sqrt{n}} + o(n^{-1/2})$$

を仮定すると, 次が得られる:

$$P(k|n, p) = \underbrace{\frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{x^2}{2p(1-p)}\right)}_{\text{pdf of normal distribution}} \underbrace{\frac{1}{\sqrt{n}}}_{dx/dk} (1 + o(1)).$$

これより,

$$K_n \sim \text{Binomial}(n, p), \quad X_n = \frac{K_n - np}{\sqrt{n}}, \quad X_\infty \sim \text{Normal}(0, p(1-p))$$

ならば

$$\lim_{n \rightarrow \infty} E[f(X_n)] \rightarrow E[f(X_\infty)]$$

となることがわかる.

この結果は, n が大きくなるとき, $K_n \sim \text{Binomial}(n, p)$ に対する $X_n = (K_n - np)/\sqrt{n}$ が従う分布が $\text{Normal}(0, p(1-p))$ で近似されることを意味している($p(1-p)$ は分布 $\text{Bernoulli}(p)$ の分散に等しい). この結果を二項分布の中心極限定理と呼ぶ.

2.3 (1) Stirlingの公式を使った二項分布の確率密度関数の近似

n ごとに決まっている k は $n \rightarrow \infty$ で

$$\frac{k}{n} = p + o(1) \quad \left(\Leftrightarrow 1 - \frac{k}{n} = 1 - p + o(1) \right)$$

という条件を満たしていると仮定する. $0 < p < 1$ と仮定していたので, $n \rightarrow \infty$ のとき $k \rightarrow \infty$ だけでなく, $n - k \rightarrow \infty$ も成立している. ゆえに $n, k, n - k$ について次のStirlingの公式が成立している:

$$\begin{aligned} n! &= n^n e^{-n} \sqrt{2\pi n} (1 + o(1)), \\ k! &= k^k e^{-k} \sqrt{2\pi k} (1 + o(1)), \\ (n - k)! &= (n - k)^{n-k} e^{-(n-k)} \sqrt{2\pi(n - k)} (1 + o(1)). \end{aligned}$$

これを代入すると

$$\begin{aligned} \frac{n!}{k!(n - k)!} &= \frac{n^n e^{-n} \sqrt{2\pi n}}{k^k e^{-k} \sqrt{2\pi k} (n - k)^{n-k} e^{-(n-k)} \sqrt{2\pi(n - k)}} (1 + o(1)) \\ &= \frac{1}{(k/n)^k (1 - k/n)^{n-k} \sqrt{2\pi n(k/n)(1 - k/n)}} (1 + o(1)) \\ &= \frac{1}{(k/n)^k (1 - k/n)^{n-k} \sqrt{2\pi n p(1 - p)}} (1 + o(1)) \end{aligned}$$

2つめの等号で分子の e^{-n} と分母の $e^{-k} e^{-(n-k)}$ がキャンセルさせ, 分子分母を n^k と n^{n-k} と $\sqrt{2\pi n}$ で割った. 3つめの等号で $k/n = p + o(1)$, $1 - k/n = 1 - p + o(1)$ を使った. これを二項分布の確率質量関数

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

に適用すると,

$$\begin{aligned} P(k|n, p) &= \frac{1}{\sqrt{2\pi n p(1 - p)}} \left(\frac{p}{k/n} \right)^k \left(\frac{1 - p}{1 - k/n} \right)^{n-k} (1 + o(1)) \\ &= \frac{1}{\sqrt{2\pi n p(1 - p)}} \left(\left(\frac{k/n}{p} \right)^{k/n} \left(\frac{1 - k/n}{1 - p} \right)^{1 - k/n} \right)^{-n} (1 + o(1)). \end{aligned}$$

2.4 注意: Kullback-Leibler情報量とSanovの定理との関係

前節でわざわざ最後の形に変形した理由は, 対数を取ると

$$\log \left(\left(\left(\frac{k/n}{p} \right)^{k/n} \left(\frac{1 - k/n}{1 - p} \right)^{1 - k/n} \right)^{-n} \right) = -n \left(\frac{k}{n} \log \frac{k/n}{p} + \left(1 - \frac{k}{n} \right) \log \frac{1 - k/n}{1 - p} \right)$$

となることを分かり易くするためである. これを $-n$ で割った結果は

$$D(Q||P) = \sum_{i=1}^r Q(i) \log \frac{Q(i)}{P(i)} \quad \left(P(i), Q(i) > 0, \sum_{i=1}^r P(i) = \sum_{i=1}^r Q(i) = 1 \right)$$

の特別な場合になっている($r = 2$, $P(1) = p$, $P(2) = 1 - p$, $Q(1) = k/n$, $Q(2) = 1 - k/n$). この $D(Q||P)$ は **Kullback-Leibler 情報量 (KL情報量)** と呼ばれ, そして, 上の計算結果は $r = 2$ の場合のKL情報量のSanovの証明にもなっている. Sanovの定理の易しい解説が以下の場所にある:

- [Kullback-Leibler情報量とSanovの定理 \(https://genkuroki.github.io/documents/20160616KullbackLeibler.pdf\)](https://genkuroki.github.io/documents/20160616KullbackLeibler.pdf)

Sanovの定理の内容を要約すると, Kullback-Leibler情報量 $D(Q||P)$ は確率分布 P に従う乱数の繰り返し生成において確率分布 Q がどれだけ出て来難いかを表しているという主張になる. 上の二項分布の場合には, Q に対応する k/n が生じる確率が $\exp(-nD(Q||P) + O(\log n))$ の形をしていることが示されていることになるので(分母の $\sqrt{2\pi np(1-p)}$ は $\exp(O(\log n))$ になる), KL情報量 $D(Q||P)$ が大きいほど k/n が生じる確率は急激に小さくなる.

Sanovの定理は統計学でのモデル選択における情報量規準の考え方の基礎になっている. KL情報量は確率論では **大偏差原理** に出て来る. KL情報量は物理的には統計力学的な相対エントロピーの -1 倍になる. このような分野を超えた広がりについては, **赤池情報量規準 (AIC)**で有名な赤池弘次氏による1980年の2つの論説も参照せよ:

- 赤池弘次, エントロピーとモデルの尤度(<講座>物理学周辺の確率統計), 日本物理学会誌, 1980年第35巻7号, pp. 608-614. [link \(https://www.jstage.jst.go.jp/article/butsuri1946/35/7/35_7_608/article-char/ja/\)](https://www.jstage.jst.go.jp/article/butsuri1946/35/7/35_7_608/article-char/ja/)
- 赤池弘次, 統計的推論のパラダイムの変遷について, 統計数理研究所彙報, 1980年第27巻第1号, pp. 5-12. [link \(https://ismrepo.ism.ac.jp/index.php?active_action=repository_view_main_item_detail&page_id=13&block_id=21&item_id=32568&item_no=1\)](https://ismrepo.ism.ac.jp/index.php?active_action=repository_view_main_item_detail&page_id=13&block_id=21&item_id=32568&item_no=1)

赤池弘次さんは前者のp.612に「筆者によって導入されたこの統計量はAIC(an information criterionの略記*)と呼ばれ」と書いているが, それに対して会誌編集委員会が「一般には Akaike's information criterionの略と解されている」と脚注を付けている. このことを確認するだけでもダウンロードする価値があるだろう.

上で紹介した計算は二項分布の確率質量関数にStirlingの公式を適用する単純な計算に過ぎないのだが, その先には分野を超えた非常に面白い世界が広がっている.

2.5 (2) 二項分布の確率質量関数から正規分布の密度関数が出て来ること

n ごとに決まっている k は $n \rightarrow \infty$ で

$$\frac{k}{n} = p + o(1)$$

という条件を満たしているならば

$$P(k|n, p) = \frac{1}{\sqrt{2\pi np(1-p)}} \left(\left(\frac{k/n}{p} \right)^{k/n} \left(\frac{1-k/n}{1-p} \right)^{1-k/n} \right)^{-n} (1 + o(1))$$

となることを上で示した. 以下では $k/n = p + o(1)$ よりも強い条件

$$\frac{k - np}{\sqrt{n}} = x + \varepsilon_n, \quad \lim_{n \rightarrow \infty} \varepsilon_n = 0$$

を仮定する. このとき,

$$\frac{k}{n} = p + \frac{x + \varepsilon_n}{\sqrt{n}}, \quad 1 - \frac{k}{n} = 1 - p - \frac{x + \varepsilon_n}{\sqrt{n}}.$$

以上の仮定のもとで, $\log(1+t) = t - t^2/2 + t^3/3 - \dots$ を使うと,

$$\begin{aligned}
\log\left(\frac{k/n}{p}\right)^{k/n} &= \frac{k}{n} \log \frac{k/n}{p} = \left(p + \frac{x + \varepsilon_n}{\sqrt{n}}\right) \log\left(1 + \frac{x + \varepsilon_n}{\sqrt{n} p}\right) \\
&= \left(p + \frac{x + \varepsilon_n}{\sqrt{n}}\right) \left(\frac{x + \varepsilon_n}{\sqrt{n} p} - \frac{x^2}{2np^2} + o(1/n)\right) \\
&= \frac{x + \varepsilon_n}{\sqrt{n}} - \frac{x^2}{2np} + \frac{x^2}{np^2} + o(1/n) \\
&= \frac{x + \varepsilon_n}{\sqrt{n}} + \frac{x^2}{2np} + o(1/n), \\
\log\left(1 - \frac{k/n}{1-p}\right)^{1-k/n} &= \left(1 - \frac{k}{n}\right) \log \frac{k/n - 1}{p} = \left(1 - p - \frac{x + \varepsilon_n}{\sqrt{n}}\right) \log\left(1 - \frac{x + \varepsilon_n}{\sqrt{n}(1-p)}\right) \\
&= \left(1 - p - \frac{x + \varepsilon_n}{\sqrt{n}}\right) \left(-\frac{x + \varepsilon_n}{\sqrt{n}(1-p)} - \frac{x^2}{2n(1-p)^2} + o(1/n)\right) \\
&= -\frac{x + \varepsilon_n}{\sqrt{n}} - \frac{x^2}{2n(1-p)} + \frac{x^2}{n(1-p)^2} + o(1/n) \\
&= -\frac{x + \varepsilon_n}{\sqrt{n}} + \frac{x^2}{2n(1-p)} + o(1/n).
\end{aligned}$$

ゆえに,

$$\begin{aligned}
&\log\left(\left(\frac{k/n}{p}\right)^{k/n} \left(\frac{1-k/n}{1-p}\right)^{1-k/n}\right)^{-n} \\
&= -n \left(\log\left(\frac{k/n}{p}\right)^{k/n} + \log\left(1 - \frac{k/n}{1-p}\right)^{1-k/n}\right) \\
&= -\frac{x^2}{2p} - \frac{x^2}{2(1-p)} + o(1) = -\frac{x^2}{2p(1-p)} + o(1).
\end{aligned}$$

したがって,

$$P(k|n, p) = \frac{1}{\sqrt{2\pi p(1-p)}} \exp\left(-\frac{x^2}{2p(1-p)}\right) \frac{1}{\sqrt{n}} (1 + o(1)).$$

この公式の $1/\sqrt{n}$ より前の部分は平均 0, 分散 $p(1-p)$ の正規分布の確率密度関数であり, $1/\sqrt{n}$ の因子は $k = np + \sqrt{n}(x + \varepsilon_n)$ から形式的に得られる $dk = \sqrt{n} dx$ の \sqrt{n} の因子とキャンセルすると考えると自然に見える。

これより, $K_n \sim \text{Binomial}(n, p)$, $X_n \xrightarrow{\text{law}} \text{Normal}(0, p(1-p))$ のとき,

$$X_n = \frac{K_n - np}{\sqrt{n}}, \quad x_n(k) = \frac{k - np}{\sqrt{n}}, \quad \Delta x_n = \frac{1}{\sqrt{n}}$$

とおくと, 有界な連続関数 $f(x)$ について, $n \rightarrow \infty$ のとき

$$\begin{aligned}
E[f(X_n)] &= \sum_{k=0}^n f(x_n(k)) P(k|n, p) \\
&= \sum_{k=0}^n f(x_n(k)) \frac{1}{\sqrt{2\pi p(1-p)}} \exp\left(-\frac{x_n(k)^2}{2p(1-p)}\right) \Delta x_n (1 + o(1)) \\
&= \int_{-\infty}^{\infty} f(x) \frac{1}{\sqrt{2\pi p(1-p)}} \exp\left(-\frac{x^2}{2p(1-p)}\right) dx = E[f(X_\infty)].
\end{aligned}$$

以上の議論は厳密には少しギャップがあるのだが, 二項分布に関して中心極限定理が成立する理由の本質は十分に分かる内容になっている。

2.6 問題: 二項分布の中心極限定理を使った簡単な計算問題

$0 < p < 1$ と $m = 0, 1, 2, \dots$ について以下の極限を求めよ:

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n \left(\frac{k - np}{\sqrt{np(1-p)}}\right)^{2m} \binom{n}{k} p^k (1-p)^{n-k}.$$

解答例: 二項分布の中心極限定理より,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sum_{k=0}^n \left(\frac{k - np}{\sqrt{np(1-p)}} \right)^{2m} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \int_{-\infty}^{\infty} \left(\frac{x}{\sqrt{p(1-p)}} \right)^{2m} \frac{1}{\sqrt{2\pi p(1-p)}} \exp\left(-\frac{x^2}{2p(1-p)}\right) dx \\ &= \int_{-\infty}^{\infty} z^{2m} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-z^2/2} z^{2m} dz \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-y/2} y^{m+1/2-1} dy = \frac{1}{\sqrt{2\pi}} 2^{m+1/2} \Gamma(m+1/2) \\ &= \frac{1}{\sqrt{2\pi}} 2^{m+1/2} \frac{2m-1}{2} \dots \frac{3}{2} \frac{1}{2} \Gamma(1/2) = 1 \cdot 3 \dots (2m-1). \end{aligned}$$

最初の等号で二項分布の中心極限定理を使い, 2番目の等号で $x = \sqrt{p(1-p)} z$ とおき, 4番目の等号で $z = \sqrt{y}$ とおいた. 5番目の等号では $\int_0^{\infty} e^{-y/2} y^{\alpha-1} dy = \theta^{\alpha} \Gamma(\alpha)$ を使った.

解答終

上の問題の結果を数値的に確認してみよう.

In [17]:

```
1 g(n, p, m, k) = ((k - n*p)/√(n*p*(1-p)))^(2m)
2 p = 0.4
3 @show p
4 for m in 0:5
5     println()
6     @show m
7     @show n = 10^(m+1)
8     @show prod(1:2:2m-1)
9     @show sum(g(n, p, m, k)*pdf(Binomial(n, p), k) for k in 0:n)
10 end;

p = 0.4

m = 0
n = 10 ^ (m + 1) = 10
prod(1:2:2m - 1) = 1
sum((g(n, p, m, k) * pdf(Binomial(n, p), k) for k = 0:n)) = 1.0000000000000001

m = 1
n = 10 ^ (m + 1) = 100
prod(1:2:2m - 1) = 1
sum((g(n, p, m, k) * pdf(Binomial(n, p), k) for k = 0:n)) = 1.00000000000000095

m = 2
n = 10 ^ (m + 1) = 1000
prod(1:2:2m - 1) = 3
sum((g(n, p, m, k) * pdf(Binomial(n, p), k) for k = 0:n)) = 2.9981666666666703

m = 3
n = 10 ^ (m + 1) = 10000
prod(1:2:2m - 1) = 15
sum((g(n, p, m, k) * pdf(Binomial(n, p), k) for k = 0:n)) = 14.997416790272869

m = 4
n = 10 ^ (m + 1) = 100000
prod(1:2:2m - 1) = 105
sum((g(n, p, m, k) * pdf(Binomial(n, p), k) for k = 0:n)) = 104.99661670588064

m = 5
n = 10 ^ (m + 1) = 1000000
prod(1:2:2m - 1) = 945
sum((g(n, p, m, k) * pdf(Binomial(n, p), k) for k = 0:n)) = 944.9952750033706
```

3 中心極限定理

中心極限定理 (central limit theorem)は「中心に収束する極限定理」というような意味ではなく、「確率論における中心的な極限定理」という意味である.

注意: 特性函数, モーメント母函数, キュムラント母函数, 歪度(わいど), 尖度(せんど)などについては「標本分布について」のノートに詳しい解説がある.

3.1 中心極限定理のラフな説明

X_1, \dots, X_n は各々が期待値 μ , 分散 σ^2 を持つ分布に従う n 個の独立同分布確率変数達であるとし, n は十分に大きいと仮定する.

中心極限定理は以下のように同値な言い方が色々ある:

(1) それらの和 $X_1 + \dots + X_n$ が従う分布は期待値 $n\mu$, 分散 $n\sigma^2$ の正規分布で近似される:

$$\sum_{i=1}^n X_i \sim \text{Normal}(n\mu, \sqrt{n}\sigma) \quad \text{approximately.}$$

(2) それらの加法平均 $\bar{X}_n = (X_1 + \dots + X_n)/n$ が従う分布は期待値 μ , 分散 σ^2/n の正規分布で近似される:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \sim \text{Normal}(\mu, \sigma/\sqrt{n}) \quad \text{approximately.}$$

(3) $\sqrt{n}(\bar{X}_n - \mu)$ が従う分布は期待値 0, 分散 σ^2 の正規分布で近似される:

$$\sqrt{n}(\bar{X}_n - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \sim \text{Normal}(0, \sigma) \quad \text{approximately.}$$

(4) 次の Z_n が従う分布は標準正規分布で近似される:

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} \sim \text{Normal}(0, 1) \quad \text{approximately.}$$

これらはどれも便利であり, 今後, 自由に使われることになるだろう.

注意・警告: 統計学に中心極限定理を応用する場合には, $n \rightarrow \infty$ とできる数学的に理想的な状況における中心極限定理ではなく, 有限の固定された n における近似として中心極限定理は使われる. だから, 場合ごとにその近似の誤差が問題になる. この問題は目的ごとに異なる許容される誤差との兼ね合いの問題になるので, きれいな一般論で解決できる話ではなく, 中心極限定理のユーザー自身が自分自身の目的に合わせて誤差が大きくなるリスクをどこまで許容するかを決定する必要がある. そのためには, 中心極限定理による近似の精度がどのような場合に悪くなりそうかについて前もって理解しておく必要がある.

補足: 分布 D_n が分布 D_∞ で近似されるとは, 分布 D_n に従う確率変数 X_n と分布 D_∞ に従う確率変数 X_∞ と適切なクラスに含まれる任意の函数 $f(x)$ について

$$\lim_{n \rightarrow \infty} E[f(X_n)] = E[f(X_\infty)]$$

が成立することだと, 大雑把に定義しておく. (函数 $f(x)$ としてどのような函数を許すかについて詳細な説明をするためにはさらなる数学的道具立てが必要で, 雑に扱っても害はほとんどない場合にもなっているので, このノートで説明しない. 所謂「分布収束」「法収束」の話になる.) 統計学に必要な確率論の計算は期待値を取る操作経由で可能なのでこのような定義にしておいても困ることはない.

補足の補足: 数学的詳細を理解できそうもないと感じる人であっても, 中心極限定理の良いユーザーになることは可能である. そのためには, 正規分布で近似したい分布のグラフとそれを近似すると期待される正規分布のグラフを重ねて描く作業を十分に沢山行えばよい. 数学的詳細を理解可能な人であっても, そういう視覚化による理解をサボってしまうと, 有限の n でどのように正規分布による近似の精度が悪くなるかについての感覚が身に付かなくなってしまう. 数学的詳細を理解していなくても, 中心極限定理の利用でどのようなリスクが発生するかを具体例を沢山見ることによって把握していれば, その人は十分に良い中心極限定理ユーザーになれるだろう.

3.2 中心極限定理の特性函数を使った証明

中心極限定理: X_1, X_2, X_3, \dots が独立同分布な確率変数の列であるとき, $\mu = E[X_k]$ が定義されていて, $\sigma^2 = \text{var}(X_k) = E[(X_k - \mu)^2] < \infty$ でかつ, $E[|X_k - \mu|^3] < \infty$ となっていると仮定する. このとき,

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k, \quad Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

とおくと, $n \rightarrow \infty$ で Z_n の分布は標準正規分布に近づく.

証明: X_k の標準化を $Y_k = (X_k - \mu)/\sigma$ と書くことにする. Y_1, Y_2, \dots も独立同分布になり, $E[Y_k] = 0$, $E[Y_k^2] = 1$ が成立している. ゆえに Y_k の特性函数 $\varphi(t)$ は k によらず,

$$\varphi(t) = E[e^{itY_k}] = 1 + iE[Y_k]t - E[Y_k^2]\frac{t^2}{2} + O(t^3) = 1 - \frac{t^2}{2} + O(t^3)$$

という形になる。そして、

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k = \frac{\sqrt{n}}{\sigma} \frac{1}{n} \sum_{k=1}^n (X_k - \mu) = \frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu) = Z_n$$

なので、 Z_n の特性関数は、 $n \rightarrow \infty$ で

$$\begin{aligned} \varphi_{Z_n}(t) &= E \left[\exp \left(it \frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k \right) \right] = E \left[\prod_{k=1}^n \exp \left(i \frac{t}{\sqrt{n}} Y_k \right) \right] \\ &= \prod_{k=1}^n E \left[\exp \left(i \frac{t}{\sqrt{n}} Y_k \right) \right] = \varphi \left(\frac{t}{\sqrt{n}} \right)^n = \left(1 - \frac{t^2}{2n} + O(n^{-3/2}) \right)^n \rightarrow e^{-t^2/2}. \end{aligned}$$

と標準正規分布の特性関数 $e^{-t^2/2}$ に収束する。(3つめの等号で Y_1, \dots, Y_n の独立性を使った.)

ゆえに Z_n の分布は $n \rightarrow \infty$ で標準正規分布に近づく。

証明終

注意: $f(x)$ のFourier変換を

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

と定義するとき、逆Fourier変換によって、

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

が成立しているので、確率変数 X について $f(X)$ の期待値は、 X の特性関数 $\varphi_X(t) = E[e^{itX}]$ を使って、

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

と書ける(積分と期待値を取る操作を交換できる)。このことから、確率変数 X の特性関数と確率変数 Y の特性関数が近ければ、 $f(X)$ と $f(Y)$ の期待値も近くなる。このことから、分布の近似は特性関数の近似で扱えることがわかる。

3.3 中心極限定理の収束の速さと歪度

前節の証明より、独立同分布確率変数列 X_1, X_2, \dots に対して、

$$\mu = E[X_k], \quad \sigma = \sqrt{E[(X - \mu)^2]}, \quad Y_k = \frac{X_k - \mu}{\sigma}, \quad \varphi(t) = E[e^{itY_k}]$$

とおいたときの、 $n \rightarrow \infty$ での $\varphi(t/\sqrt{n})^n \rightarrow e^{-t^2/2}$ という収束の速さを調べれば、中心極限定理による正規分布への収束の速さがわかる。

$$\bar{\mu}_3 = E[Y_k^3] = E \left[\left(\frac{X_k - \mu}{\sigma} \right)^3 \right]$$

とおくと、

$$\varphi \left(\frac{t}{\sqrt{n}} \right) = E[e^{itY_k}] = 1 - \frac{t^2}{2n} - i\bar{\mu}_3 \frac{t^3}{6n\sqrt{n}} + O(n^{-2})$$

なので、

$$\log \varphi \left(\frac{t}{\sqrt{n}} \right)^n = n \log \left(1 - \frac{t^2}{2n} - i\bar{\mu}_3 \frac{t^3}{6n\sqrt{n}} + O(n^{-2}) \right) = -\frac{t^2}{2} - i\bar{\mu}_3 \frac{t^3}{6\sqrt{n}} + O(n^{-1}).$$

これは $n \rightarrow \infty$ での $\log \varphi(t/\sqrt{n})^n \rightarrow -t^2/2$ の収束の速さは、 $Y_k = (X_k - \mu)/\sigma$ の3次のモーメント $\bar{\mu}_3$ の絶対値の大きさが大体決まっていることがわかる。 $\bar{\mu}_3$ の絶対値が小さいほど収束が速く、大きいほど収束が遅い。

$\bar{\mu}_3$ は X_k の分布の期待値 μ を中心とする非対称性の σ によって適切に正規化した尺度になっている。 $\bar{\mu}_3$ は $Y_k = (X_k - \mu)/\sigma$ の3次のキュムラントにも一致している:

$$K_{Y_k}(t) = \log E[e^{tY_k}] = \log \left(1 + \frac{t^2}{2} + \bar{\mu}_3 \frac{t^3}{3!} + O(t^4) \right) = \frac{t^2}{2} + \bar{\mu}_3 \frac{t^3}{3!} + O(t^4).$$

ここでの $t^3/3!$ の係数 $\bar{\kappa}_3 = \bar{\mu}_3$ は X_k の **歪度** (skewness) と呼ばれるのであった。

3.4 中心極限定理のキュムラント母函数を使った証明

中心極限定理: X_1, X_2, X_3, \dots が独立同分布な確率変数の列であるとき, $\mu = E[X_k]$ が定義されていて, $\sigma^2 = \text{var}(X_k) = E[(X_k - \mu)^2] < \infty$ でかつ, $E[|X_k - \mu|^3] < \infty$ となっており, さらに各 X_k のキュムラント母函数がうまく定義されているとする. このとき,

$$\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k, \quad Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

とおくと, $n \rightarrow \infty$ で Z_n の分布は標準正規分布に近づく.

証明: X_k の標準化を $Y_k = (X_k - \mu)/\sigma$ と書くことにする. Y_1, Y_2, \dots も独立同分布になり, $E[Y_k] = 0$, $E[Y_k^2] = 1$ が成立している. ゆえに Y_k のキュムラント母函数 $K(t)$ は k によらず,

$$K(t) = K_{Y_k}(t) = \frac{t^2}{2} + O(t^3)$$

という形になる. そして,

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k = \frac{\sqrt{n}}{\sigma} \frac{1}{n} \sum_{k=1}^n (X_k - \mu) = \frac{\sqrt{n}}{\sigma} (\bar{X}_n - \mu) = Z_n$$

なので, Z_n のキュムラント母函数は, $n \rightarrow \infty$ で

$$\begin{aligned} K_{Z_n}(t) &= K_{Y_1/\sqrt{n} + \dots + Y_n/\sqrt{n}}(t) = K_{Y_1} \left(\frac{t}{\sqrt{n}} \right) + \dots + K_{Y_1} \left(\frac{t}{\sqrt{n}} \right) \\ &= nK \left(\frac{t}{\sqrt{n}} \right) = n \left(\frac{t^2}{2n} + O(n^{-3/2}) \right) = \frac{t^2}{2} + O(n^{-1/2}) \rightarrow \frac{t^2}{2} \end{aligned}$$

と標準正規分布のキュムラント母函数 $t^2/2$ に収束する.

ゆえに Z_n の分布は $n \rightarrow \infty$ で標準正規分布に近づく.

証明終

注意: このようにキュムラント母函数を使うと証明が非常にシンプルになる.

3.5 中心極限定理の収束の速さと歪度と尖度

前節の証明より, 独立同分布確率変数列 X_1, X_2, \dots に対して,

$$\mu = E[X_k], \quad \sigma = \sqrt{E[(X_k - \mu)^2]}, \quad K(t) = K_{(X_k - \mu)/\sigma}(t) = \log E \left[\left(\frac{X_k - \mu}{\sigma} \right)^t \right]$$

とおいたときの, $n \rightarrow \infty$ での $nK(t/\sqrt{n}) \rightarrow t^2/2$ という収束の速さを調べれば, 中心極限定理による正規分布への収束の速さがわかる. X_k の歪度(わいど) $\bar{\kappa}_3$ と尖度(せんど) $\bar{\kappa}_4$ は

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

と表され, X_k の標準化のキュムラント母函数 $K(t)$ の展開の係数になっているのであった:

$$K(t) = \frac{t^2}{2} + \bar{\kappa}_3 \frac{t^3}{3!} + \bar{\kappa}_4 \frac{t^4}{4!} + O(t^5).$$

このとき,

$$nK \left(\frac{t}{\sqrt{n}} \right) = n \left(\frac{t^2}{2n} + \bar{\kappa}_3 \frac{t^3}{3! n^{3/2}} + \bar{\kappa}_4 \frac{t^4}{4! n^2} + O(n^{-5/2}) \right) = \frac{t^2}{2} + \bar{\kappa}_3 \frac{t^3}{3! \sqrt{n}} + \bar{\kappa}_4 \frac{t^4}{4! n} + O(n^{-3/2})$$

これが $t^2/2$ に収束する速さは $\bar{\kappa}_3 \neq 0$ ならば $O(n^{-1/2})$ のオーダーになり、歪度 $\bar{\kappa}_3$ の絶対値が大きいほど遅くなる。そして、 $\bar{\kappa}_3 = 0$ ならば $O(n^{-1})$ のオーダーでの収束になり、尖度 $\bar{\kappa}_4$ の絶対値が大きいほど収束は遅くなる。

3.6 中心極限定理のTaylorの定理のみを使う証明

以上では中心極限定理を特性函数やキュムラント母函数を使って証明する方法を(論理的には大雑把に)紹介した。

その方法は本質的にFourier解析に依存している。

実はFourier解析に依存せずに、本質的にTaylorの定理(Taylor展開の剰余項版)しか使わない初等的な証明も存在する。

しかし、読者の中にはTaylorの定理も証明をよく知らない人もいるかもしれないので、以下では丁寧にTaylorの定理の証明の概略から説明して行くことにする。

3.6.1 多重積分の書き方

以下では多重積分での括弧を略して書くために、次の式の右辺を左辺のように書くことにする：

$$\int_a^b dx f(x) = \int_a^b f(x) dx.$$

このように書くと、次の右辺のように括弧が沢山必要になる多重積分の式を左辺のようにシンプルに書くことができる：

$$\int_a^x dx_1 \int_a^{x_1} dx_2 \int_a^{x_2} dx_3 g(x_3) = \int_a^x \left(\int_a^{x_1} \left(\int_a^{x_2} g(x_3) dx_3 \right) dx_2 \right) dx_1$$

しかも、この式の右辺の書き方だと、積分変数の x_i 達の指定が式の最後の方でされているので、 $\int_a^{x_i}$ の部分の解釈を確定させるためには式を最後の方まで全部見る必要が生じてしまう。左辺の書き方だとそのような問題が生じない。

3.6.2 積分剰余項型のTaylorの定理

以下では次の結果を示そう：

$$R_4(x, a) = \int_a^x dx_1 \int_a^{x_1} dx_2 \int_a^{x_2} dx_3 \int_a^{x_3} dx_4 f^{(4)}(x_4)$$

とおくと、

$$f(x) = f(a) + f'(a)(x-a) + f''(a)\frac{(x-a)^2}{2} + f'''(a)\frac{(x-a)^3}{3!} + R_4(x, a).$$

これを示すためには4階の導函数 $f^{(4)}(x)$ を4回不定積分すればよい。すなわち、「微分して不定積分するときに適切に積分定数を決めてやればもとの函数に戻る」という結果

$$g(x) = g(a) + \int_a^x dx' g'(x')$$

を4回使うと、

$$\begin{aligned}
f'''(x_3) &= f'''(a) + \int_a^{x_3} dx_4 f^{(4)}(x_4), \\
f''(x_2) &= f''(a) + \int_a^{x_2} dx_3 f'''(x_3) \\
&= f''(a) + f'''(a)(x_2 - a) + \int_a^{x_2} dx_3 \int_a^{x_3} dx_4 f^{(4)}(x_4) \\
f'(x_1) &= f'(a) + \int_a^{x_1} dx_2 f''(x_2) \\
&= f'(a) + f''(a)(x_1 - a) + f'''(a) \frac{(x_1 - a)^2}{2} \\
&\quad + \int_a^{x_1} dx_2 \int_a^{x_2} dx_3 \int_a^{x_3} dx_4 f^{(4)}(x_4) \\
f(x) &= f(a) + \int_a^x dx_1 f'(x_1) \\
&= f(a) + f'(a)(x - a) + f''(a) \frac{(x - a)^2}{2} + f'''(a) \frac{(x - a)^3}{3!} \\
&\quad + \underbrace{\int_a^x dx_1 \int_a^{x_1} dx_2 \int_a^{x_2} dx_3 \int_a^{x_3} dx_4 f^{(4)}(x_4)}_{=R_4(x,a)}.
\end{aligned}$$

不定積分を繰り返す過程で、

$$\int_a^x dx' \frac{(x' - a)^k}{k!} = \frac{(x - a)^{k+1}}{(k + 1)!}$$

となることから、次の形の式が得られた:

$$f(a) + f'(a)(x - a) + f''(a) \frac{(x - a)^2}{2} + f'''(a) \frac{(x - a)^3}{3!}$$

以上とまったく同様にして次を示すことができる:

$$R_n(x, a) = \int_a^x dx_1 \int_a^{x_1} dx_2 \cdots \int_a^{x_{n-1}} dx_n f^{(n)}(x_n)$$

とおくと、

$$f(x) = \sum_{k=0}^{n-1} f^{(k)}(a) \frac{(x - a)^k}{k!} + R_n(x, a).$$

この結果を **積分剰余項型のTaylorの定理** と呼ぶことにする。

もしかしら、大学1年生のときの講義で剰余項の形が別のTaylorの定理について複雑な証明を聞いた人が多いかもしれないが、Taylorの定理の本質は「沢山微分して沢山積分すればもとの函数に戻る」ということに過ぎず、直観的にはほとんど明らかな結果に過ぎない。明らかな結果について「わけがわからないがとにかく成立する定理」だと認識してしまうことは数学を理解するときに避けるべきことである。

3.6.3 積分剰余項の別の表示

剰余項 $R_n(x, t)$ は次の式で定義されたと考えてよい:

$$R_n(x, t) = f(x) - \sum_{k=0}^{n-1} f^{(k)}(t) \frac{(x - t)^k}{k!}.$$

このとき、 $R_n(x, x) = 0$ であつ、両辺を t で偏微分すると、

$$\begin{aligned}
\frac{\partial}{\partial t} R_n(x, t) &= \sum_{k=1}^{n-1} f^{(k)}(t) \frac{(x - t)^{k-1}}{(k - 1)!} - \sum_{k=0}^{n-1} f^{(k+1)}(t) \frac{(x - t)^k}{k!} \\
&= \sum_{k=1}^{n-1} f^{(k)}(t) \frac{(x - t)^{k-1}}{(k - 1)!} - \sum_{k=1}^n f^{(k)}(t) \frac{(x - t)^{k-1}}{(k - 1)!} \\
&= -f^{(n)}(t) \frac{(x - t)^{n-1}}{(n - 1)!}.
\end{aligned}$$

この偏導函数を t について x から a まで積分することによって(a から x まで積分して -1 倍することと同じ)、次が得られる:

$$R_n(x, a) = \int_a^x f^{(n)}(t) \frac{(x-t)^{n-1}}{(n-1)!} dt.$$

以上によっても積分剰余項型のTaylorの定理がシンプルに証明されたことになる.

Taylorの定理の証明は易しい!

注意: $h = x - a, t = a + sh$ とおくと, $R_n(x, a)$ は次のようにも表される:

$$R_n(a+h, a) = h^n \int_0^1 f^{(n)}(a+sh) \frac{(1-s)^{n-1}}{(n-1)!} ds.$$

すなわち,

$$f(a+h) = \sum_{k=0}^{n-1} f^{(k)}(a) \frac{h^k}{k!} + h^n \int_0^1 f^{(n)}(a+sh) \frac{(1-s)^{n-1}}{(n-1)!} ds.$$

後でTaylorの定理をこの形で使うことになる.

注意: 上で求めた積分剰余項の公式と前節で求めた公式の関係. 前節では

$$R_n(x, a) = \int_a^x dx_1 \int_a^{x_1} dx_2 \cdots \int_a^{x_{n-1}} dx_n f^{(n)}(x_n)$$

を示した. 簡単のため $a \leq x$ と仮定する($a \geq x$ の場合も同様である). そのとき, すぐ上の式の右辺の積分の範囲は

$$a \leq x_n \leq x_{n-1} \leq x_{n-2} \leq \cdots \leq x_2 \leq x_1 \leq x$$

なので, x_n による積分を一番外側に出すと次のように書き直される:

$$R_n(x, a) = \int_a^x f^{(n)}(x_n) \left(\int_{x_n}^x dx_1 \int_{x_n}^{x_1} dx_2 \cdots \int_{x_n}^{x_{n-2}} dx_{n-1} \right) dx_n.$$

$\int_{x_n}^x dx_1 \int_{x_n}^{x_1} dx_2 \cdots \int_{x_n}^{x_{n-2}} dx_{n-1}$ の部分は 1 の $n-1$ 回の不定積分なので $(x-x_n)^{n-1}/(n-1)!$ に等しい:

$$\int_{x_n}^x dx_1 \int_{x_n}^{x_1} dx_2 \cdots \int_{x_n}^{x_{n-2}} dx_{n-1} = \frac{(x-x_n)^{n-1}}{(n-1)!}.$$

これより, 前節で求めた積分剰余項の公式から上で求めた公式が得られることがわかった.

3.6.4 微分剰余項型のTaylorの定理

$f(t)$ は C^n 級であると仮定する.

簡単のため $a \leq x$ と仮定する($a \geq x$ の場合も同様である). a 以上 x 以下の t についての $f^{(n)}(t)$ の最小値と最大値をそれぞれ $f^{(n)}(\alpha), f^{(n)}(\beta)$ ($a \leq \alpha, \beta \leq x$) と書くと,

$$f^{(n)}(\alpha) \frac{(x-a)^n}{n!} \leq R_n(x, a) = \int_a^x f^{(n)}(t) \frac{(x-t)^{n-1}}{(n-1)!} dt \leq f^{(n)}(\beta) \frac{(x-a)^n}{n!}.$$

ゆえに, 中間値の定理より, α と β のあいだのある実数 ξ が存在して,

$$f^{(n)}(\xi) \frac{(x-a)^n}{n!} = R_n(x, a).$$

これで積分を使わずに微分だけを使った剰余項の表示が得られた.

3.6.5 Taylorの定理を使った中心極限定理の証明

X_1, X_2, X_3, \dots は同分布独立確率変数列であるとし, それらの共通の期待値 $\mu = E[X_i]$ は 0 であるとし, 共通の分散 $\sigma^2 = E[(X_i - \mu)^2]$ は 1 であるとし, $E[|X_i|^m] < \infty$ ($m = 3, 4, 5$) であると仮定し, X_i の歪度と尖度を次のように書くことにする:

$$\bar{\kappa}_3 = E[X_i^3], \quad \bar{\kappa}_4 = E[X_i^4] - 3.$$

このとき, Z を標準正規分布に従う確率変数であるとし,

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i.$$

とおくと、有限区間の外で 0 になる C^5 級関数 $f(z)$ に対して、 $m = 3, 4, 5$ に関する $|f^{(m)}(z)|$ の最大値を M_m と書くと、 $n \rightarrow \infty$ において次が成立している:

$$\begin{aligned} & |E[f(Z_n)] - E[f(Z)]| \\ & \leq \frac{M_3}{3! \sqrt{n}} |\bar{\kappa}_3| + \frac{M_4}{4! n} |\bar{\kappa}_4| + \frac{M_5}{5! n \sqrt{n}} (E[|X_i|^5] + E[|Z|^5]). \end{aligned}$$

特に次が成立している:

$$\lim_{n \rightarrow \infty} E[f(Z_n)] = E[f(Z)].$$

注意: 必要ならば X_i を $(X_i - \mu)/\sigma$ で置き換えることによって、上のように仮定しても一般性が失われないことに注意せよ。

証明:

Step 1. Y_1, Y_2, Y_3, \dots は X_1, X_2, X_3, \dots とは別の独立同分布確率変数列であるとし、 X_i, Y_i の全体は独立でかつ $E[Y_i] = 0$, $E[Y_i^2] = 1$, $E[|Y_i|^3] < \infty$ を満たしていると仮定する。

$W_n, Z_n^{(k)}, A_n^{(k)}$ を次のように定める:

$$\begin{aligned} W_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i, \\ Z_n^{(k)} &= \frac{1}{\sqrt{n}} (X_1 + \dots + X_k + Y_{k+1} + \dots + Y_n), \\ A_n^{(k)} &= \frac{1}{\sqrt{n}} (X_1 + \dots + X_{k-1} + Y_{k+1} + \dots + Y_n). \end{aligned}$$

このとき,

$$Z_n = Z_n^{(n)}, \quad W_n = Z_n^{(0)}, \quad Z_n^{(k)} = A_n^{(k)} + \frac{X_k}{\sqrt{n}}, \quad Z_n^{(k-1)} = A_n^{(k)} + \frac{Y_k}{\sqrt{n}}.$$

以下では、大雑把に言う、 $Z_n^{(k)}$ の分布と $Z_n^{(k-1)}$ の分布の違いは小さいことを示す。

まず、 $G(a, h)$ を次のように定める:

$$G(a, h) = \int_0^1 f^{(5)}(a + sh) \frac{(1-s)^4}{4!} ds.$$

$m = 3, 4, 5$ に対する $|f^{(m)}(z)|$ の最大値を M_m と書くと,

$$|G(a, h)| \leq M_5 \int_0^1 \frac{(1-s)^4}{4!} ds = \frac{M_5}{5!}.$$

$f(z)$ に Taylor の定理を適用すると,

$$f(a + h) = f(a) + hf'(a) + \frac{h^2}{2} f''(a) + \frac{h^3}{3!} f'''(a) + \frac{h^4}{4!} f^{(4)}(a) + h^5 G(a, h).$$

これを $a = A_n^{(k)}$, $h = X_k/\sqrt{n}, Y_k/\sqrt{n}$ に適用すると,

$$\begin{aligned} f(Z_n^{(k)}) &= f(A_n^{(k)}) + \frac{X_k}{\sqrt{n}} f'(A_n^{(k)}) + \frac{X_k^2}{2n} f''(A_n^{(k)}) \\ &\quad + \frac{X_k^3}{3! n \sqrt{n}} f'''(A_n^{(k)}) + \frac{X_k^4}{4! n^2} f^{(4)}(A_n^{(k)}) + \frac{X_k^5}{n^2 \sqrt{n}} G\left(A_n^{(k)}, \frac{X_k}{\sqrt{n}}\right), \\ f(Z_n^{(k-1)}) &= f(A_n^{(k)}) + \frac{Y_k}{\sqrt{n}} f'(A_n^{(k)}) + \frac{Y_k^2}{2n} f''(A_n^{(k)}) \\ &\quad + \frac{Y_k^3}{3! n \sqrt{n}} f'''(A_n^{(k)}) + \frac{Y_k^4}{4! n^2} f^{(4)}(A_n^{(k)}) + \frac{Y_k^5}{n^2 \sqrt{n}} G\left(A_n^{(k)}, \frac{Y_k}{\sqrt{n}}\right). \end{aligned}$$

$A_n^{(k)}, X_k, Y_k$ は独立で、 $E[X_k] = E[Y_k] = 0$, $E[X_k^2] = E[Y_k^2] = 1$ であることより,

$$\begin{aligned}
& E[f(Z_n^{(k)})] - E[f(Z_n^{(k-1)})] \\
&= \frac{E[f'''(A_n^{(k)})]}{3! n \sqrt{n}} (E[X_k^3] - E[Y_k^3]) + \frac{E[f^{(4)}(A_n^{(k)})]}{4! n^2} (E[X_k^4] - E[Y_k^4]) \\
&+ \frac{1}{n^2 \sqrt{n}} \left(E \left[X_k^5 G \left(A_n^{(k)}, \frac{Y_k}{\sqrt{n}} \right) \right] + E \left[Y_k^5 G \left(A_n^{(k)}, \frac{Y_k}{\sqrt{n}} \right) \right] \right)
\end{aligned}$$

$m = 3, 4, 5$ に対する $|f^{(m)}(z)|$ の最大値を M_m と書くことにしてあったので,

$$\begin{aligned}
& |E[f(Z_n^{(k)})] - E[f(Z_n^{(k-1)})]| \\
&\leq \frac{M_3}{3! n \sqrt{n}} |E[X_k^3] - E[Y_k^3]| + \frac{M_4}{4! n^2} |E[X_k^4] - E[Y_k^4]| \\
&+ \frac{M_5}{5! n^2 \sqrt{n}} (E[|X_k|^5] + E[|Y_k|^5]).
\end{aligned}$$

Step 2. X_k, Y_k の分布が k によらないことに注意しながら, これを $k = 1, 2, \dots, n$ について足し上げることによって次が得られる:

$$\begin{aligned}
& |E[f(Z_n)] - E[f(W_n)]| \\
&\leq \frac{M_3}{3! \sqrt{n}} |E[X_1^3] - E[Y_1^3]| + \frac{M_4}{4! n} |E[X_1^4] - E[Y_1^4]| \\
&+ \frac{M_5}{5! n \sqrt{n}} (E[|X_1|^5] + E[|Y_1|^5]).
\end{aligned}$$

Step 3. Y_1, Y_2, Y_3, \dots の各々が標準正規分布に従っている場合. そのとき, $\sum_{i=1}^n Y_i$ は平均 0, 分散 n の正規分布に従うので, $W_n = (1/\sqrt{n}) \sum_{i=1}^n Y_i$ は標準正規分布に従う. ゆえに $E[f(W_n)] = E[f(Z)]$. さらにこのとき, $E[Y_1^3] = 0$, $E[Y_1^4] = 3$ となっていることにも注意せよ. ゆえに, この場合に Step 2 の結果を適用すると次が得られる:

$$\begin{aligned}
& |E[f(Z_n)] - E[f(Z)]| \\
&\leq \frac{M_3}{3! \sqrt{n}} |\bar{\kappa}_3| + \frac{M_4}{4! n} |\bar{\kappa}_4| + \frac{M_5}{5! n \sqrt{n}} (E[|X_i|^5] + E[|Z|^5]).
\end{aligned}$$

これが示したい不等式であった.

証明終

注意: Z を標準正規分布に従う確率変数とすると,

$$E[|Z|^m] = \frac{\Gamma((m+1)/2)}{\sqrt{2\pi}} = \begin{cases} 1 \cdot 3 \cdots (2k-1) & (m = 2k \in 2\mathbb{Z}_{\geq 0} + 1) \\ 2^{k+1} k! / \sqrt{2\pi} & (m = 2k+1 \in 2\mathbb{Z}_{\geq 0}) \end{cases}$$

```
In [18]: 1 [(2*quadgk(z -> z^(2k)*pdf(Normal(), z), 0, Inf)[1],
          2 2^(k+1/2)*gamma(k+1/2)/sqrt(2pi), prod(1:2:2k-1)) for k in 0:4]
```

```
Out[18]: 5-element Vector{Tuple{Float64, Float64, Int64}}:
 (0.9999999999999998, 1.0000000000000002, 1)
 (1.0000000000000009, 1.0000000000000002, 1)
 (3.0000000000000005, 3.0000000000000004, 3)
 (15.000000000000029, 15.000000000000004, 15)
 (105.00000000000988, 105.00000000000001, 105)
```

```
In [19]: 1 [(2*sqrt(2pi)*quadgk(z -> z^(2k+1)*pdf(Normal(), z), 0, Inf)[1],
          2 2^(k+1)*gamma(k+1), 2^(k+1)*factorial(k)) for k in 0:4]
```

```
Out[19]: 5-element Vector{Tuple{Float64, Float64, Int64}}:
 (1.9999999999999993, 2.0, 2)
 (4.0, 4.0, 4)
 (16.000000000000001, 16.0, 16)
 (96.000000000000463, 96.0, 96)
 (767.9999999999999, 768.0, 768)
```

3.6.6 中心極限定理の収束の速さと歪度と尖度(再)

前節で示した不等式

$$|E[f(Z_n)] - E[f(Z)]| \leq \frac{M_3}{3! \sqrt{n}} |\bar{\kappa}_3| + \frac{M_4}{4! n} |\bar{\kappa}_4| + \frac{M_5}{5! n \sqrt{n}} (E[|X_i|^5] + E[|Z|^5]).$$

より, 再び, X_i 達に共通の歪度 $\bar{\kappa}_3$ と尖度 $\bar{\kappa}_4$ の大きさが中心極限定理の収束の速さにどのように関係しているかがわかる.

$\bar{\kappa}_3 = 0$ ならば中心極限定理の収束の速さのオーダーは $O(1/\sqrt{n})$ から $O(1/n)$ に速くなる. $\bar{\kappa}_3 = 0$ の場合には $\bar{\kappa}_4$ が 0 に近いと中心極限定理の収束はさらに速くなる.

3.7 問題: 中心極限定理の収束の様子グラフ

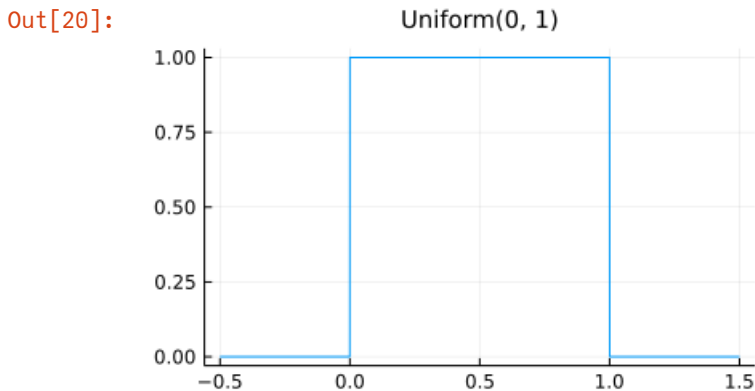
中心極限定理による正規分布への収束の様子をコンピュータでグラフを描いて確認せよ. 収束が速い場合($n = 10$ ですでに正規分布に十分に近い場合)と遅い場合の両方の例を作れ.

注意: 特に収束が遅い場合の様子を確認することが重要である. 中心極限定理は統計学において空気のごとく使用されるが, 数学的には収束が遅い場合があるので, 現実の分析で中心極限定理を使うことには注意を要する. しかし, 単に注意を要することを知っているだけでは役に立たない. 中心極限定理による収束が遅い場合について具体的な例を知っていれば, 危険を察知し易くなるだろう.

解答例は以下の通り.

```
In [20]: 1 # 中心極限定理による収束が速い場合1
2 # distを左右対称な分布でかつ「おとなしめなもの」とする
3 dist = Uniform(0, 1)
4 @show skewness(dist) kurtosis(dist)
5 plot(dist, -0.5, 1.5; label="")
6 title!("Uniform(0, 1)")
```

```
skewness(dist) = 0.0
kurtosis(dist) = -1.2
```



```
In [21]: 1 # 分布distの独立同分布確率変数達(n個)の実現値(要するに乱数)を大量に(L個)生成
2 n = 10
3 L = 10^6
4 Xs = rand(dist, n, L) # Xs の s は X = (X_1, ..., X_n) 達意味(複数という意味)
```

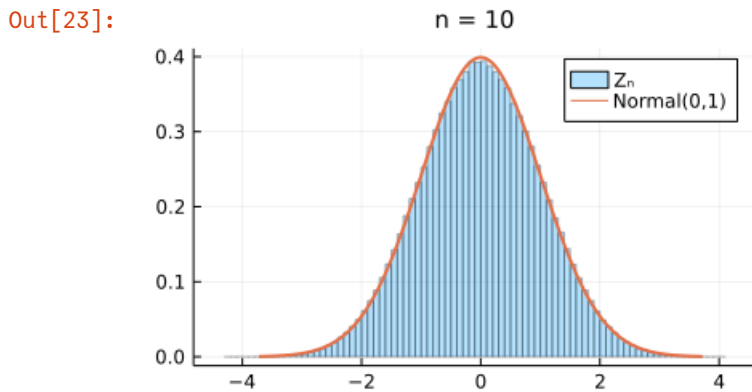
Out[21]: 10×1000000 Matrix{Float64}:

0.382097	0.889019	0.50413	0.273555	...	0.728294	0.901365	0.136964	0.315117
0.698131	0.855675	0.171981	0.892106		0.577322	0.248402	0.493341	0.320716
0.12181	0.384233	0.34591	0.625776		0.817476	0.404155	0.431969	0.106353
0.623901	0.281624	0.91358	0.407791		0.619469	0.70791	0.287514	0.28366
0.800837	0.00683821	0.327856	0.624212		0.76786	0.936917	0.468953	0.310032
0.375193	0.780737	0.283468	0.46302	...	0.322631	0.786465	0.225806	0.733639
0.237805	0.250958	0.421322	0.922456		0.239982	0.852617	0.470971	0.0218825
0.732326	0.777625	0.759354	0.856429		0.286465	0.781914	0.697131	0.343346
0.814374	0.249683	0.355184	0.474089		0.576051	0.851128	0.70772	0.01895
0.742497	0.980567	0.84963	0.97921		0.515878	0.877632	0.975434	0.0848473

```
In [22]: 1 #  $Z_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$  を大量に計算
2  $\mu = \text{mean}(\text{dist})$ 
3  $\sigma = \text{std}(\text{dist})$ 
4  $Zns = [\sqrt{n} * (\text{mean}(X) - \mu) / \sigma \text{ for } X \text{ in eachcol}(Xs)]$  #  $Zns$  は  $Z_n$  達という意味
5  $\text{first}(Zns, 5)$ 
```

```
Out[22]: 5-element Vector{Float64}:
 0.5794592164665237
 0.5005729047211307
-0.07403666130137182
 1.663592359743632
-0.4974519171725785
```

```
In [23]: 1 #  $Z_n$  達のヒストグラムと標準正規分布の密度函数を比較
2  $\text{histogram}(Zns; \text{norm}=\text{true}, \text{alpha}=0.3, \text{bin}=100, \text{label}="Z_n")$ 
3  $\text{plot!}(\text{Normal}(0, 1); \text{label}="Normal(0,1)", \text{lw}=2)$ 
4  $\text{title!}("n = \$n")$ 
```



In [24]:

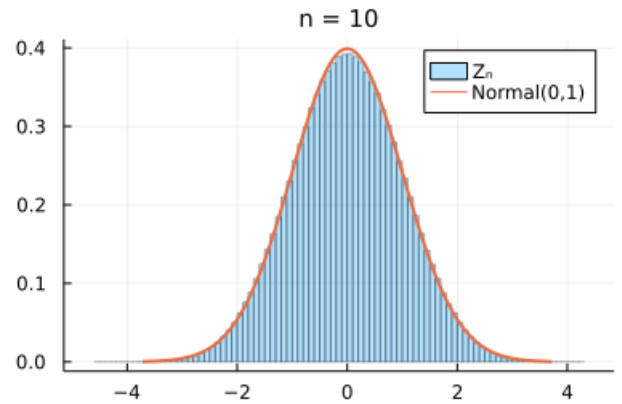
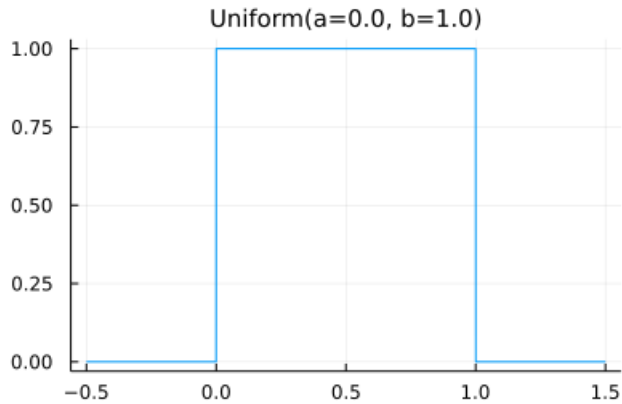
```
1 # 以上の手続きを函数化
2 # さらに
3 # * メモリアロケーションの節約と並列化による最適化
4 # * 離散分布の場合にも対応
5 # * 歪度と尖度を混合モデルの場合にも表示
6
7 distname(dist::Distribution) = replace(string(dist), r"{.*}" => "")
8 myskewness(dist) = skewness(dist)
9 mykurtosis(dist) = kurtosis(dist)
10 function standardized_moment(dist::ContinuousUnivariateDistribution, m)
11     μ, σ = mean(dist), std(dist)
12     quadgk(x -> (x - μ)^m * pdf(dist, x), extrema(dist)...)[1] / σ^m
13 end
14 myskewness(dist::MixtureModel{Univariate, Continuous}) = standardized_moment(dist, 3)
15 mykurtosis(dist::MixtureModel{Univariate, Continuous}) = standardized_moment(dist, 4) - 3
16
17 function plot_central_limit_theorem(dist, n;
18     L=10^6,
19     μ = mean(dist),
20     σ = std(dist),
21     a = max(minimum(dist), μ - 5σ),
22     b = min(maximum(dist), μ + 5σ),
23     disttitle = distname(dist),
24     bin = 100,
25     kwargs...
26 )
27     println("skewness(dist) = ", myskewness(dist))
28     println("kurtosis(dist) = ", mykurtosis(dist))
29
30     # 分布distをプロット
31     if dist isa DiscreteUnivariateDistribution
32         P1 = bar(round(Int, a):round(Int, b), x -> pdf(dist, x), ; alpha=0.3, label="")
33     else
34         P1 = plot(x -> pdf(dist, x), a, b; label="")
35     end
36     title!(disttitle)
37
38     # 分布distの独立同分布確率変数達(n個)の実現値(要するに乱数)を大量に(L個)生成し,
39     #  $Z_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$  を大量に計算することを並列化で実行
40     # このような計算には並列化が非常に有効である
41     Zns = Vector{Float64}(undef, L)
42     tmp = [Vector{Float64}(undef, n) for _ in 1:Threads.nthreads()]
43     Threads.@threads for i in 1:L # Threads.@threads マクロで並列化
44         X = rand!(dist, tmp[Threads.threadid()]) # rand! を使ってアロケーションを節約
45         Zns[i] = sqrt(n)*(mean(X) - μ)/σ
46     end
47
48     # Z達のヒストグラムと標準正規分布の密度函数を比較
49     if dist isa DiscreteUnivariateDistribution
50         summin = max(n*minimum(dist), round(n*μ - 6*sqrt(n*σ)))
51         summax = min(n*maximum(dist), round(n*μ + 6*sqrt(n*σ)))
52         sumran = summin-0.5:summax+0.5
53         bin = @. sqrt(n)*(sumran/n - μ)/σ
54         P2 = histogram(Zns; norm=true, alpha=0.3, bin, label="Z_n")
55     else
56         P2 = histogram(Zns; norm=true, alpha=0.3, bin, label="Z_n")
57     end
58     plot!(Normal(0,1); label="Normal(0,1)", lw=2)
59     title!("n = $n")
60
61     plot(P1, P2; size=(800, 250), kwargs...)
62 end
```

Out[24]: plot_central_limit_theorem (generic function with 1 method)

```
In [25]: 1 # 中心極限定理による収束が速い場合1の再現
          2 plot_central_limit_theorem(Uniform(0, 1), 10; a=-0.5, b=1.5)
```

```
skewness(dist) = 0.0
kurtosis(dist) = -1.2
```

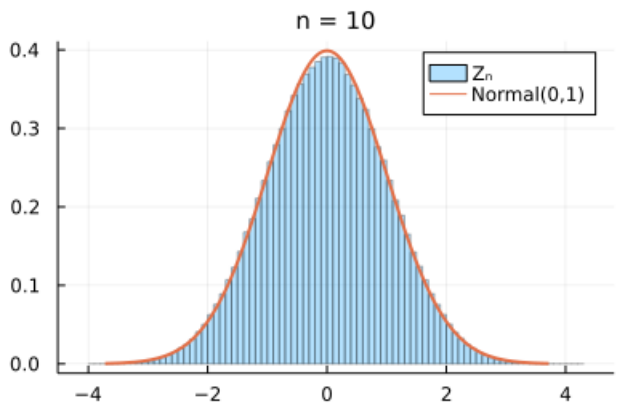
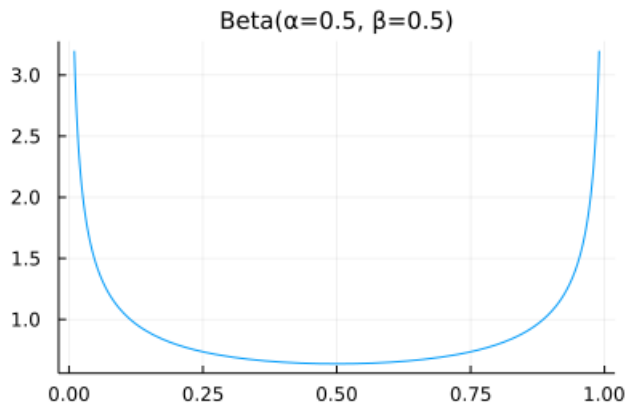
Out[25]:



```
In [26]: 1 # 中心極限定理による収束が速い場合2
          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
```

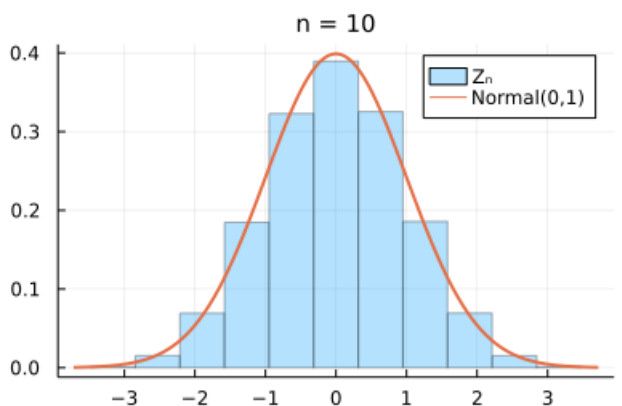
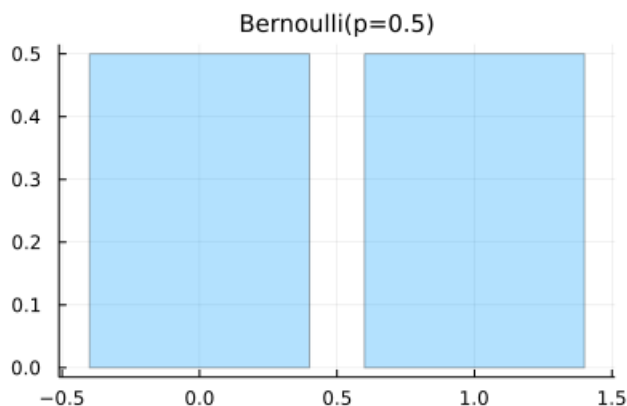
Out[26]:



```
In [27]: 1 # 中心極限定理による収束が速い場合3
          2 plot_central_limit_theorem(Bernoulli(0.5), 10)
```

```
skewness(dist) = 0.0
kurtosis(dist) = -2.0
```

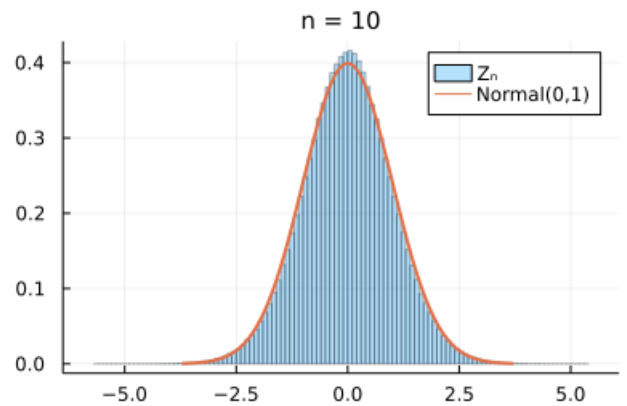
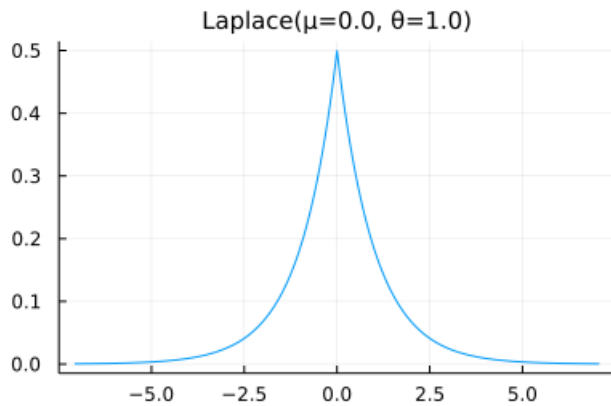
Out[27]:



```
In [28]: 1 # 中心極限定理による収束が速い場合4
        2 plot_central_limit_theorem(Laplace(), 10)
```

```
skewness(dist) = 0.0
kurtosis(dist) = 3.0
```

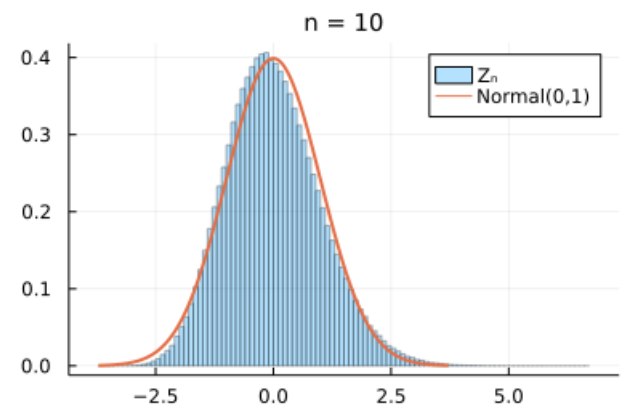
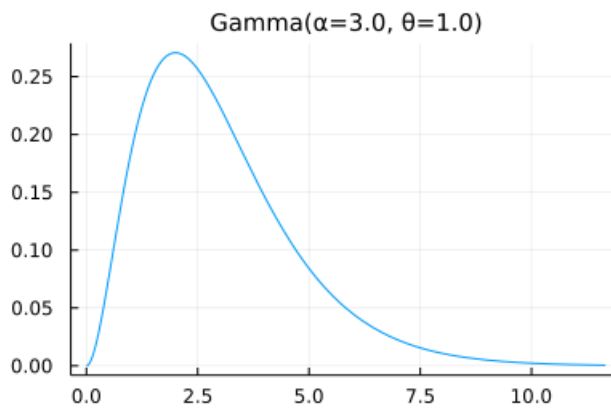
Out[28]:



```
In [29]: 1 # 左右非対称な分布の場合1
        2 plot_central_limit_theorem(Gamma(3, 1), 10)
```

```
skewness(dist) = 1.1547005383792517
kurtosis(dist) = 2.0
```

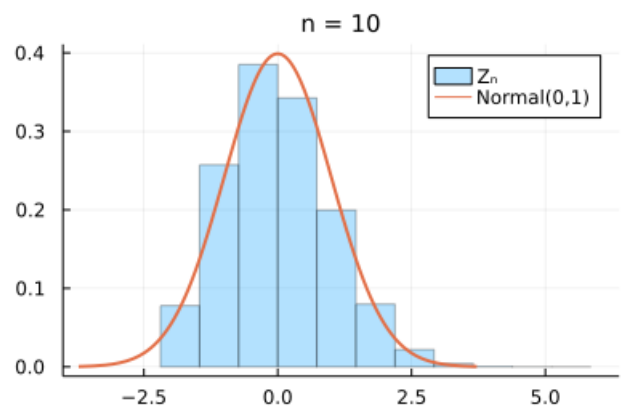
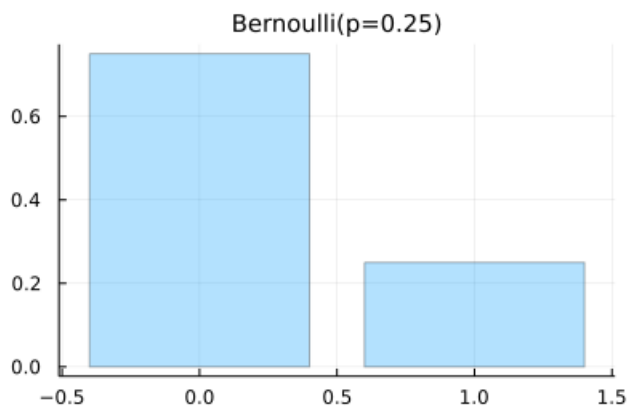
Out[29]:



```
In [30]: 1 # 左右非対称な分布の場合2
        2 plot_central_limit_theorem(Bernoulli(0.25), 10)
```

```
skewness(dist) = 1.1547005383792517
kurtosis(dist) = -0.6666666666666667
```

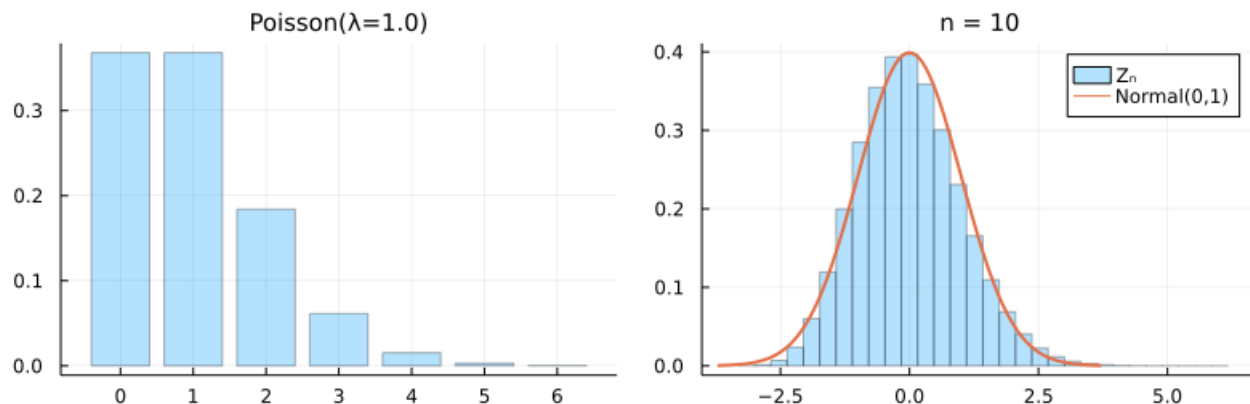
Out[30]:



```
In [31]: 1 # 左右非対称な分布の場合3
        2 plot_central_limit_theorem(Poisson(1), 10)
```

```
skewness(dist) = 1.0
kurtosis(dist) = 1.0
```

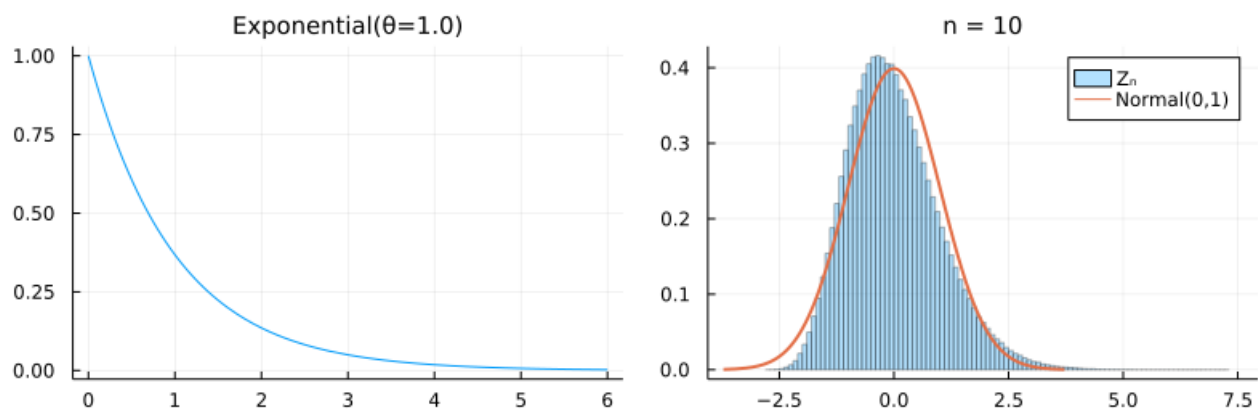
Out[31]:



```
In [32]: 1 # 中心極限定理による収束が遅い場合1
        2 # 左右の非対称性が大きな分布を試してみる
        3 # 指数分布は左右の非対称性が大きな分布になっている
        4 plot_central_limit_theorem(Exponential(), 10)
```

```
skewness(dist) = 2.0
kurtosis(dist) = 6.0
```

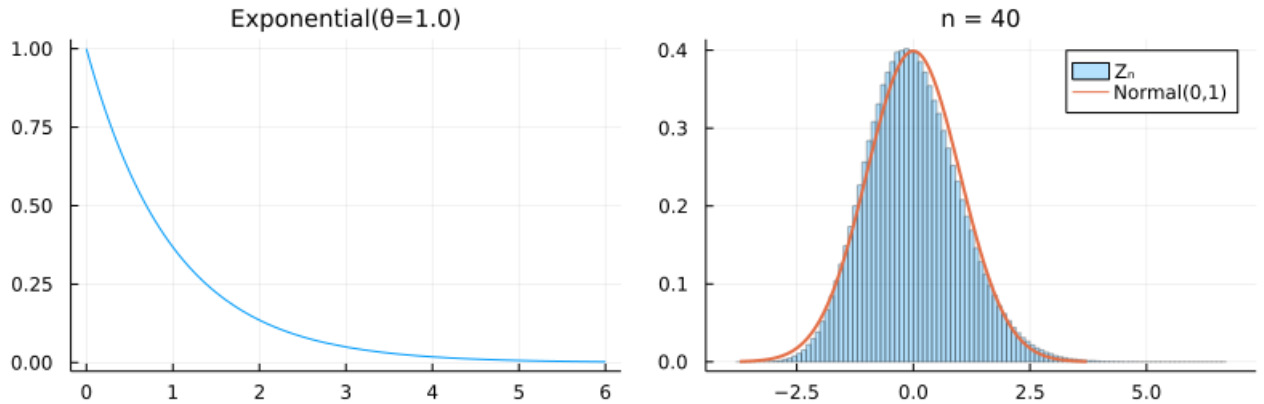
Out[32]:




```
In [33]: 1 plot_central_limit_theorem(Exponential(), 40)
```

```
skewness(dist) = 2.0  
kurtosis(dist) = 6.0
```

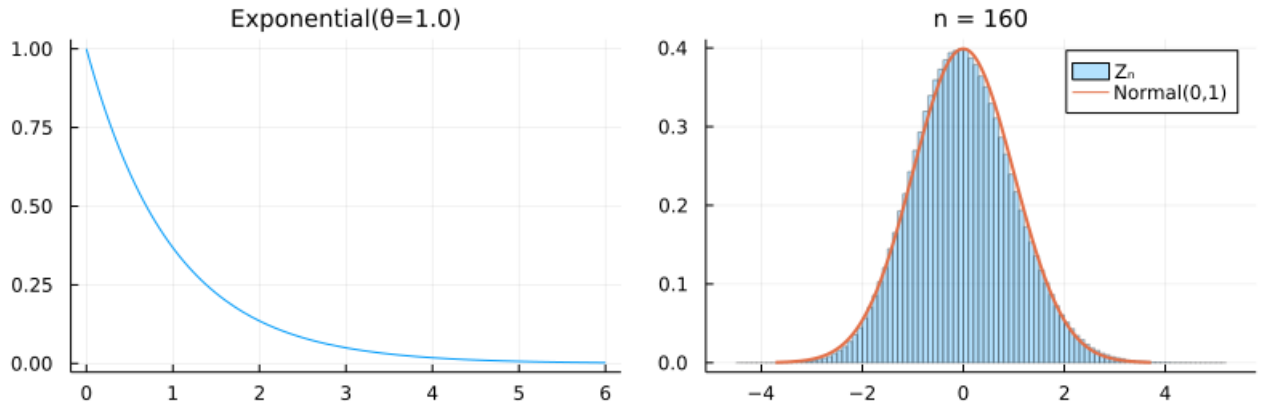
Out[33]:



```
In [34]: 1 plot_central_limit_theorem(Exponential(), 160)
```

```
skewness(dist) = 2.0  
kurtosis(dist) = 6.0
```

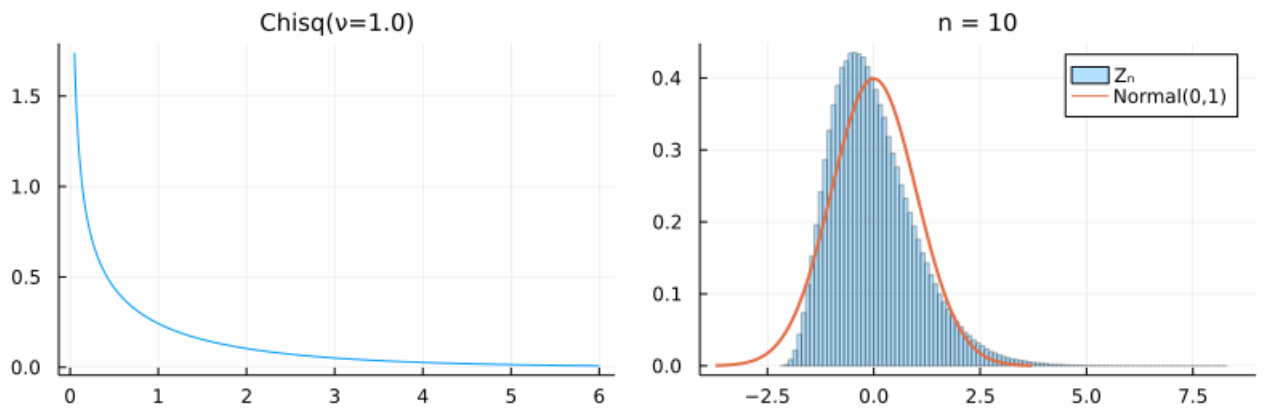
Out[34]:



```
In [35]: 1 # 中心極限定理による収束が遅い場合2  
2 # 自由度1の $\chi^2$ 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  
3 plot_central_limit_theorem(Chisq(1), 10; a=0.05, b=6)
```

```
skewness(dist) = 2.8284271247461903  
kurtosis(dist) = 12.0
```

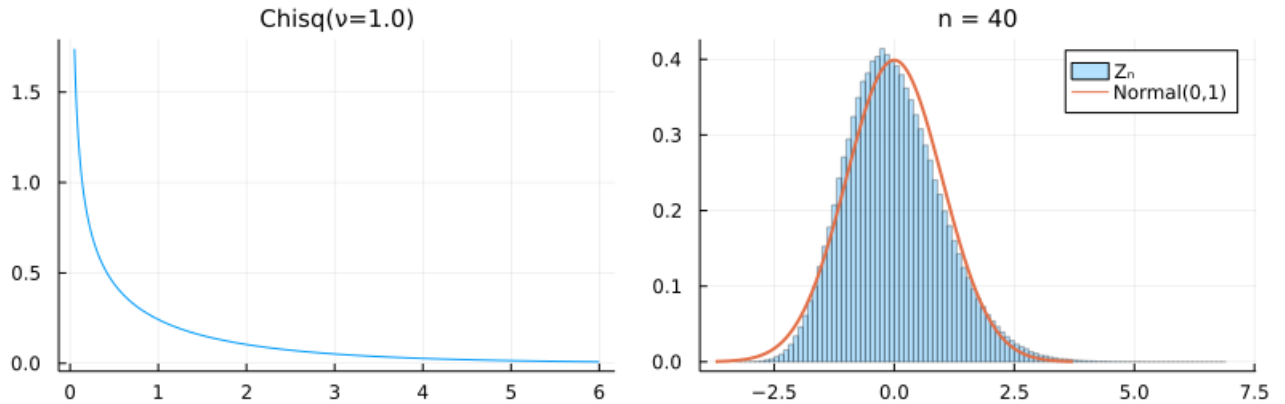
Out[35]:



```
In [36]: 1 plot_central_limit_theorem(Chisq(1), 40; a=0.05, b=6)
```

```
skewness(dist) = 2.8284271247461903  
kurtosis(dist) = 12.0
```

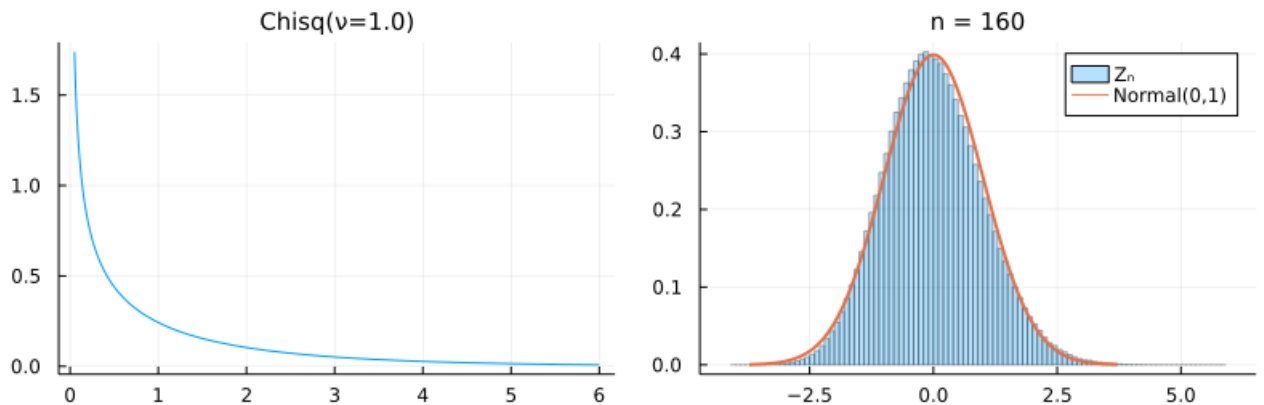
Out[36]:



```
In [37]: 1 plot_central_limit_theorem(Chisq(1), 160; a=0.05, b=6)
```

```
skewness(dist) = 2.8284271247461903  
kurtosis(dist) = 12.0
```

Out[37]:



以下で使う, 分布 $\text{MixtureModel}([\text{Normal}(0, 1), \text{Normal}(20, 1)], [0.95, 0.05])$ の確率密度関数は次の形になる:

$$p(x) = 0.95 \frac{e^{-x^2/2}}{\sqrt{2\pi}} + 0.05 \frac{e^{-(x-20)^2/2}}{\sqrt{2\pi}}.$$

これは標準正規分布に割合が 5% の極端な外れ値を付け加えた分布になっている. 何らかの原因で分布に極端な外れ値が混ざっている場合には中心極限定理を使うときに注意を要する.

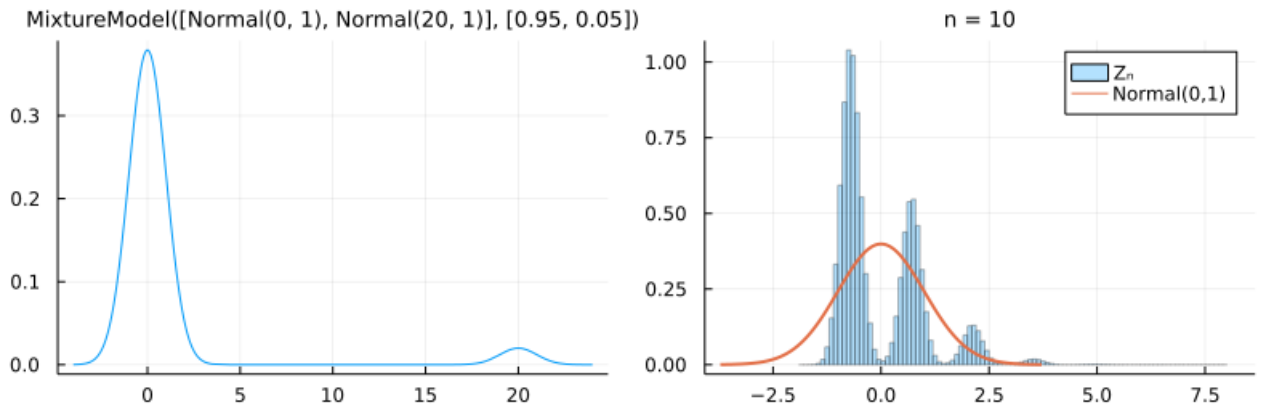
このように確率密度関数が複数の正規分布の確率密度関数の一次結合になっている分布は **混合正規分布** (mixture normal distribution) と呼ばれている.

In [38]:

```
1 # 中心極限定理による収束が遅い場合3
2 # 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
3 dist = MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])
4 disttitle = "MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])"
5 titlefontsize = 9
6 a, b = -4, 24
7 plot_central_limit_theorem(dist, 10; a, b, disttitle, titlefontsize)
```

```
skewness(dist) = 3.8236762415246486
kurtosis(dist) = 13.584999999990963
```

Out[38]:

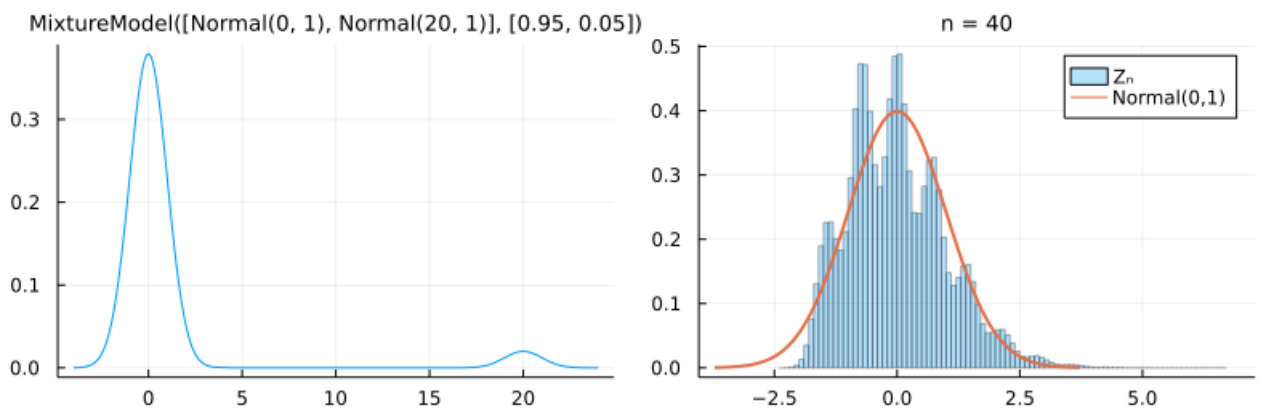


In [39]:

```
1 plot_central_limit_theorem(dist, 40; a, b, disttitle, titlefontsize)
```

```
skewness(dist) = 3.8236762415246486
kurtosis(dist) = 13.584999999990963
```

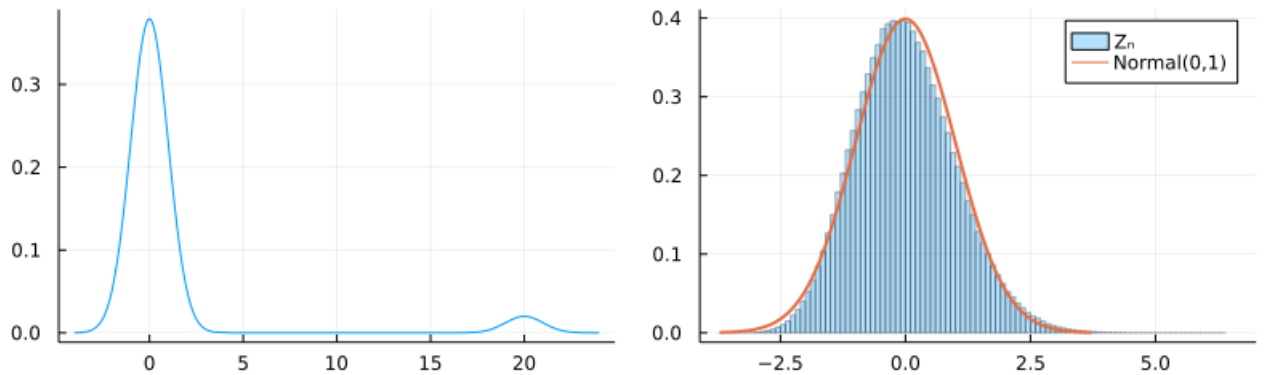
Out[39]:



```
In [40]: 1 plot_central_limit_theorem(dist, 160; a, b, disttitle, titlefontsize)
```

```
skewness(dist) = 3.8236762415246486  
kurtosis(dist) = 13.584999999990963
```

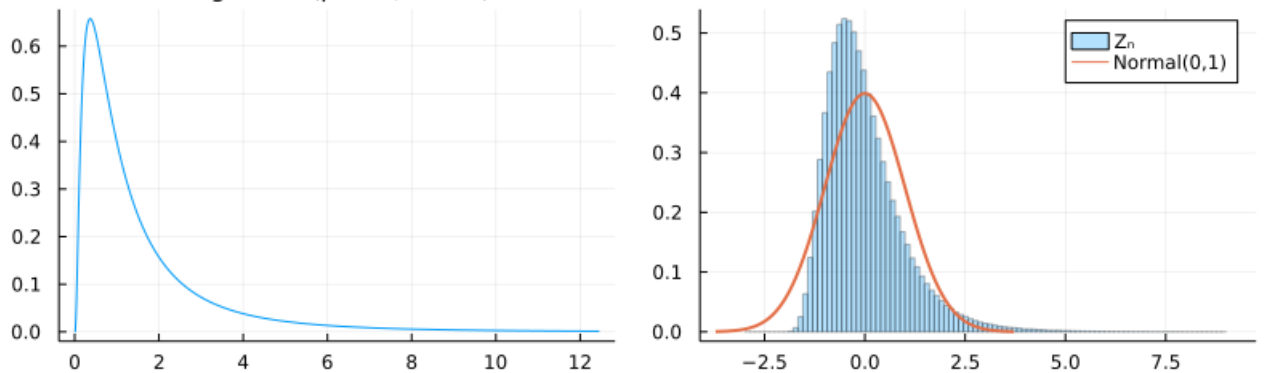
Out[40]: MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])



```
In [41]: 1 # 中心極限定理による収束が遅い場合4  
2 # 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  
3 # 右側の裾が太く、外れ値が出やすい  
4 plot_central_limit_theorem(LogNormal(), 10; bin=range(-3, 9, 100))
```

```
skewness(dist) = 6.184877138632554  
kurtosis(dist) = 110.9363921763115
```

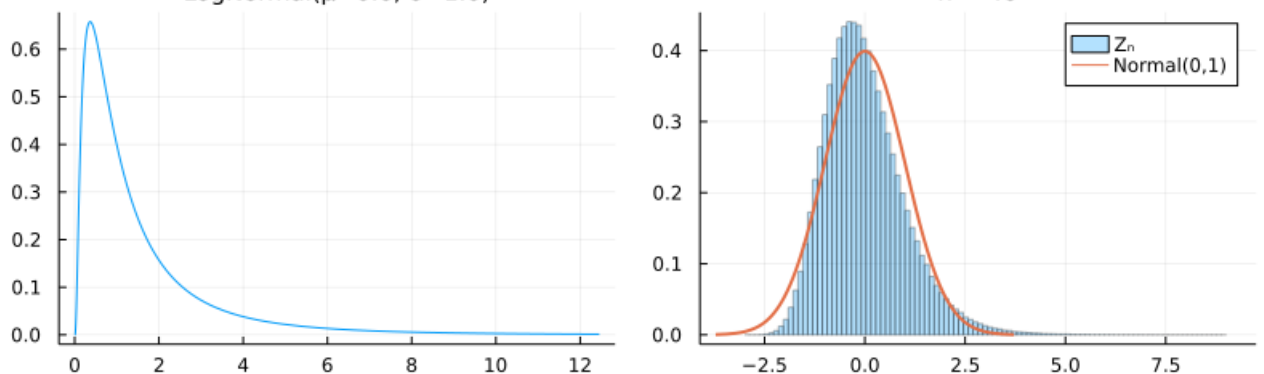
Out[41]: LogNormal($\mu=0.0$, $\sigma=1.0$)



```
In [42]: 1 plot_central_limit_theorem(LogNormal(), 40; bin=range(-3, 9, 100))
```

```
skewness(dist) = 6.184877138632554  
kurtosis(dist) = 110.9363921763115
```

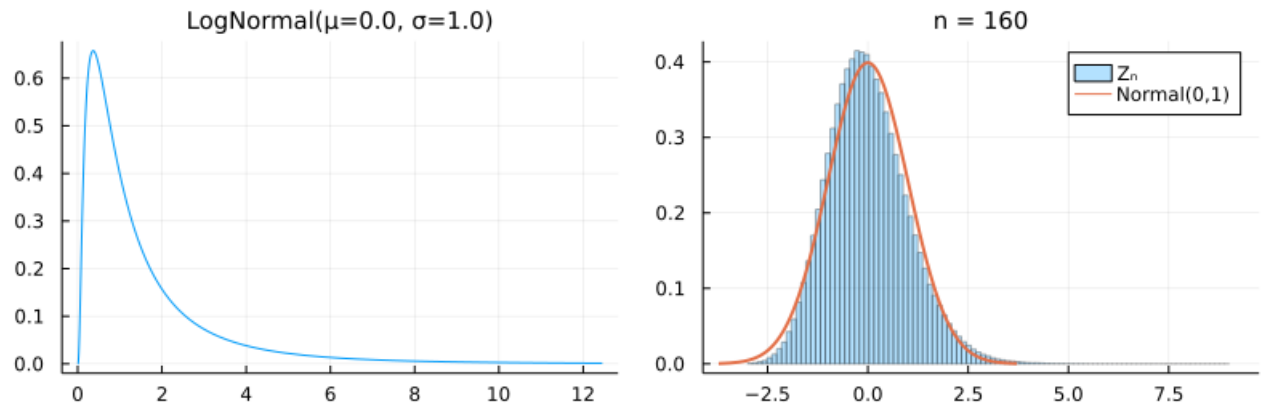
Out[42]: LogNormal($\mu=0.0$, $\sigma=1.0$)



```
In [43]: 1 plot_central_limit_theorem(LogNormal(), 160; bin=range(-3, 9, 100))
```

```
skewness(dist) = 6.184877138632554  
kurtosis(dist) = 110.9363921763115
```

Out[43]:



3.8 問題: デルタ法 (実は単なる一次近似)

```
In [ ]: 1
```