

中心極限定理などについて

- 黒木玄
- 2022-04-11~2022-04-12

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

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

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

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

目次

[1 大数の法則](#)

▼ [2 モーメントとその母関数と特性関数とキュムラント母関数](#)

- [2.1 モーメントとその母関数と特性関数とキュムラント母関数の定義](#)
- [2.2 特性関数による期待値の表示](#)
- [2.3 問題: キュムラントのロケーションスケール変換](#)
- [2.4 問題: 標準正規分布のモーメント母関数と特性関数とキュムラント母関数](#)
- [2.5 確率変数の標準化と標準化キュムラントと歪度と尖度](#)
- [2.6 問題: 独立な確率変数達の和のモーメント母関数と特性関数とキュムラント母関数](#)

▼ [3 中心極限定理](#)

- [3.1 中心極限定理の特性関数を使った証明](#)
- [3.2 中心極限定理の収束の速さと歪度](#)
- [3.3 中心極限定理のキュムラント母関数を使った証明](#)
- [3.4 中心極限定理の収束の速さと歪度と尖度](#)
- [3.5 問題: 中心極限定理の収束の様子](#)
- [3.6 標本平均と不偏分散の定義](#)
- [3.7 問題: 標本平均の期待値と分散](#)
- [3.8 問題: 不偏分散の期待値と分散](#)

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

1 大数の法則

2 モーメントとその母関数と特性関数とキュムラント母関数

2.1 モーメントとその母関数と特性関数とキュムラント母関数の定義

確率変数 X と $m = 0, 1, 2, \dots$ について

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

を X の m 次の **モーメント**(moment, 積率) と呼び、

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

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

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

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

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

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

モーメント母関数だけではなく、特性関数もモーメント達の母関数になっている。

モーメント母関数の対数

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

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

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

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

2.2 特性関数による期待値の表示

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

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

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

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

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

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

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

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

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

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

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

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

2.3 問題: キュムラントのロケーションスケール変換

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

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

注意: キュムラントの変換公式は非常に単純な形になる. $\kappa_1(aX+b) = a\kappa_1(X) + b$ は $\kappa_1(X) = E[X]$ だったので当然である. 2 次以上のキュムラントは a^m 倍されるだけになる. ここにもわざわざ対数を取ってキュムラント母関数とキュムラントを定義することの利点が現れている.

注意: この結果は空気のごとく使われる.

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

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

ゆえに, $aX+b$ のキュムラントは次の形になる:

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

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

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

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

$K_{aX+b}(t)$ の展開結果は

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

になるので,

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

となることがわかる.

解答終

2.4 問題: 標準正規分布のモーメント母関数と特性函数とキュムラント母関数

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

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

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

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

4つめの等号で $z = w + t$ とおいた. ゆえに,

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

解答終

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

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

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

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

ゆえに

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

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

2.5 確率変数の標準化と標準化キュムラントと歪度と尖度

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

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

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

X の標準化のモーメントやキュムラントをそれぞれ **標準化モーメント**, **標準化キュムラント** と呼び、それぞれを $\mu'_m(X)$, $\kappa'_m(X)$ と表す。詳しくは以下の通り:

$$\begin{aligned}\mu'_m(X) &= \mu'_m(Z) = E \left[\left(\frac{X - \mu}{\sigma} \right)^m \right], \\ M_Z(t) &= E \left[\exp \left(t \frac{X - \mu}{\sigma} \right) \right] = \sum_{m=0}^{\infty} \mu'_m(X) \frac{t^m}{m!} = 1 + \frac{t^2}{2} + \mu'_3(X) \frac{t^3}{3!} + \mu'_4(X) \frac{t^4}{4!} + \dots, \\ K_Z(t) &= \log M_Z(t) = \sum_{m=1}^{\infty} \kappa'_m(X) \frac{t^m}{m!} = \frac{t^2}{2} + \kappa'_3(X) \frac{t^3}{3!} + \kappa'_4(X) \frac{t^4}{4!} + \dots.\end{aligned}$$

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

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

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

$$\begin{aligned}\log \left(1 + \frac{t^2}{2} + \mu'_3(X) \frac{t^3}{3!} + \mu'_4(X) \frac{t^4}{4!} + O(t^5) \right) &= \frac{t^2}{2} + \mu'_3(X) \frac{t^3}{3!} + \mu'_4(X) \frac{t^4}{4!} - \frac{1}{2} \left(\frac{t^2}{2} \right)^2 + O(t^5) \\ &= \frac{t^2}{2} + \mu'_3(X) \frac{t^3}{3!} + (\mu'_4(X) - 3) \frac{t^4}{4!} + O(t^5).\end{aligned}$$

$\kappa'_3(X)$ を X もしくは X が従う分布の **歪度** (わいど, skewness) と呼び、 $\kappa'_4(X)$ を **尖度** (せんど, kurtosis) と呼び、以下のように書くことにする:

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

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

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

ゆえに、3 次以上の標準化キュムラントは分布が正規分布からどれだけ離れているかを表していると考えられる。その最初の2つが上で定義した歪度 $\kappa'_3(X)$ と尖度 $\kappa'_4(X)$ である。

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

注意: $\kappa'_4(X) = \mu'_4(X) - 3$ ではなく、3 を引く前の $\mu'_4(X)$ を尖度と定義する流儀もあるが、このノートでは **正規分布の尖度を 0 にしたので、3 を引いた側の κ'_4 を尖度の定義として採用する。**

注意: $\kappa'_4(X) = \mu'_4(X) - 3$ は **過剰尖度 (excess kurtosis)** と呼ばれることも多い。正規分布の尖度を κ'_4 ではなく、 μ'_4 の方の 3 とするときに、そこからどれだけ分布の尖り具合が増したかを $\kappa'(X)$ が表しているため、過剰尖度と呼ぶのである。

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

```
Out[2]: t^2/2 + t^3 μ3/6 + t^4 (μ4/24 - 1/8) + t^5 (-μ3/12 + μ5) + O(t^6)
```

2.6 問題: 独立な確率変数達の和のモーメント母関数と特性関数とキウムラント母関数

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

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

注意: この結果は空気のごとく使われる.

解答例: 独立な確率変数達 X_1, \dots, X_n について

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

が成立することより,

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

ゆえに

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

解答終

3 中心極限定理

3.1 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

3.2 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かかる。

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

とおくと、

$$\varphi\left(\frac{t}{\sqrt{n}}\right) = E[e^{itY_k}] = 1 - \frac{t^2}{2n} - i\mu'_3 \frac{t^3}{6n\sqrt{n}} + O(n^{-2})$$

なので、

$$\log \varphi\left(\frac{t}{\sqrt{n}}\right)^n = n \log\left(1 - \frac{t^2}{2n} - i\mu'_3 \frac{t^3}{6n\sqrt{n}} + O(n^{-2})\right) = -\frac{t^2}{2} - i\mu'_3 \frac{t^3}{6\sqrt{n}} + O(n^{-1}).$$

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 μ'_3 の絶対値の大ききさで大体決まっていることがわかる。 μ'_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。

μ'_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。 μ'_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:

$$K_{Y_k}(t) = \log E[e^{itY_k}] = \log\left(1 + \frac{t^2}{2} + \mu'_3 \frac{t^3}{3!} + O(t^4)\right) = \frac{t^2}{2} + \mu'_3 \frac{t^3}{3!} + O(t^4).$$

ここでの $t^3/3!$ の係数 $\kappa'_3 = \mu'_3$ は X_k の **歪度** (skewness) と呼ばれるのであった。

3.3 中心極限定理のキュムラント母関数を使った証明

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、 $E[(X_k - \mu)^3] < \infty$ となっており、さらに各 X_k のキュムラント母関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 のキュムラント母関数 $K(t)$ は k によらず、

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

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 のキュムラント関数は、 $n \rightarrow \infty$ 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.

3.4 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,

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

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 の標準化のキウムラント母函数は X_k の歪度(わいど) κ'_3 と尖度(せんど)は

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

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:

$$K(t) = \frac{t^2}{2} + \kappa'_3 \frac{t^3}{3!} + \kappa'_4 \frac{t^4}{4!} + O(t^5).$$

このとき,

$$nK \left(\frac{t}{\sqrt{n}} \right) = n \left(\frac{t^2}{2n} + \kappa'_3 \frac{t^3}{3! n^{3/2}} + \kappa'_4 \frac{t^4}{4! n^2} + O(n^{-5/2}) \right) = \frac{t^2}{2} + \kappa'_3 \frac{t^3}{3! n^{1/2}} + \kappa'_4 \frac{t^4}{4! n} + O(n^{-3/2})$$

これが $t^2/2$ に収束する速さは $\kappa'_3 \neq 0$ ならば $O(n^{-1/2})$ のオーダーになり, 歪度 κ'_3 の絶対値が大きいほど遅くなる. そして, $\kappa'_3 = 0$ ならば $O(n^{-1})$ のオーダーでの収束になり, 尖度 κ'_4 の絶対値が大きいほど遅くなる.

3.5 問題: 中心極限定理の収束の様子

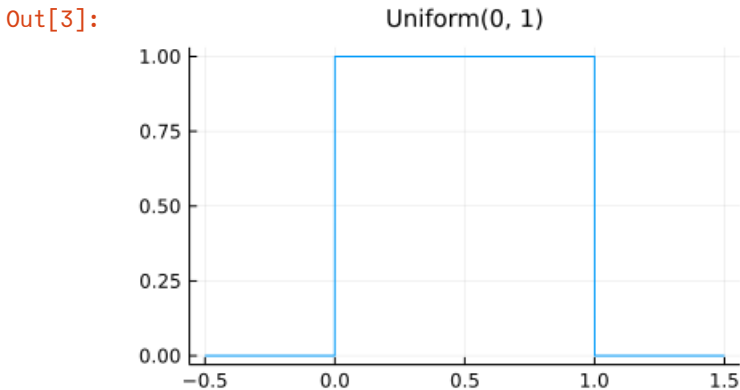
中心極限定理による正規分布への収束の様子をコンピュータでグラフを描いて確認せよ. 収束が速い場合($n = 10$ ですでに正規分布に十分に近い場合)と遅い場合の両方の例を作れ.

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 [3]: 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 [4]: 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[4]: 10×1000000 Matrix{Float64}:

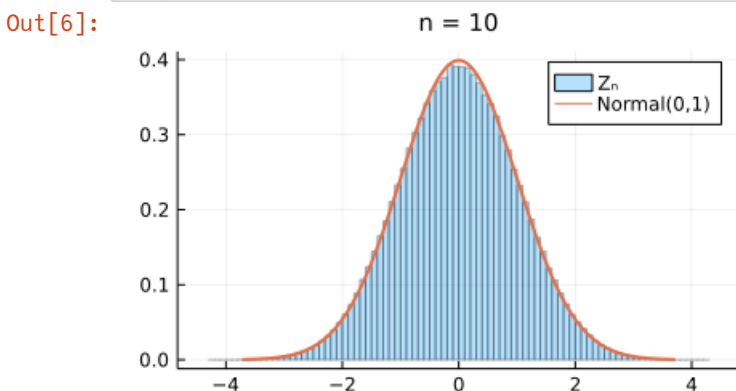
0.676799	0.805504	0.935768	0.947193	...	0.633127	0.259868	0.0311475	0.672738
0.256619	0.481629	0.121879	0.490844		0.595994	0.931575	0.442106	0.348156
0.826526	0.41462	0.905168	0.591054		0.474323	0.865503	0.75942	0.0741175
0.220369	0.286867	0.0568799	0.135445		0.827834	0.633611	0.289928	0.927236
0.18158	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 [5]: 1 # Z_n = sqrt(n)(X_bar_n - mu)/sigma を大量に計算
2 mu = mean(dist)
3 sigma = std(dist)
4 Zns = [sqrt(n)*(mean(X) - mu)/sigma for X in eachcol(Xs)] # Zns はZ_n達という意味
5 first(Zns, 5)
```

Out[5]: 5-element Vector{Float64}:

```
-0.8371864051509108
 0.6718409315870623
-0.5824878305721526
 0.204906668905459
 0.450280311050892
```

```
In [6]: 1 # Z_n達のヒストグラムと標準正規分布の密度関数を比較
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 [7]:

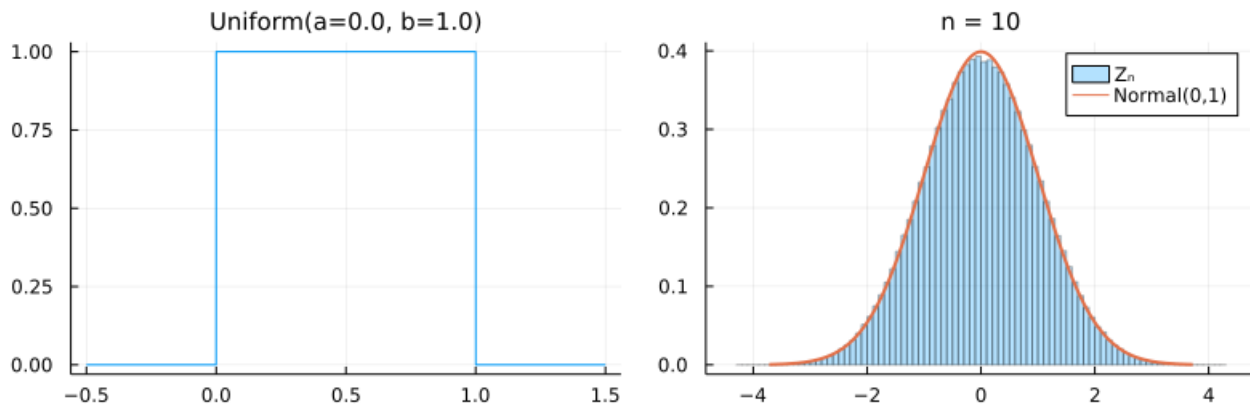
```
1 # 以上の手続きを函数化
2 # さらに
3 # * メモリアロケーションの節約と並列化による最適化
4 # * 離散分布の場合にも対応
5 # * 歪度と尖度を混合モデルの場合にも表示
6
7 distname(dist::Distribution) = replace(string(dist), r"{.*}" => "")
8 myskewness(dist) = skewness(dist)
9 mykurtosis(dist) = kurtosis(dist)
10 function standardized_moment(dist::ContinuousUnivariateDistribution,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

```
In [8]: 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
```

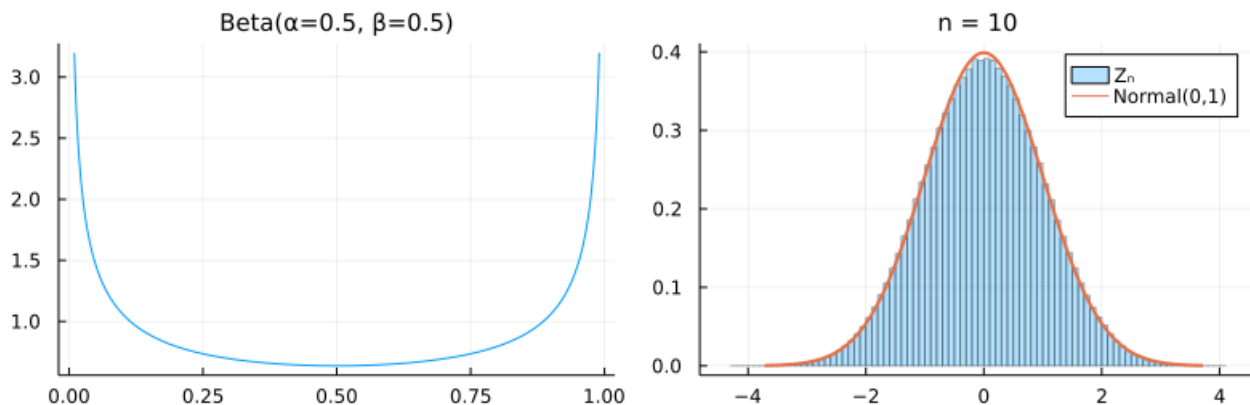
Out[8]:



```
In [9]: 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
```

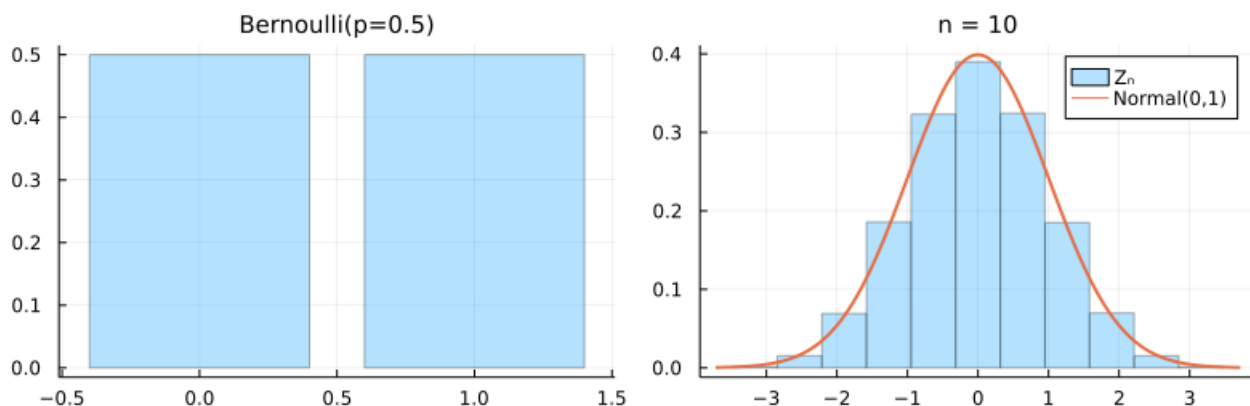
Out[9]:



```
In [10]: 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
```

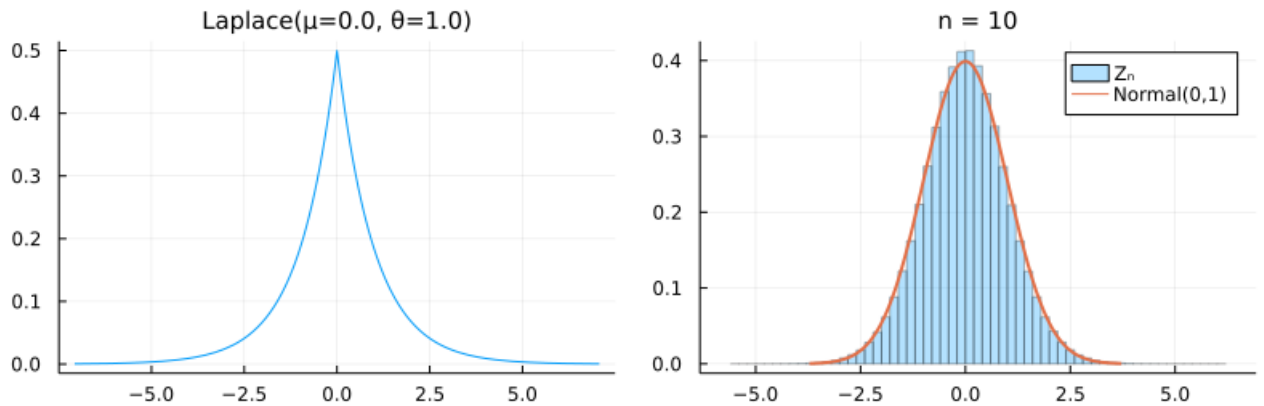
Out[10]:



```
In [11]: 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
```

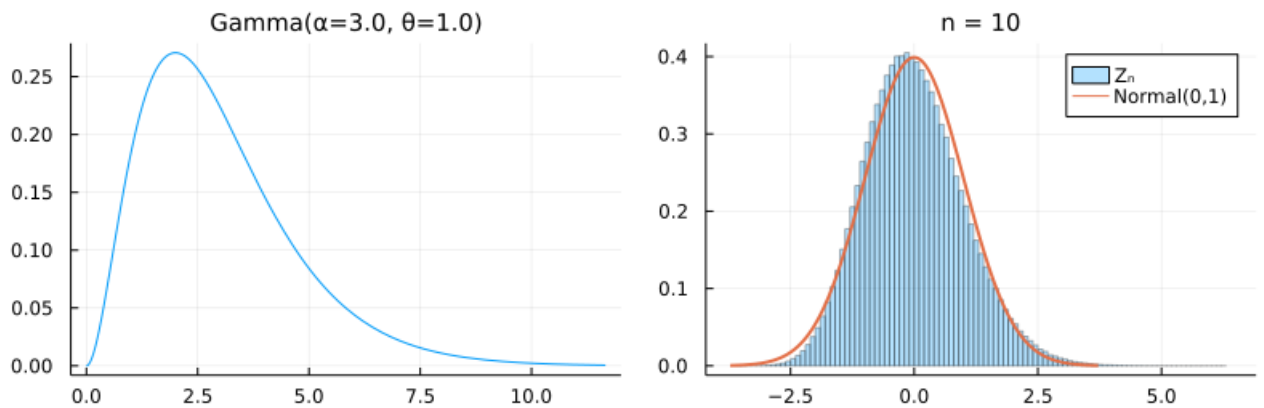
Out[11]:



```
In [12]: 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
```

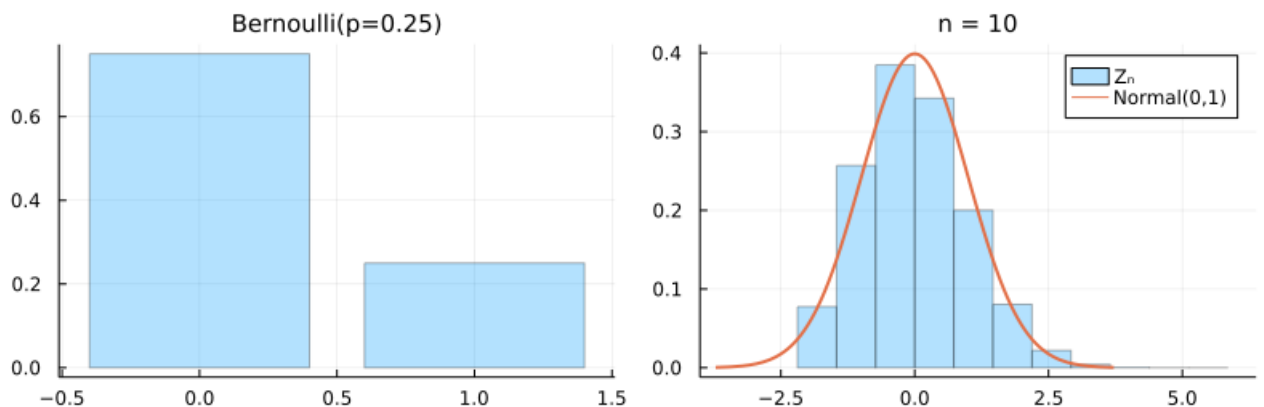
Out[12]:



```
In [13]: 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
```

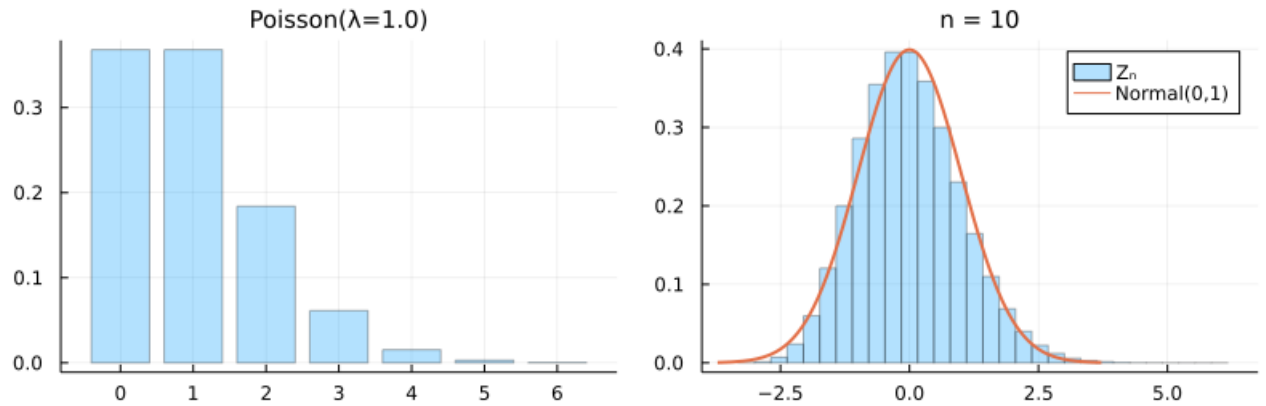
Out[13]:



```
In [14]: 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
```

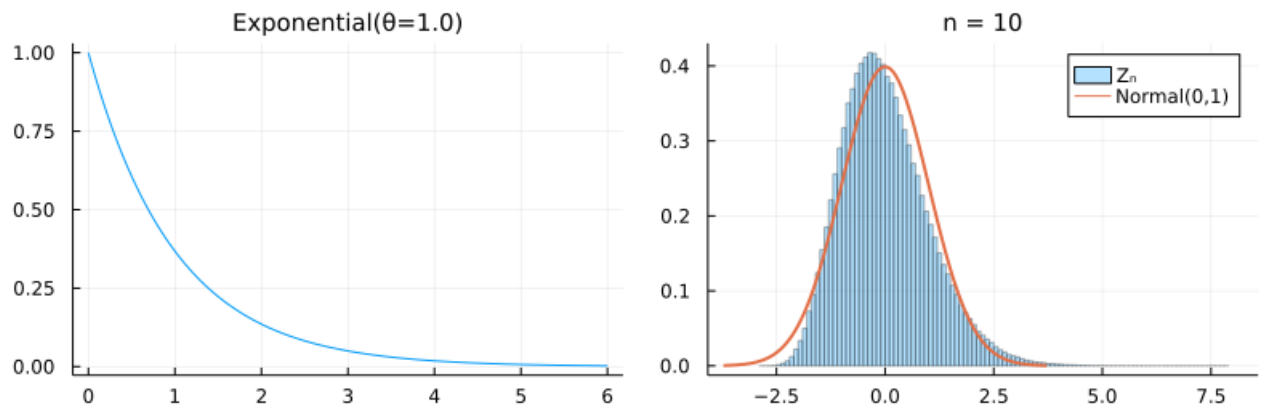
Out[14]:



```
In [15]: 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
```

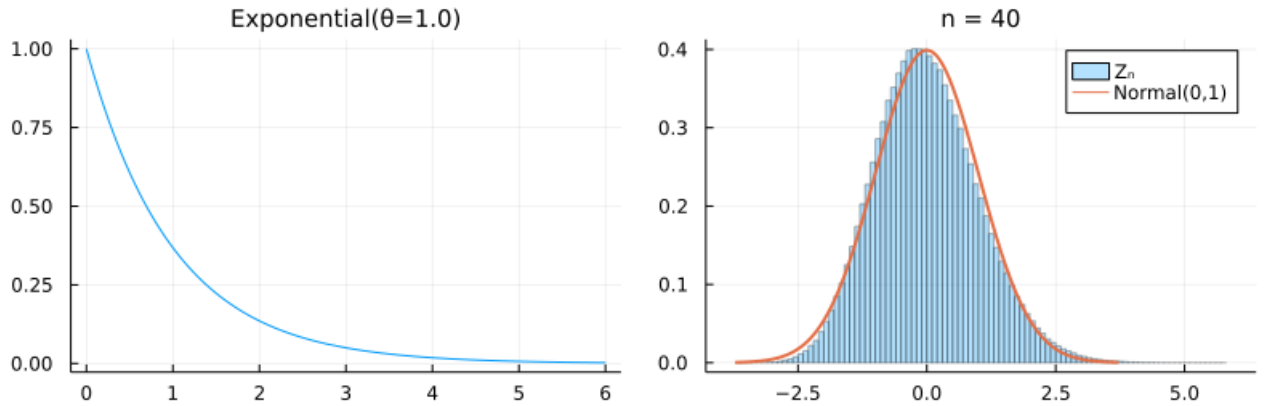
Out[15]:



```
In [16]: 1 plot_central_limit_theorem(Exponential(), 40)
```

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

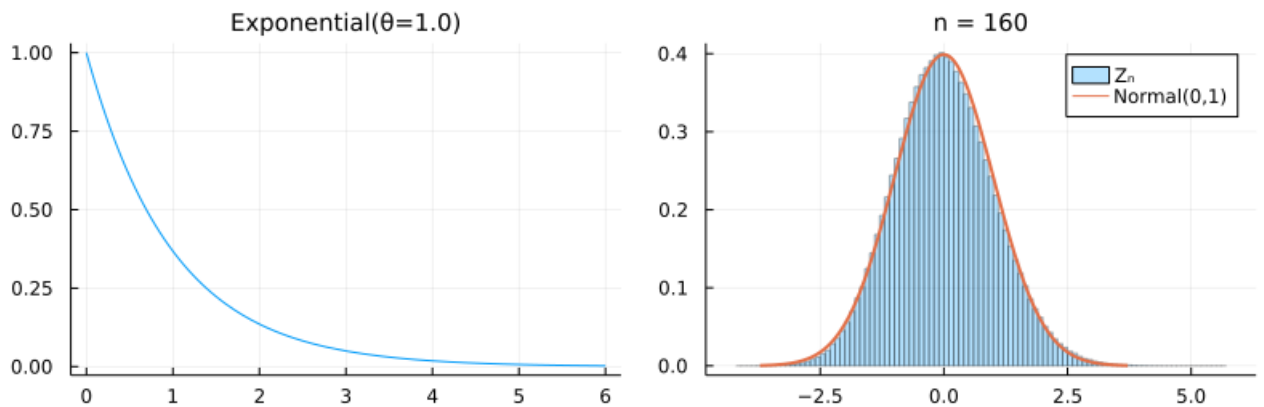
Out[16]:



```
In [17]: 1 plot_central_limit_theorem(Exponential(), 160)
```

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

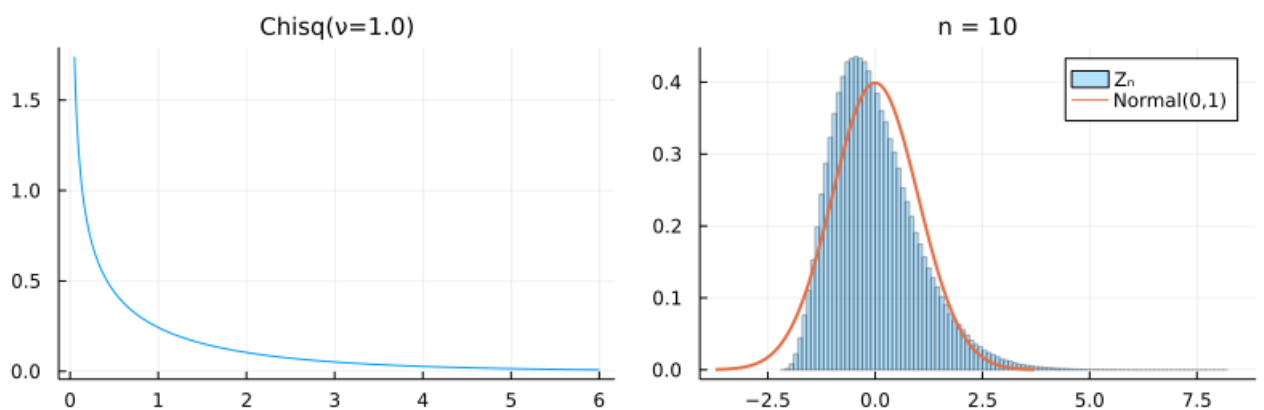
Out[17]:



```
In [18]: 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
```

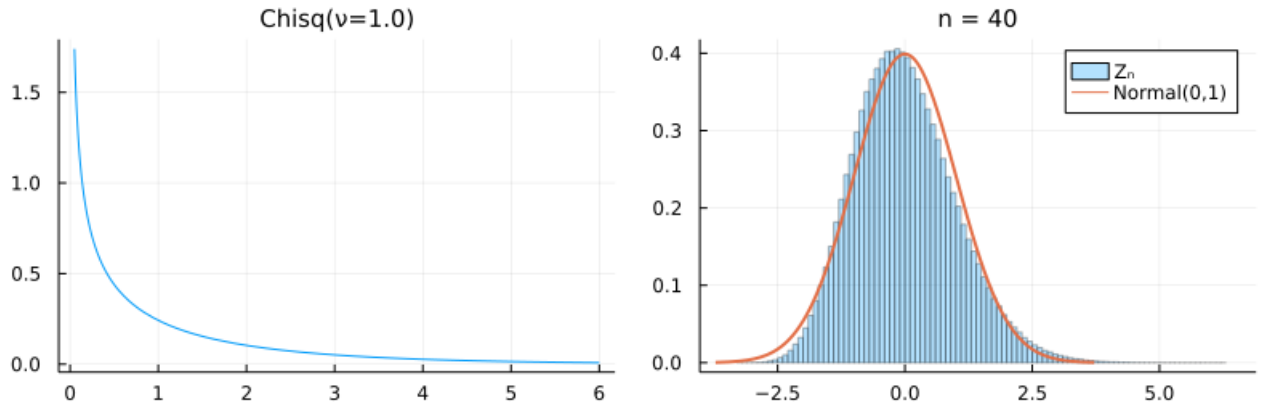
Out[18]:



```
In [19]: 1 plot_central_limit_theorem(Chisq(1), 40; a=0.05, b=6)
```

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

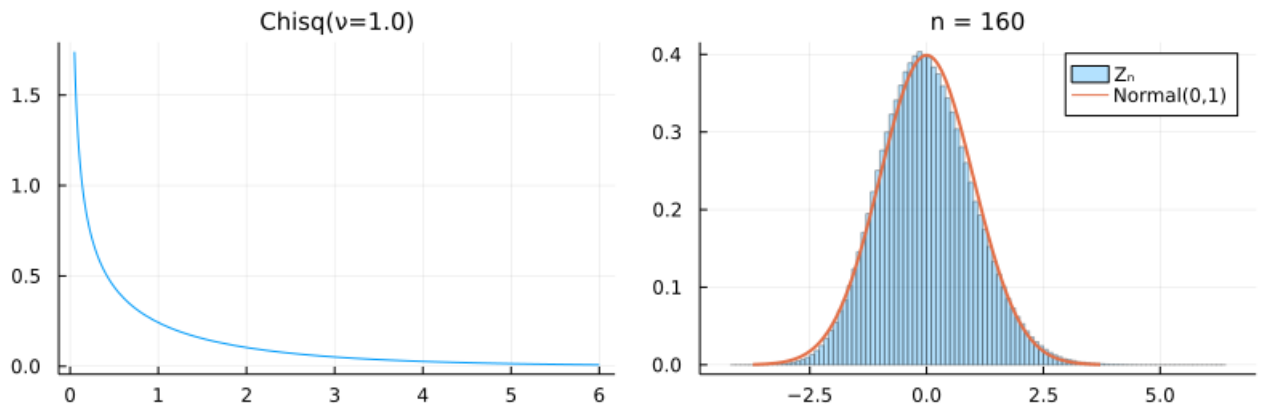
Out[19]:



```
In [20]: 1 plot_central_limit_theorem(Chisq(1), 160; a=0.05, b=6)
```

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

Out[20]:



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:

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

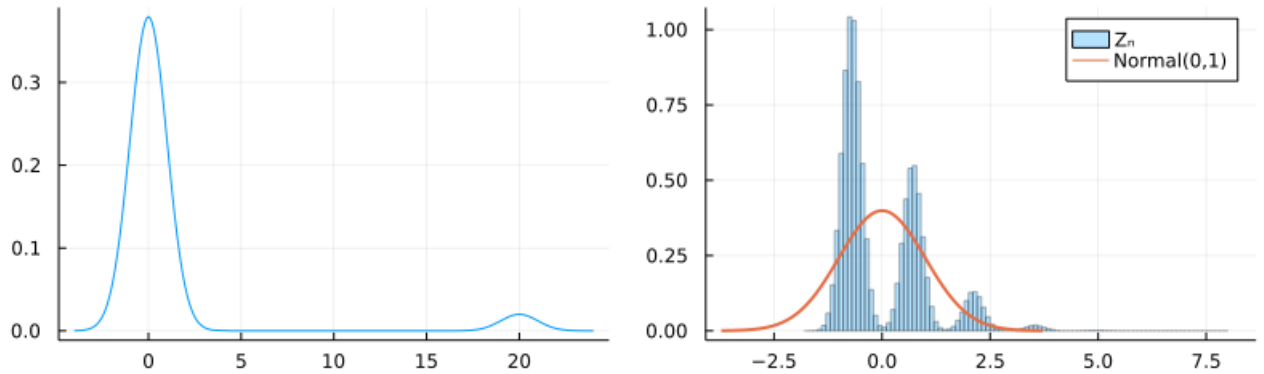
これは標準正規分布に割合が 5% の極端な外れ値を付け加えた分布になっている. 何らかの原因で分布に極端な外れ値が混ざっている場合には中心極限定理を使うときに注意を要する.

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 [21]: 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
```

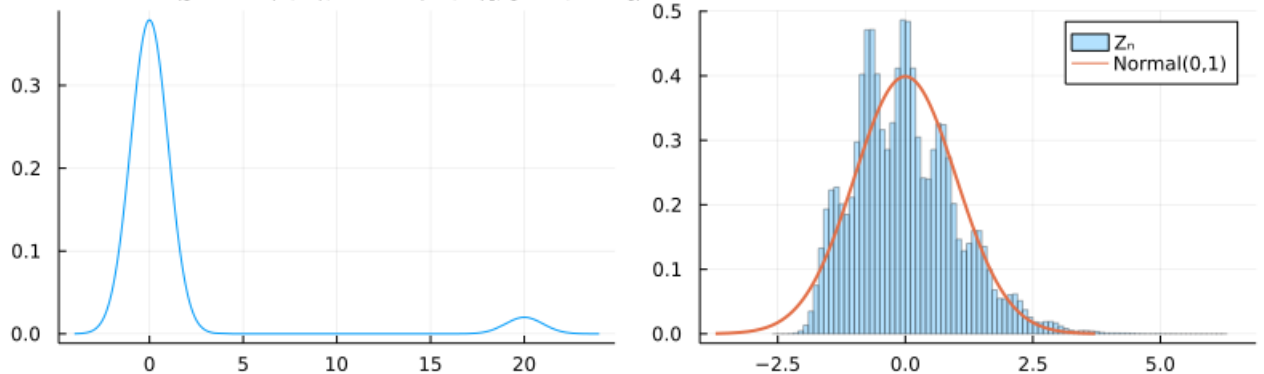
Out[21]: MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])



```
In [22]: 1 plot_central_limit_theorem(dist, 40; a, b, disttitle, titlefontsize)
```

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

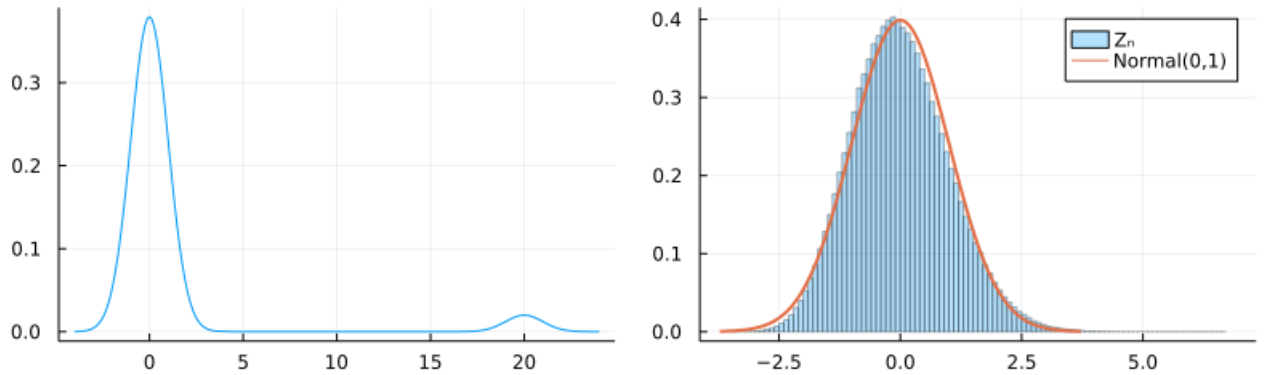
Out[22]: MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])



```
In [23]: 1 plot_central_limit_theorem(dist, 160; a, b, disttitle, titlefontsize)
```

```
skewness(dist) = 3.8236762415246486  
kurtosis(dist) = 13.5849999999990963
```

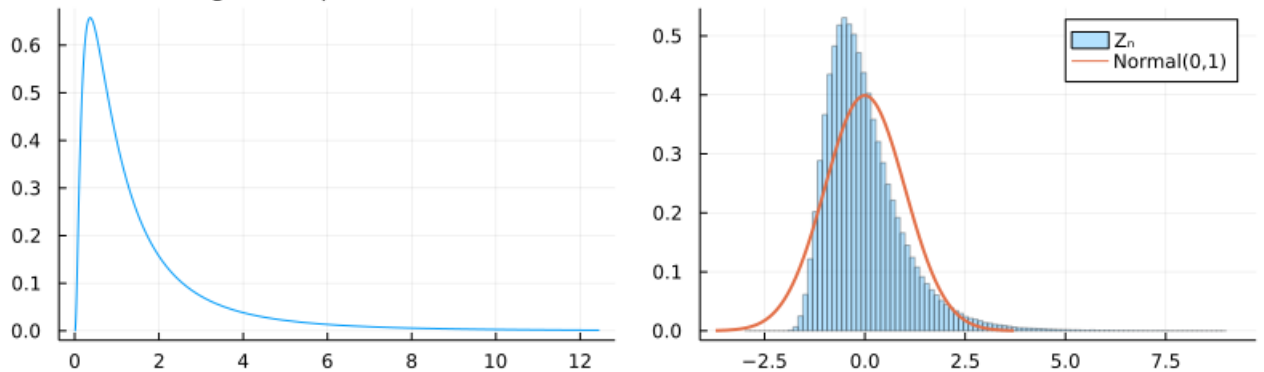
Out[23]: MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])



```
In [24]: 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
```

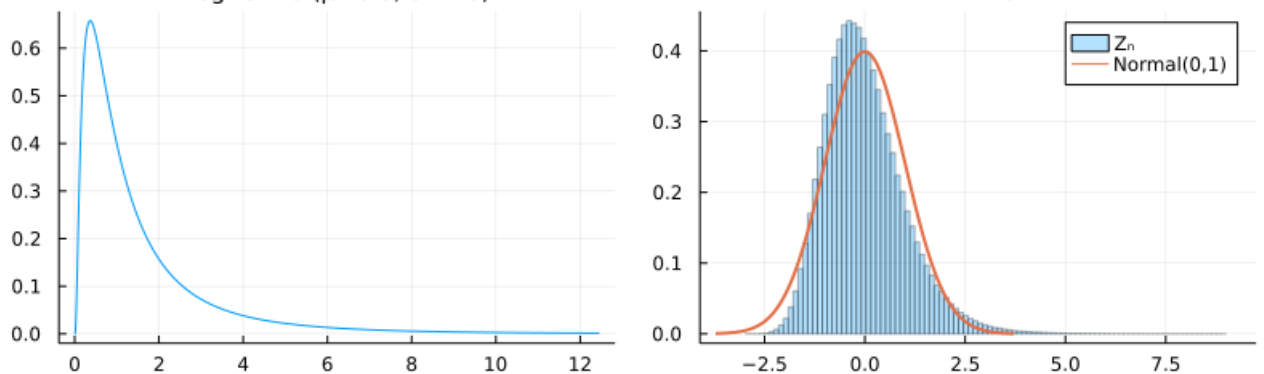
Out[24]: LogNormal($\mu=0.0$, $\sigma=1.0$)



```
In [25]: 1 plot_central_limit_theorem(LogNormal(), 40; bin=range(-3, 9, 100))
```

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

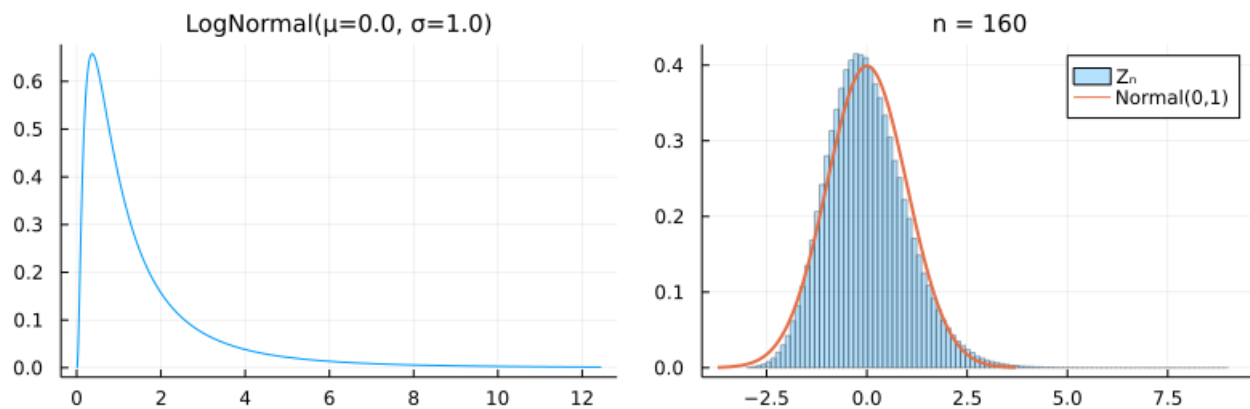
Out[25]: LogNormal($\mu=0.0$, $\sigma=1.0$)




```
In [26]: 1 plot_central_limit_theorem(LogNormal(), 160; bin=range(-3, 9, 100))
```

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

Out[26]:



3.6 標本平均と不偏分散の定義

3.7 問題: 標本平均の期待値と分散

3.8 問題: 不偏分散の期待値と分散

```
In [ ]: 1
```