

中心極限定理などについて

- 黒木玄
- 2022-04-11~2022-04-29

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

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

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

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

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

目次

- 1 大数の法則
- ▼ 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
- ▼ 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
 - 3.7 中心極限定理の収束の速さと歪度と尖度(再)
 - 3.8 問題: 中心極限定理の収束の様子のグラフ
 - 3.9 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
```

1 大数の法則

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

二項分布が正規分布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。

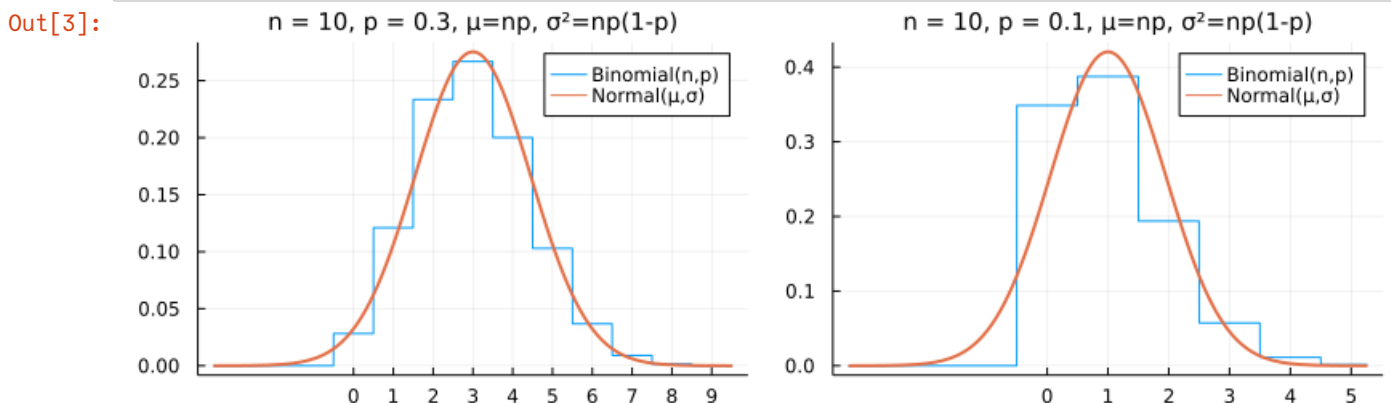
以下は、試行回数 n 、成功確率 p の二項分布(期待値と分散はそれぞれ np , $np(1-p)$ になる)の確率質量函数と平均 np 、分散 $np(1-p)$ を持つ正規分布の確率密度函数の同時プロットである。それらを見れば、証明しなくても、二項分布の中心極限定理が成立していることは明らかだろう。(証明する前にコンピュータでグラフを確認するべき!)

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。)

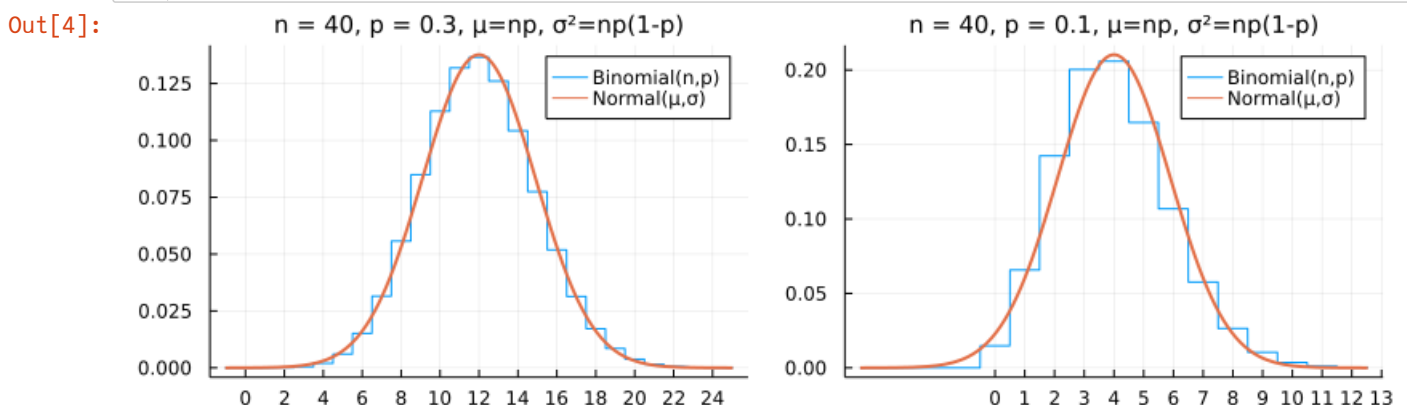
```
In [2]: 1 function plot_binomial_clt(n, p, s = 1; c = 4.5)
2         μ, σ = n*p, √(n*p*(1-p))
3         xlim = (μ-c*σ, μ+c*σ)
4         plot(x → pdf(Binomial(n, p), round(x)), xlim...; label="Binomial(n,p)")
5         plot!(Normal(μ, σ), xlim...; label="Normal(μ,σ)", lw=2)
6         # plot!(x → pdf(Poisson(μ), 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[2]: plot_binomial_clt (generic function with 2 methods)

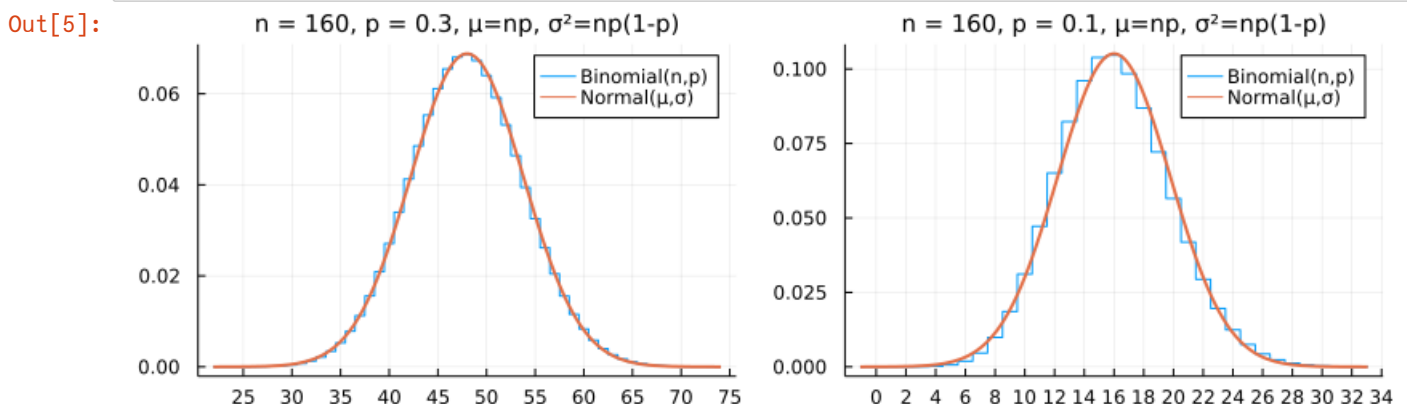
```
In [3]: 1 plot(plot_binomial_clt(10, 0.3), plot_binomial_clt(10, 0.1); size=(800, 250))
```



```
In [4]: 1 plot(plot_binomial_clt(40, 0.3, 2), plot_binomial_clt(40, 0.1); size=(800, 250))
```



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



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

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

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, $K_n \sim \text{Binomial}(n, p)$ 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 $k \rightarrow \infty$ だけではなく, $n - k \rightarrow \infty$ 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 $\sqrt{2\pi n p(1-p)}$ 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

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」のような意味ではなく、「確率論における中心的な極限定理」というような意味である。

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関数 $f(x)$ 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。(関数 $f(x)$ としてどのような関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。) 統計学で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関数を使った証明

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 の特性関数 $\varphi(t)$ 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 中心極限定理のキウ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 のキウ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 のキウムラント母函数 $K(t)$ 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 のキウムラント母函数は, $n \rightarrow \infty$ 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ウムラント母函数 $t^2/2$ 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: このようにキウ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,

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

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 の標準化のキウムラント母函数 $K(t)$ 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,

$$nK \left(\frac{t}{\sqrt{n}} \right) = n \left(\frac{t^2}{2n} + \bar{\kappa}_3 \frac{t^3}{3! n^{3/2}} + \bar{\kappa}_4 \frac{t^4}{4! n^2} + O(n^{-5/2}) \right) = \frac{t^2}{2} + \bar{\kappa}_3 \frac{t^3}{3! n^{1/2}} + \bar{\kappa}_4 \frac{t^4}{4! n} + O(n^{-3/2})$$

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

3.7 中心極限定理の収束の速さと歪度と尖度(再)

3.8 問題: 中心極限定理の収束の様子グラフ

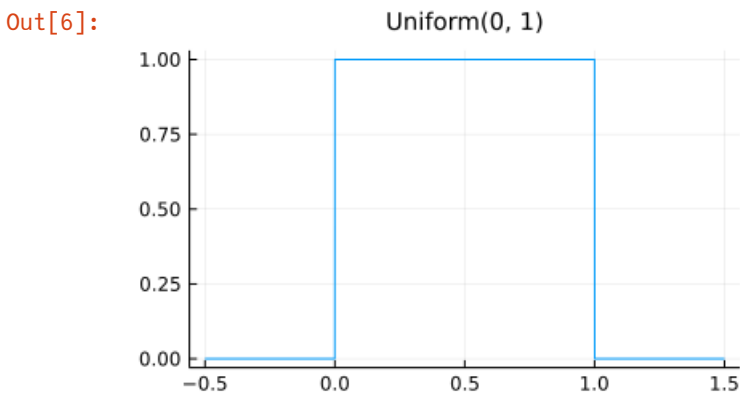
中心極限定理による正規分布への収束の様子をコンピュータでグラフを描いて確認せよ。収束が速い場合($n = 10$ ですでに正規分布に十分に近い場合)と遅い場合の両方の例を作れ。

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 [6]: 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 [7]: 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)
```

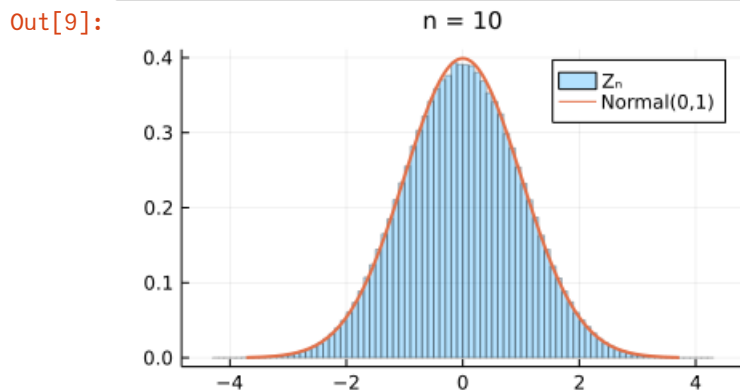
```
Out[7]: 10x1000000 Matrix{Float64}:
0.676799 0.805504 0.935768 0.947193 ... 0.633127 0.259868 0.0311475 0.672738
0.256619 0.481629 0.121879 0.490844 ... 0.595994 0.931575 0.442106 0.348156
0.826526 0.41462 0.905168 0.591054 ... 0.474323 0.865503 0.75942 0.0741175
0.220369 0.286867 0.0568799 0.135445 ... 0.827834 0.633611 0.289928 0.927236
0.18158 0.927432 0.656976 0.0526441 ... 0.987021 0.492324 0.093395 0.518172
0.159018 0.0227525 0.0700251 0.474866 ... 0.72756 0.893995 0.037462 0.246558
0.541567 0.614833 0.309085 0.275011 ... 0.181238 0.475485 0.936195 0.84376
0.976197 0.727598 0.119523 0.766493 ... 0.0903061 0.801676 0.788165 0.0638446
0.362752 0.658416 0.426256 0.537249 ... 0.499248 0.244568 0.664221 0.0784209
0.0343311 0.673652 0.866703 0.916254 ... 0.174507 0.702053 0.143359 0.809005
```



```
In [8]: 1 #  $Z_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$  を大量に計算
2  $\mu$  = mean(dist)
3  $\sigma$  = std(dist)
4 Zns = [ $\sqrt{n} * (\text{mean}(X) - \mu) / \sigma$  for X in eachcol(Xs)] # Zns は  $Z_n$  達という意味
5 first(Zns, 5)
```

```
Out[8]: 5-element Vector{Float64}:
-0.8371864051509108
 0.6718409315870623
-0.5824878305721526
 0.204906668905459
 0.450280311050892
```

```
In [9]: 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 histogram(Zns; norm=true, alpha=0.3, bin=100, label=" $Z_n$ ")
3 plot!(Normal(0, 1); label="Normal(0,1)", lw=2)
4 title!("n = $n")
```



In [10]:

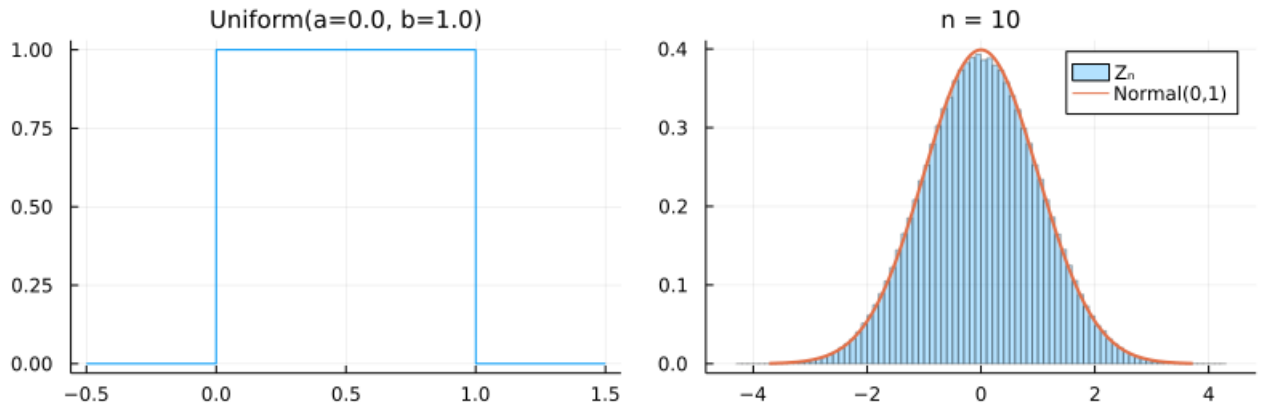
```
1 # 以上の手続きを函数化
2 # さらに
3 # * メモリアロケーションの節約と並列化による最適化
4 # * 離散分布の場合にも対応
5 # * 歪度と尖度を混合モデルの場合にも表示
6
7 distname(dist::Distribution) = replace(string(dist), r"{.*}" => "")
8 myskewness(dist) = skewness(dist)
9 mykurtosis(dist) = kurtosis(dist)
10 function standardized_moment(dist::ContinuousUnivariateDistribution, 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[10]: plot_central_limit_theorem (generic function with 1 method)

```
In [11]: 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
```

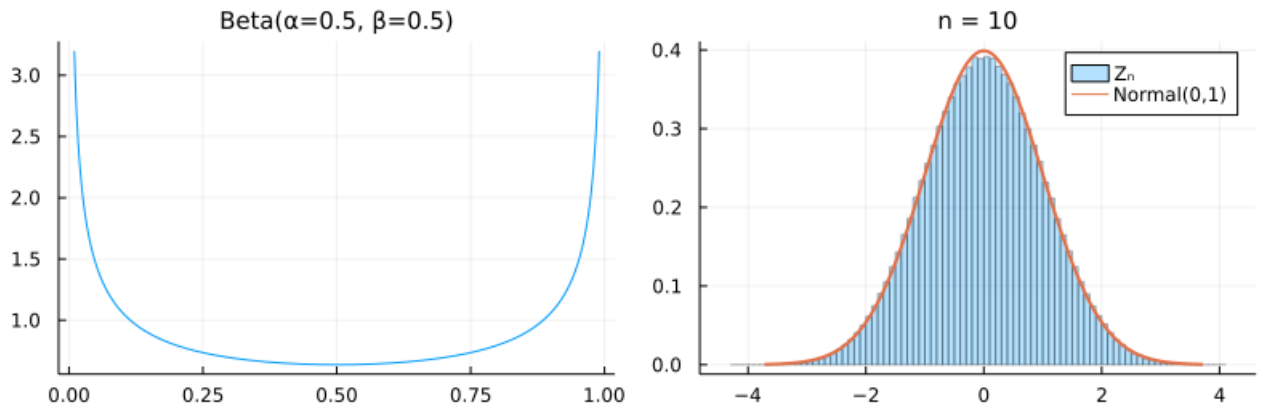
Out[11]:



```
In [12]: 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
```

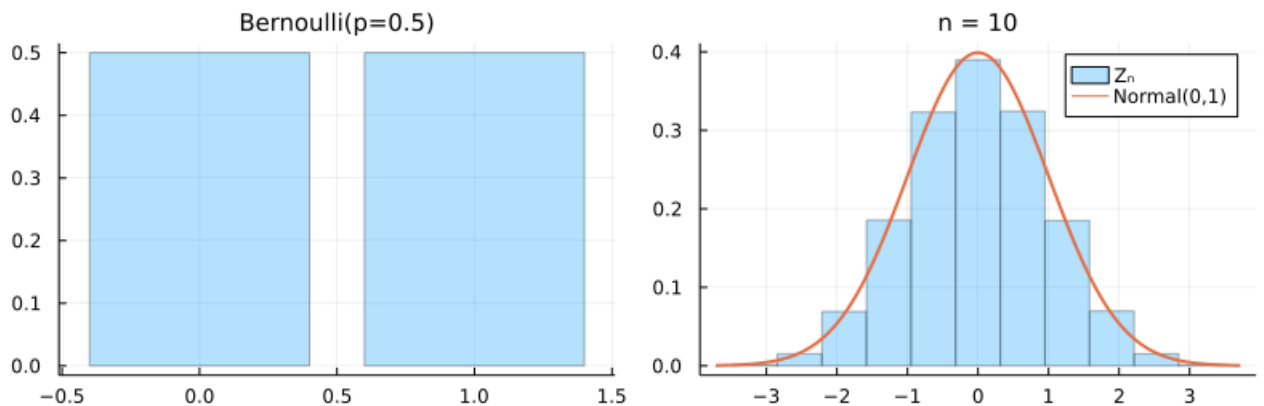
Out[12]:



```
In [13]: 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
```

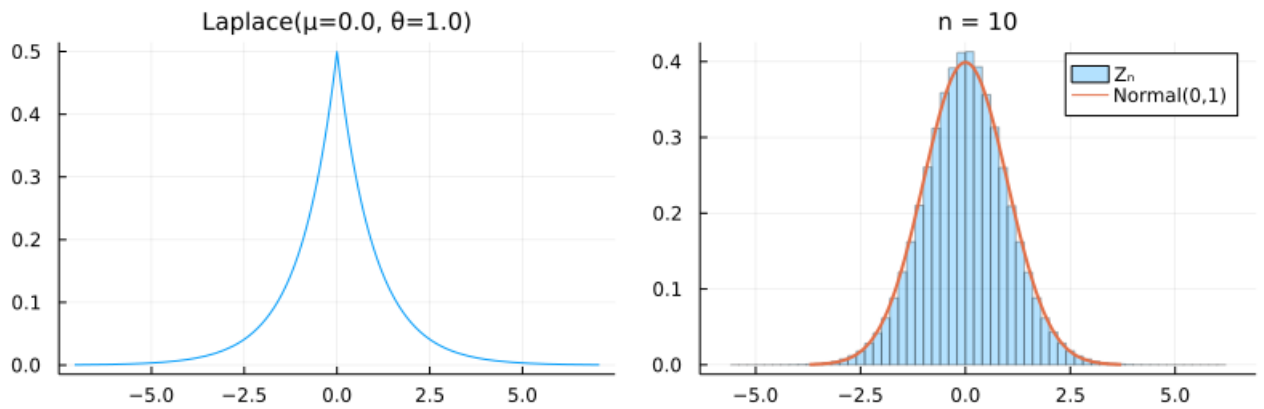
Out[13]:



```
In [14]: 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
```

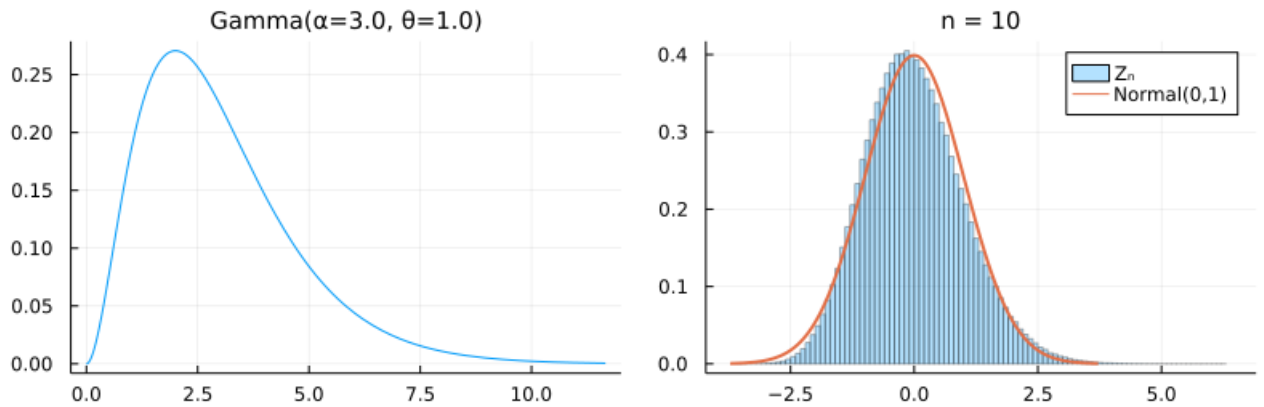
Out[14]:



```
In [15]: 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
```

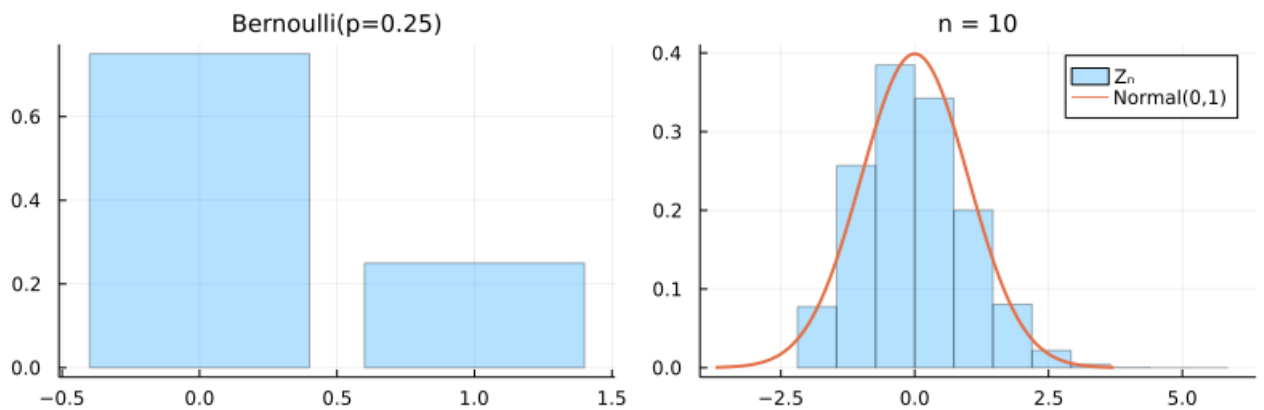
Out[15]:



```
In [16]: 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
```

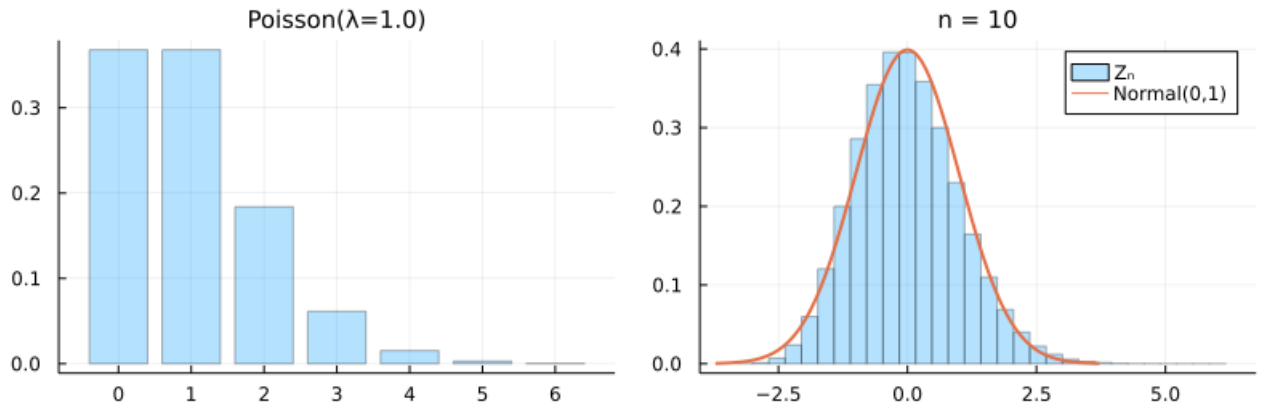
Out[16]:



```
In [17]: 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
```

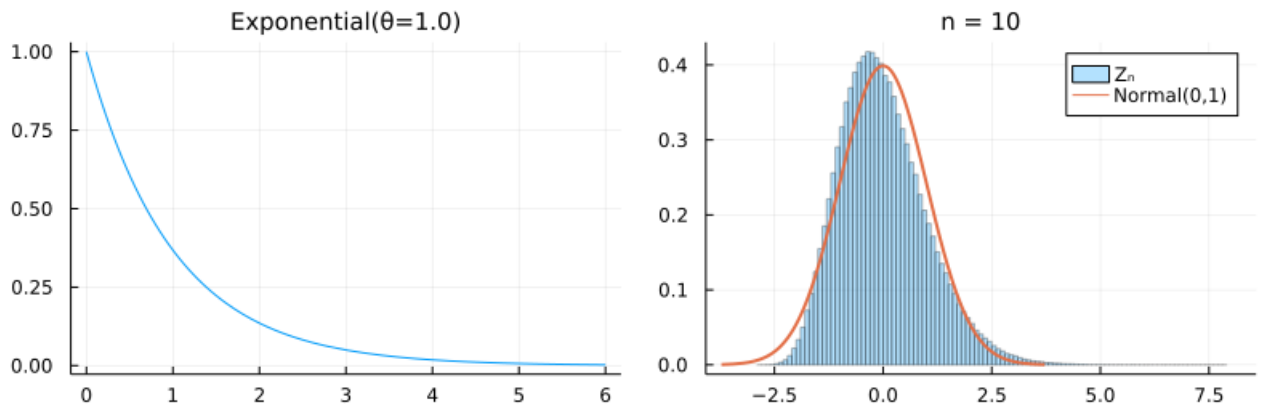
Out[17]:



```
In [18]: 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
```

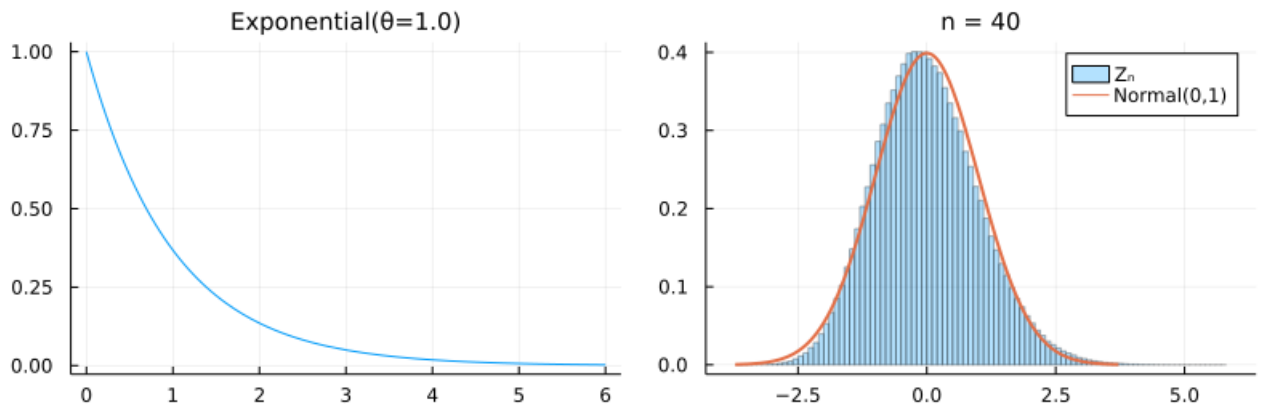
Out[18]:



```
In [19]: 1 plot_central_limit_theorem(Exponential(), 40)
```

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

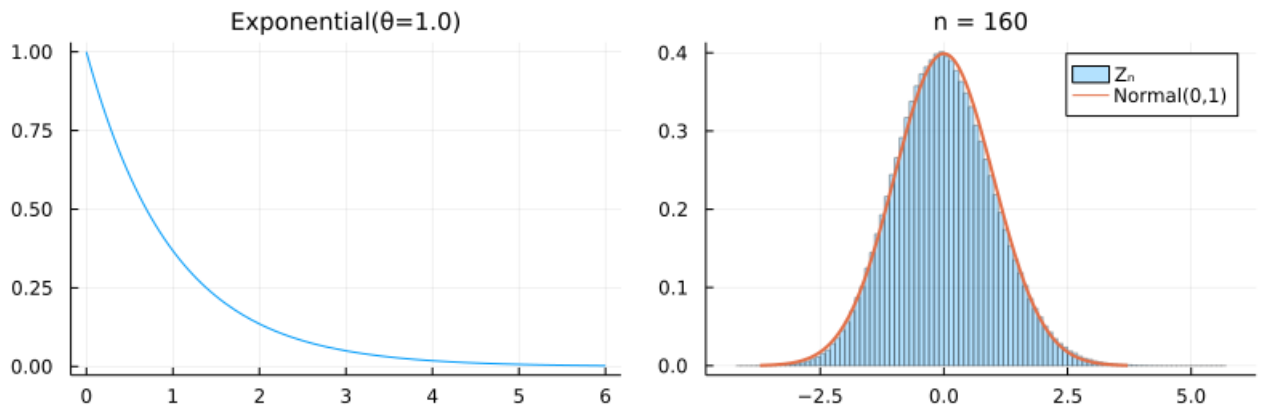
Out[19]:



```
In [20]: 1 plot_central_limit_theorem(Exponential(), 160)
```

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

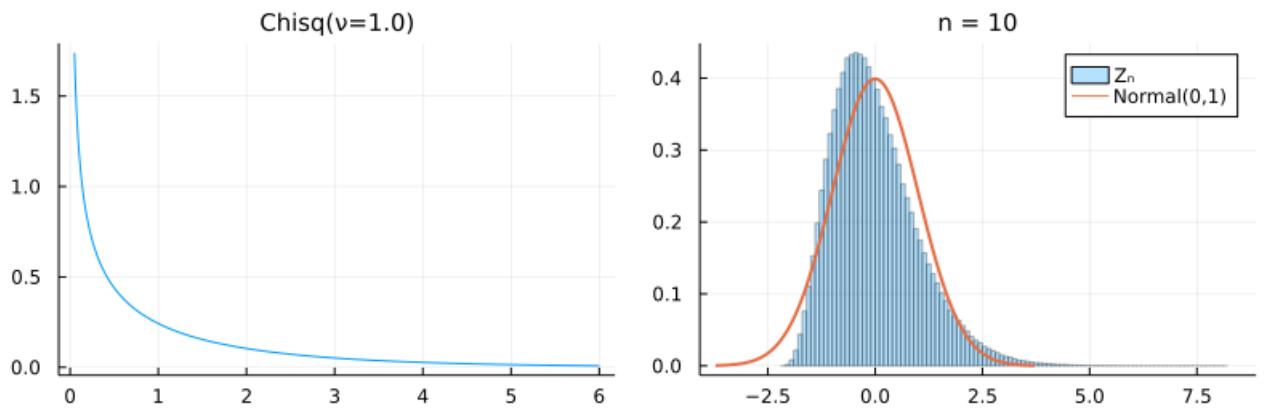
Out[20]:



```
In [21]: 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
```

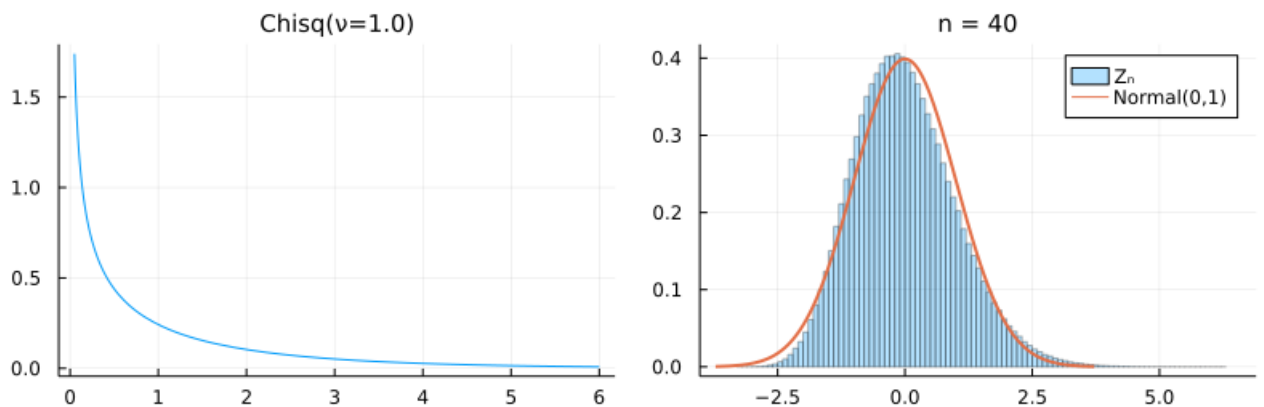
Out[21]:



```
In [22]: 1 plot_central_limit_theorem(Chisq(1), 40; a=0.05, b=6)
```

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

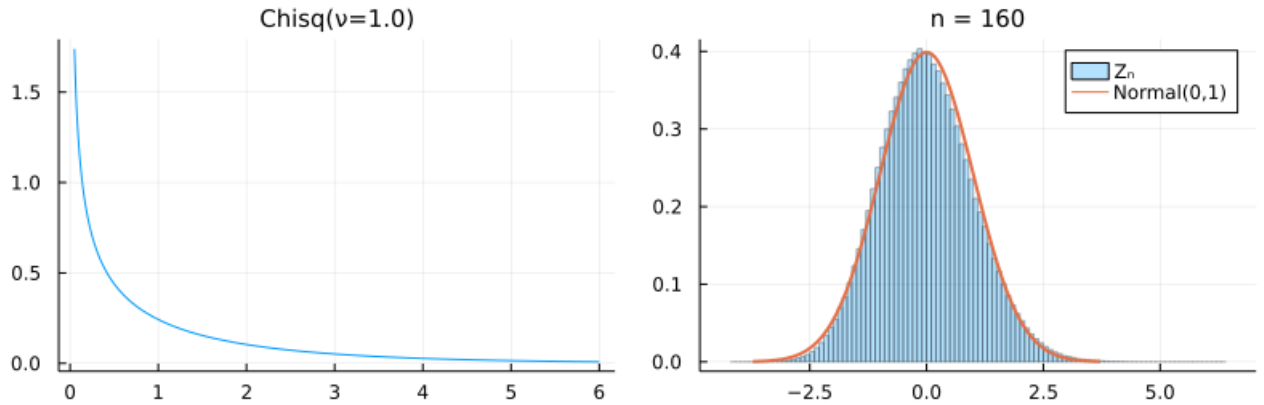
Out[22]:



```
In [23]: 1 plot_central_limit_theorem(Chisq(1), 160; a=0.05, b=6)
```

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

Out[23]:



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:

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

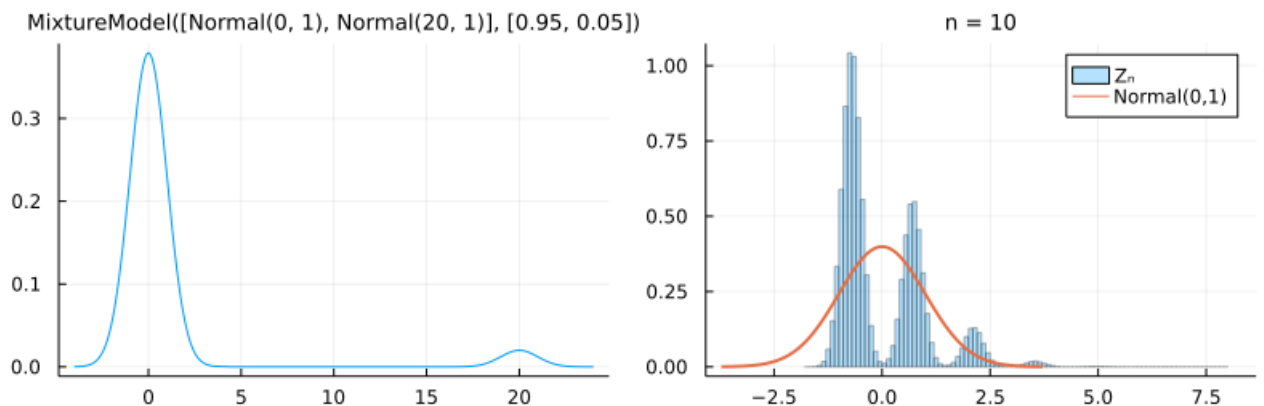
これは標準正規分布に割合が 5% の極端な外れ値を付け加えた分布になっている. 何らかの原因で分布に極端な外れ値が混ざっている場合には中心極限定理を使うときに注意を要する.

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 [24]: 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
```

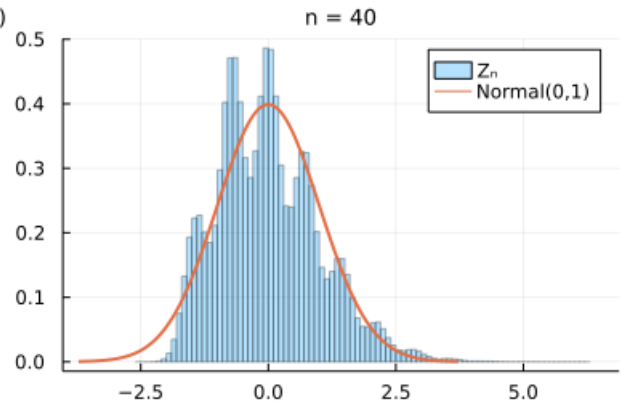
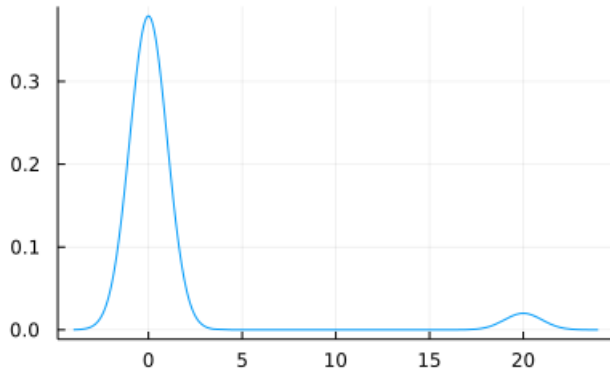
Out[24]:



```
In [25]: 1 plot_central_limit_theorem(dist, 40; a, b, disttitle, titlefontsize)
```

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

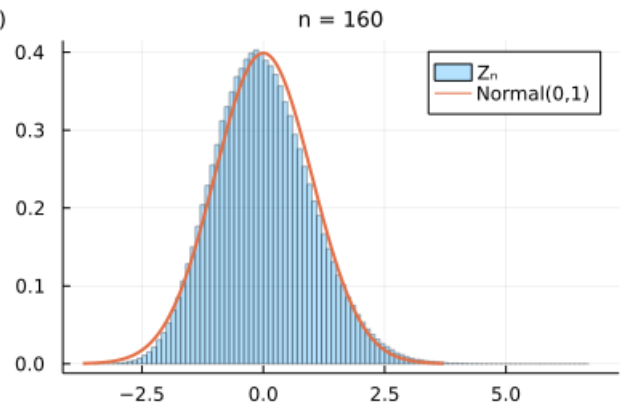
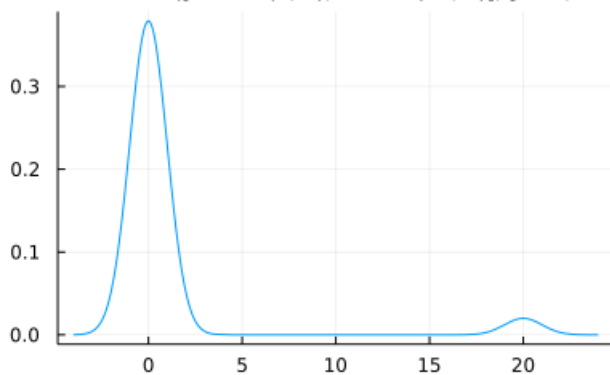
Out[25]: MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])



```
In [26]: 1 plot_central_limit_theorem(dist, 160; a, b, disttitle, titlefontsize)
```

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

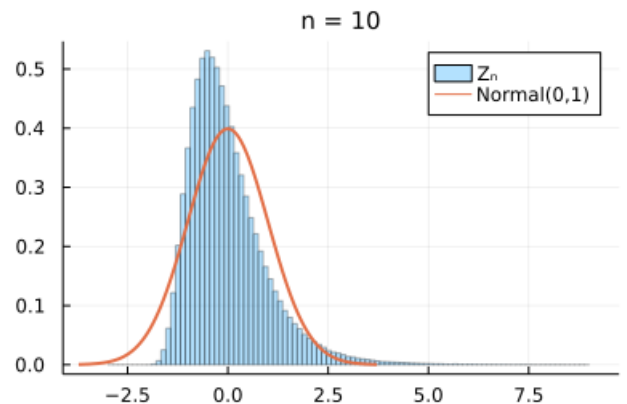
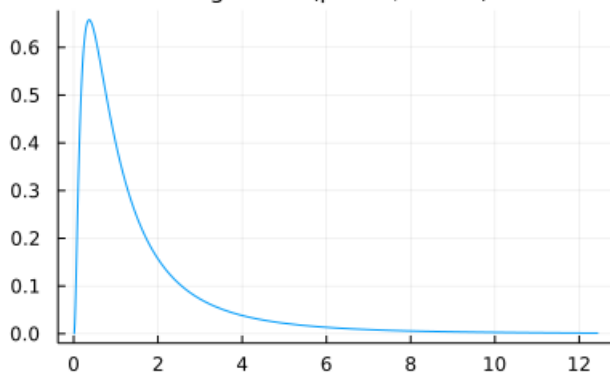
Out[26]: MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])



```
In [27]: 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
```

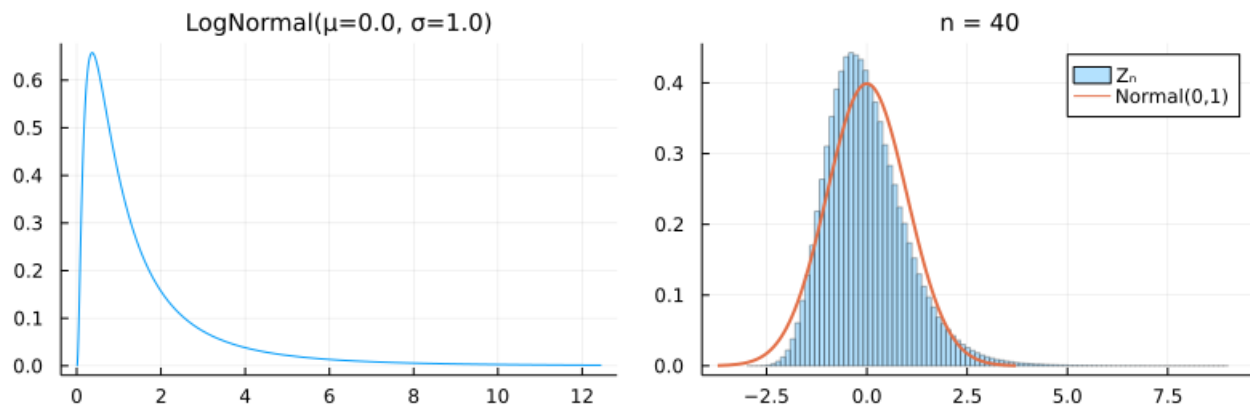
Out[27]: LogNormal($\mu=0.0$, $\sigma=1.0$)




```
In [28]: 1 plot_central_limit_theorem(LogNormal(), 40; bin=range(-3, 9, 100))
```

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

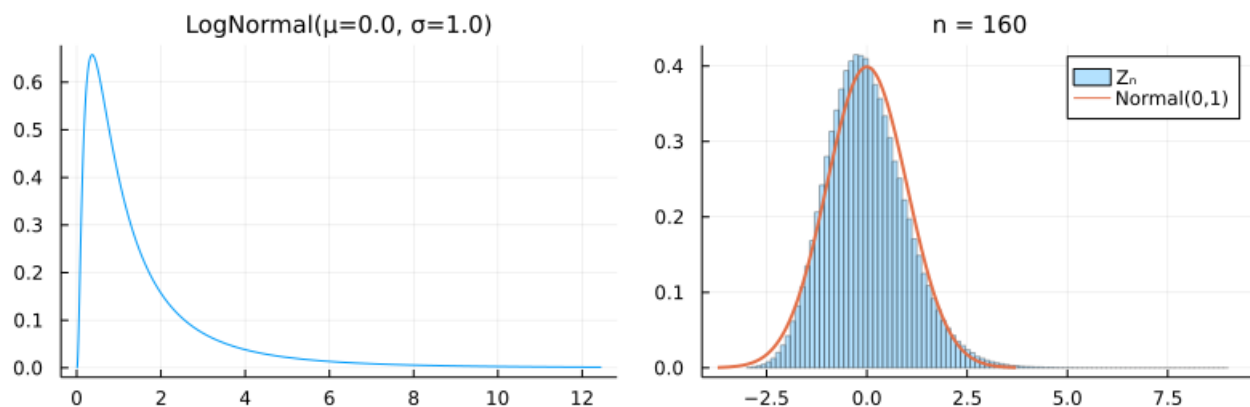
Out[28]:



```
In [29]: 1 plot_central_limit_theorem(LogNormal(), 160; bin=range(-3, 9, 100))
```

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

Out[29]:



3.9 問題: デルタ法 (実は単なる一次近似)

```
In [ ]: 1
```