

検定と信頼区間: 平均の検定と信頼区間

- 黒木玄
- 2022-05-31~2022-06-12

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 平均に関するP値と信頼区間](#)
 - [1.1 平均の検定で使われるP値の定義\(1\) 標準正規分布を使う場合](#)
 - [1.2 標本平均と不偏分散に関する中心極限定理と大数の法則の可視化](#)
 - [1.3 問題: 他の様々な分布について \$T\(\mu\)\$ の分布を確認](#)
 - [1.4 P値の定義\(1\)の標準正規分布を使う場合に対応する平均の信頼区間](#)
 - [1.5 平均の検定で使われるP値の定義\(2\) t分布を使う場合](#)
 - [1.6 標本分布におけるT統計量の分布の視覚化](#)
 - [1.7 分布の対数の標本分布のT統計量の視覚化](#)
 - [1.8 平均に関するP値の使い方](#)
 - [1.9 P値の定義\(2\)のt分布を使う場合に対応する平均の信頼区間](#)
- ▼ [2 平均に関するP値と信頼区間の計算例](#)
 - [2.1 平均に関するP値と信頼区間の計算の仕方](#)
 - ▼ [2.2 平均に関するP値と信頼区間の計算例](#)
 - [2.2.1 WolframAlphaでの平均に関するP値と信頼区間の計算例](#)
 - [2.2.2 Julia言語での平均に関するP値と信頼区間の計算例](#)
 - [2.2.3 R言語での平均に関するP値と信頼区間の計算例](#)
 - ▼ [2.3 必修練習問題: 平均に関するP値と信頼区間の計算](#)
 - [2.3.1 WolframAlphaによる必修練習問題の解き方](#)
 - [2.3.2 Julia言語による必修練習問題の解き方](#)
 - [2.3.3 R言語による必修練習問題の解き方](#)
 - [2.3.4 必修練習問題解答例](#)

```
In [1]: 1 ENV["LINES"], ENV["COLUMNS"] = 100, 100
2 using Base.Threads
3 using BenchmarkTools
4 using DataFrames
5 using Distributions
6 using LinearAlgebra
7 using Memoization
8 using Printf
9 using QuadGK
10 using RCall
11 using Random
12 Random.seed!(4649373)
13 using Roots
14 using SpecialFunctions
15 using StaticArrays
16 using StatsBase
17 using StatsFuns
18 using StatsPlots
19 default(fmt = :png, size = (400, 250),
20         titlefontsize = 10, plot_titlefontsize = 12)
21 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 " *
6         sympy.latex(x, mode="plain", fold_short_frac=false)))
7 end
8 @eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::AbstractArray{Sym})
9     function toeqnarray(x::Vector{Sym})
10         a = join(["\displaystyle " *
11             sympy.latex(x[i]) for i in 1:length(x)], "\\\")
12         """\left[ \begin{array}{r}$a\end{array} \right]"""
13     end
14     function toeqnarray(x::AbstractArray{Sym}, 2)
15         sz = size(x)
16         a = join([join("\displaystyle " .* map(sympy.latex, x[i,:]), "&")
17             for i in 1:sz[1]], "\\\")
18         """\left[ \begin{array}{r} " * repeat("r", sz[2]) * " " * a * "\end{array}\right]"
19     end
20     print(io, as_markdown(toeqnarray(x)))
21 end

```

```

In [3]: 1 safemul(x, y) = x == 0 ? x : x*y
2 safediv(x, y) = x == 0 ? x : x/y
3
4 x ≲ y = x < y || x ≈ y
5
6 mypdf(dist, x) = pdf(dist, x)
7 mypdf(dist::DiscreteUnivariateDistribution, x) = pdf(dist, round(Int, x))
8
9 distname(dist::Distribution) = replace(string(dist), r"{.*}" => "")
10 myskewness(dist) = skewness(dist)
11 mykurtosis(dist) = kurtosis(dist)
12 function standardized_moment(dist::ContinuousUnivariateDistribution, m)
13     μ, σ = mean(dist), std(dist)
14     quadgk(x → (x - μ)^m * pdf(dist, x), extrema(dist)...)[1] / σ^m
15 end
16 myskewness(dist::MixtureModel{Univariate, Continuous}) =
17     standardized_moment(dist, 3)
18 mykurtosis(dist::MixtureModel{Univariate, Continuous}) =
19     standardized_moment(dist, 4) - 3

```

Out[3]: mykurtosis (generic function with 2 methods)

In [4]:

```
1 pdflog(dist, x) = exp(logpdf(dist, exp(x)) + x)
2
3 @memoize function logmean(dist; a=-1e2, b=1e2)
4     quadgk(x → x * pdflog(dist, x), a, b)[1]
5 end
6
7 @memoize function logvar(dist; a=-1e2, b=1e2)
8     μ = logmean(dist; a, b)
9     quadgk(x → (x - μ)^2 * pdflog(dist, x), a, b)[1]
10 end
11 logstd(dist; a=-1e2, b=1e2) = √logvar(dist; a, b)
12
13 function plot_T(dist, n; L=10^6,
14     plot_chi=false, plot_t=false, logsample=false, kwargs...)
15     if logsample
16         μ, σ = logmean(dist), logstd(dist)
17     else
18         μ, σ = mean(dist), std(dist)
19     end
20     println("skewness = ", myskewness(dist), ", kurtosis = ", mykurtosis(dist))
21     ku = mykurtosis(dist)
22
23     se = σ/√n
24     Z = Vector{Float64}(undef, L)
25     S² = Vector{Float64}(undef, L)
26     T = Vector{Float64}(undef, L)
27     tmp = [Vector{eltype(dist)}(undef, n) for _ in 1:nthreads()]
28     @threads for i in 1:L
29         X = rand!(dist, tmp[threadid()])
30         if logsample
31             @. X = log(X)
32         end
33         X̄ = mean(X)
34         Z[i] = (X̄ - μ)/se
35         S²[i] = var(X)
36         T[i] = √n * (X̄ - μ)/√S²[i]
37     end
38
39     xlim = (-4.0, 5.9)
40     bin = range(-6, 6, 150)
41     P = stephist(Z; norm=true, xlim, bin, label="Z(μ)")
42     plot!(Normal(), xlim...; label="Normal(0,1)", ls=:dash)
43     plot!(); xtick=-10:10)
44
45     W = .√(S²/σ²)
46     xlim = (0, 3)
47     bin = range(quantile.(Ref(W), (0.0, 0.999))..., 150)
48     Q = stephist(W; norm=true, xlim, bin, label="√(S²/σ²)")
49     if plot_chi
50         plot!(Chi(n-1)/√(n-1), xlim...; label="Chi(n-1)/√(n-1)", ls=:dash)
51     else
52         vline!([1]; label="", ls=:dash)
53     end
54     plot!(); xtick=range(xlim...; step=0.5))
55
56     xlim = (-4.0, 5.9)
57     bin = range(-6, 6, 150)
58     R = stephist(T; norm=true, xlim, bin, label="T(μ)")
59     if plot_t
60         plot!(TDist(n-1), xlim...; label="TDist(n-1)", ls=:dash)
61     else
62         plot!(Normal(), xlim...; label="Normal(0,1)", ls=:dash)
63     end
64     plot!(); xtick=-10:10)
65
66     plot(P, Q, R; size=(800, 200), layout=(1, 3))
67     plot_title = if logsample
68         "Log of $(distname(dist)), n=$n"
69     else
70         "$ (distname(dist)), n=$n"
71     end
72     plot!(); plot_title, plot_titlefontsize=10)
73     plot!(); kwargs...)
74 end
75
76 #plot_T(Normal(2, 3), 10; plot_chi=true, plot_t=true)
77 #plot_T(Gamma(2, 3), 10; plot_chi=true, plot_t=true)
```

```
78 #plot_T(Gamma(2, 3), 10; plot_chi=true, plot_t=true, logsample=true)
```

Out[4]: plot_T (generic function with 1 method)

```
In [5]: 1 function pvalue_tdist( $\bar{x}$ ,  $s^2$ ,  $n$ ,  $\mu$ )
2      $t = (\bar{x} - \mu) / \sqrt{(s^2/n)}$ 
3      $2ccdf(TDist(n-1), abs(t))$ 
4 end
5 pvalue_tdist(X,  $\mu$ ) = pvalue_tdist(mean(X), var(X), length(X),  $\mu$ )
6
7 function confint_tdist( $\bar{x}$ ,  $s^2$ ,  $n$ ;  $\alpha = 0.05$ )
8      $c = quantile(TDist(n-1), 1-\alpha/2)$ 
9     [ $\bar{x} - c*\sqrt{(s^2/n)}$ ,  $\bar{x} + c*\sqrt{(s^2/n)}$ ]
10 end
11 confint_tdist(X;  $\alpha$ ) = confint_tdist(mean(X), var(X), length(X);  $\alpha$ )
12
13 function show_confint_of_mean(X;  $\mu = mean(x)$ ,  $\alpha = 0.05$ )
14     println("P-value = ", pvalue_tdist(X,  $\mu$ ))
15     println("confidence interval of mean = ", confint_tdist(X;  $\alpha$ ))
16 end
17
18 function plot_confint_of_mean(x;  $\alpha = 0.05$ , xtick=-100:100, kwargs...)
19      $n = length(x)$ 
20      $a, b = extrema(x)$ 
21      $a, b = a - 0.05(b-a)$ ,  $b + 0.05(b-a)$ 
22     confidence_interval = confint_tdist(x;  $\alpha$ )
23     scatter(x, zeros(n); label="", msc=:auto, alpha=0.7)
24     plot!(confidence_interval, fill(1, 2); label="", lw=8, c=:red)
25     plot!(); xlim=(a, b), xtick)
26     plot!(); ylim=(-0.5, 1.5), yaxi=false, ytick=false)
27     title!(" $\$(100(1-\alpha))\%$  confidence interval of mean for data of size  $n=\$n$ ")
28     plot!(); size=(600, 60))
29     plot!(); kwargs...)
30 end
31
32 #plot_confint_of_mean(rand(Normal(2, 3), 101);  $\alpha = 0.05$ )
```

Out[5]: plot_confint_of_mean (generic function with 1 method)

1 平均に関するP値と信頼区間

以下のようなことを行いたい.

(1) S市の中学3年生男子達から n 人を無作為抽出して身長を測って得た数値のデータ x_1, \dots, x_n から, S市の中学3年生男子達の身長の平均値を推定したい.

(2) とある店に出されるフライドポテトの長さを n 本分測って得た数値のデータ x_1, \dots, x_n から, その店に出されるフライドポテトの長さの平均値を推定したい.

このような推定を以下では **平均の推定** と呼ぶことにする.

目標は平均の信頼区間の構成である.

そのためには, 検定と信頼区間の表裏一体性より, P値を適切に定義すればよい.

1.1 平均の検定で使用するP値の定義(1) 標準正規分布を使う場合

データ: n 個の実数値 x_1, \dots, x_n .

データの標本平均と不偏分散をそれぞれ \bar{x} , s^2 と書く:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}).$$

統計モデル: 平均 μ と分散 σ^2 を持つ未知の確率分布 D のサイズ n の標本分布 D^n を統計モデルとして採用する.

D が連続分布でその確率密度関数が $p(x)$ と書かれるとき, 標本分布 D^n の確率密度関数は次のように表される:

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

D が離散分布でその確率質量関数が $P(x)$ と書かれるとき, 標本分布 D^n の確率質量関数は次のように表される:

$$P(x_1, \dots, x_n) = P(x_1) \cdots P(x_n).$$

この設定は、データ x_1, \dots, x_n がある母集団からの無作為抽出またはルーレットやサイコロなどの同分布の独立試行で得られた数値の列である場合に適切になる。

そして、未知だと想定されている分布 D は母集団の分布またはルーレットやサイコロの確率分布のモデル化になっていると考えている。

検定したい仮説: $\mu = \mu_0$ (μ_0 は具体的な数値)。

X_1, \dots, X_n は統計モデル D^n に従う確率変数達であるとする。すなわち、 X_1, \dots, X_n は独立な確率変数達であり、各々がモデルの確率分布 D に従っていると仮定する。

標本平均に関する中心極限定理: 中心極限定理によって、 n が十分に大きいならば、モデル内でランダムに生成された仮想的標本 X_1, \dots, X_n の標本平均

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

は平均 μ 、分散 σ^2/n の正規分布に近似的に従う:

$$\bar{X} \sim \text{Normal}\left(\mu, \sqrt{\sigma^2/n}\right), \text{ approximately.}$$

不偏分散に関する大数の法則: 大数の法則より、 n を十分に大きくすると、モデルの確率分布 D の分散 σ^2 の値は、モデル内でランダムに生成された仮想的標本 X_1, \dots, X_n の不偏分散(不偏推定量になるように補正された標本分散)

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

で近似される:

$$S^2 \approx \sigma^2.$$

検定で使われる T 統計量: 以上の状況の下で、 T 統計量を

$$T(\mu) := \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \approx \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim \text{Normal}(0, 1), \text{ approximately.}$$

と定める。このとき $T(\mu)$ は近似的に標準正規分布に従う:

$$T(\mu) \sim \text{Normal}(0, 1), \text{ approximately.}$$

データの t 値を次のように定める:

$$t(\mu) := \frac{\bar{x} - \mu}{\sqrt{s^2/n}}.$$

P値の定義: 仮説 $\mu = \mu_0$ 下の統計モデル内で $|T(\mu_0)| \geq |t(\mu_0)|$ となる確率を、標準正規分布に従ってランダムに生成される値の絶対値が $|t(\mu_0)|$ 以上になる確率として近似的に求めて、その値をデータ x_1, \dots, x_n に関する仮説 $\mu = \mu_0$ のP値として採用する。そのP値を次のように書く:

$$\text{pvalue}_{\text{Normal}}(\bar{x}, s^2 | n, \mu = \mu_0) = 2(1 - \text{cdf}(\text{Normal}(0, 1), |t(\mu_0)|)).$$

1.2 標本平均と不偏分散に関する中心極限定理と大数の法則の可視化

前節のP値の定義では、中心極限定理によって、平均 μ 、分散 σ^2 を持つ確率分布 D のサイズ n の標本 X_1, \dots, X_n (独立同分布な確率変数達になる)の標本平均 \bar{X} (確率変数になる)が、 n が大きくなると、平均 μ 、分散 σ^2/n を持つ正規分布に **近似的に** 従うという結果(中心極限定理)を使った。

そしてさらに、大数の法則によって、 n を大きくすると、 X_1, \dots, X_n の不偏分散が σ を近似するようになることも使った。

前節のP値に定義が意味を持つためには、 $T(\mu) = (\bar{X} - \mu) / \sqrt{S^2/n}$ が近似的に標準正規分布に従うことが必要である。

以下でそれを様々な確率分布 D について確認してみよう。

以下では

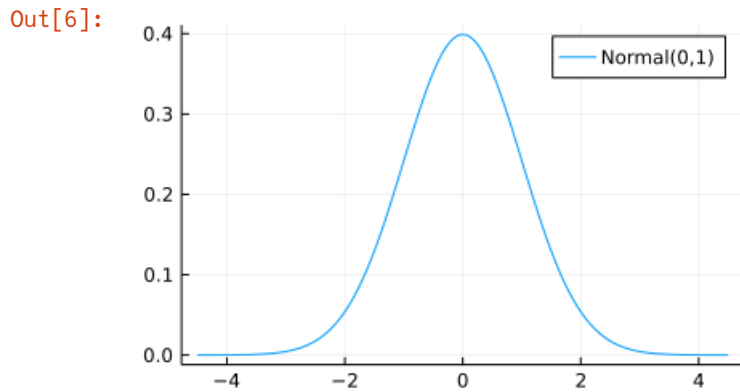
$$Z(\mu) = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}, \quad \sqrt{S^2/\sigma^2}, \quad T(\mu) = \frac{\bar{X} - \mu}{\sqrt{S^2/n}}$$

の3つの確率変数の分布をプロットする. n を大きくして行くと, 大数の法則より, $\sqrt{S^2/\sigma^2}$ の分布は1に集中して行き, 中心極限定理より, $Z(\mu)$ の分布は標準正規分布に近づく.

特に, n を大きくして行ったときの $T(\mu)$ の分布の標準正規分布による近似の誤差が小さくなる速さが, 確率分布 D の種類ごとに違うことに注目せよ.

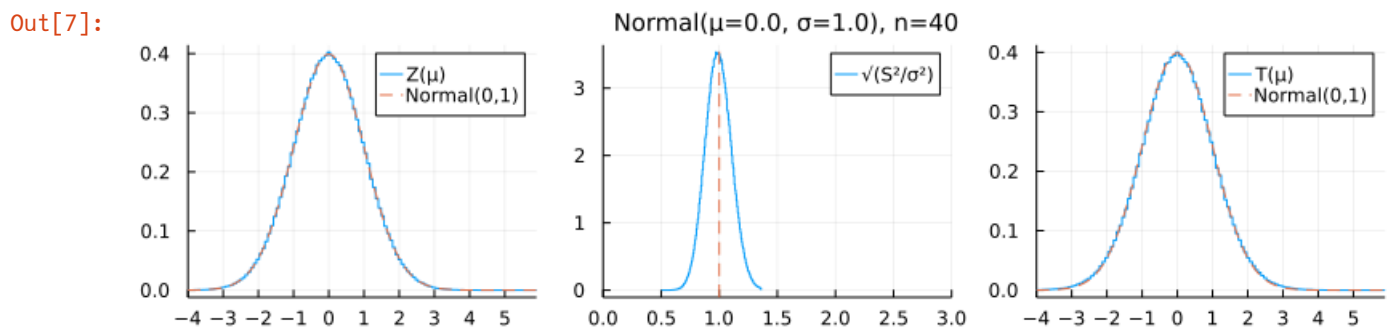
正規分布の場合

```
In [6]: 1 plot(Normal(), -4.5, 4.5; label="Normal(0,1)")
```



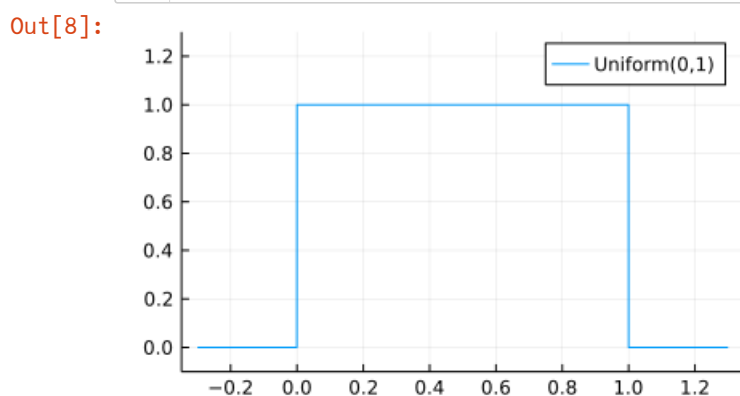
```
In [7]: 1 plot_T(Normal(), 40)
```

skewness = 0.0, kurtosis = 0.0



一様分布の場合

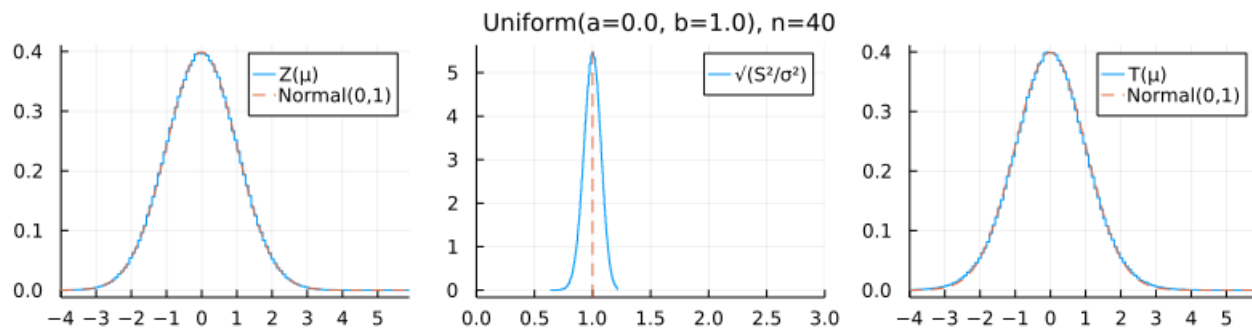
```
In [8]: 1 plot(Uniform(), -0.3, 1.3; label="Uniform(0,1)", ylim=(-0.1, 1.3))
```



```
In [9]: 1 plot_T(Uniform(), 40)
```

skewness = 0.0, kurtosis = -1.2

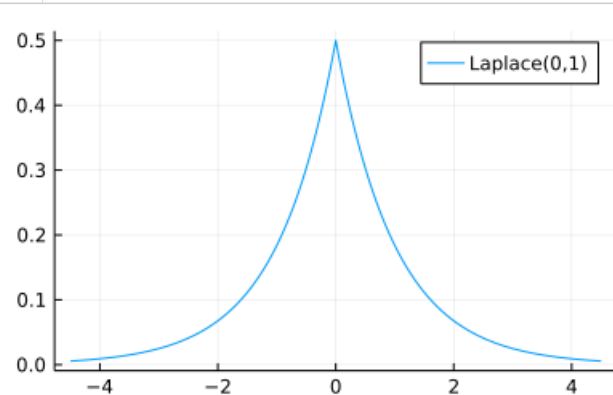
Out[9]:



Laplace分布の場合

```
In [10]: 1 plot(Laplace(), -4.5, 4.5; label="Laplace(0,1)")
```

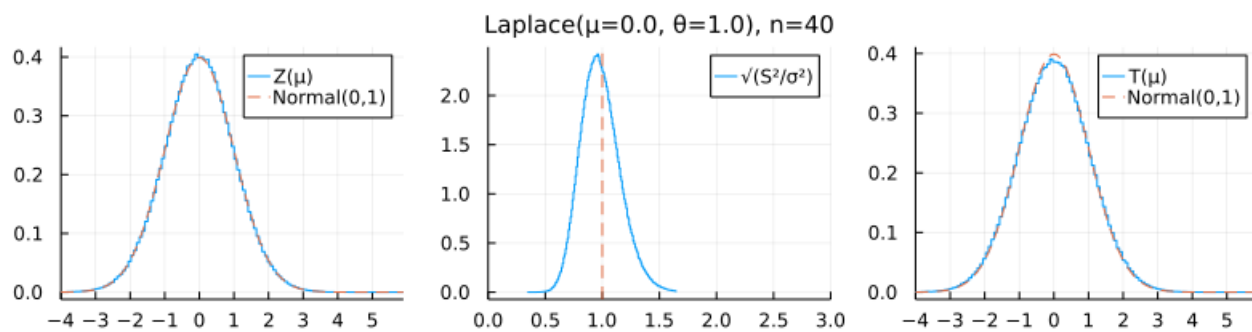
Out[10]:



```
In [11]: 1 plot_T(Laplace(), 40)
```

skewness = 0.0, kurtosis = 3.0

Out[11]:

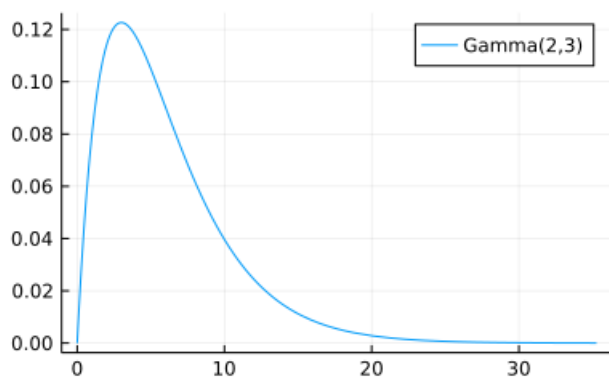


ガンマ分布の場合

ガンマ分布は左右非対称な分布の典型例である。

```
In [12]: 1 plot(Gamma(2,3); label="Gamma(2,3)")
```

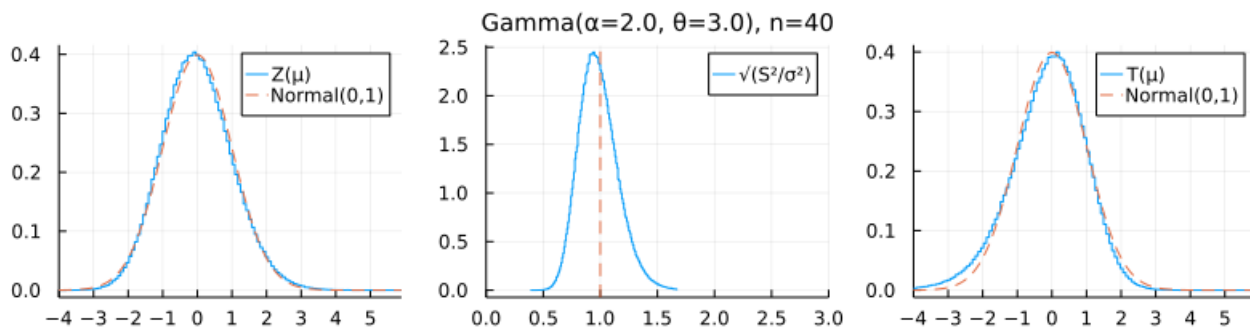
Out[12]:



```
In [13]: 1 plot_T(Gamma(2,3), 40)
```

skewness = 1.414213562373095, kurtosis = 3.0

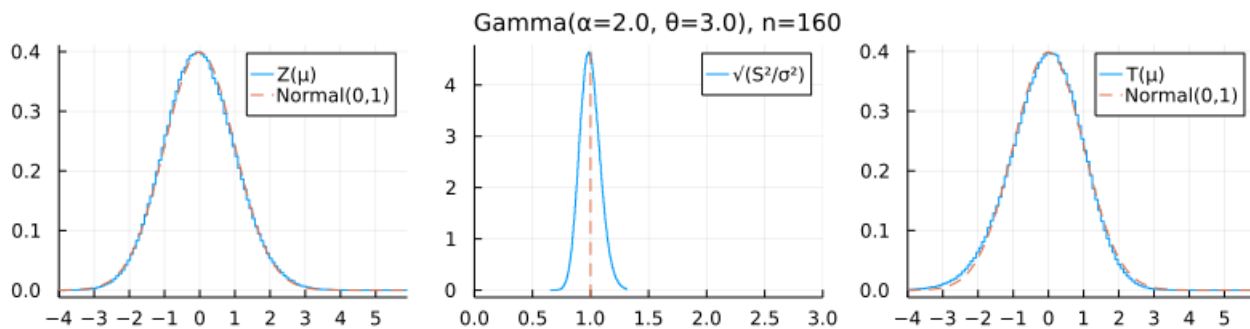
Out[13]:



```
In [14]: 1 plot_T(Gamma(2,3), 160)
```

skewness = 1.414213562373095, kurtosis = 3.0

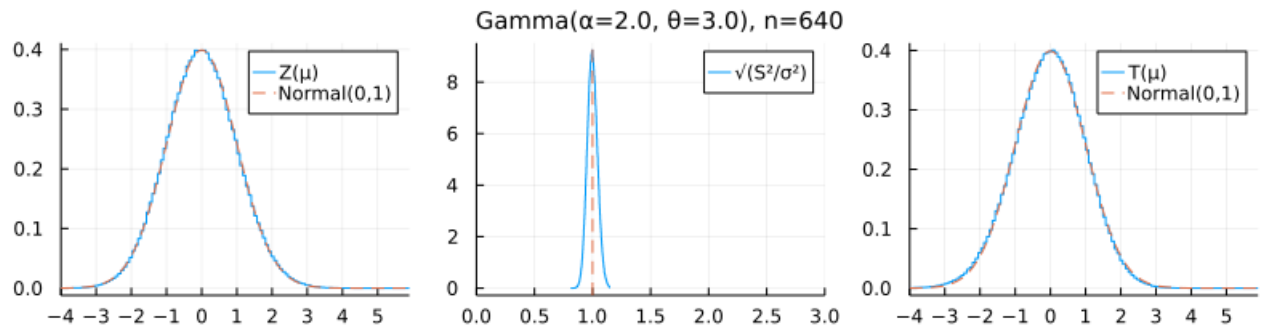
Out[14]:




```
In [15]: 1 plot_T(Gamma(2,3), 640)
```

skewness = 1.414213562373095, kurtosis = 3.0

Out[15]:



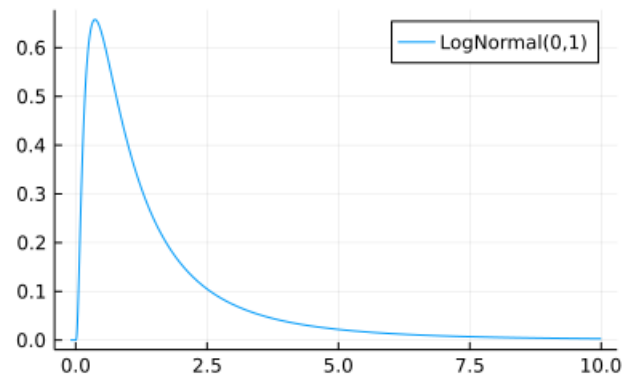
対数正規分布の場合

対数正規分布は左右非対称でかつ **外れ値が出易い** 分布の典型例である.

このような場合は n を非常に大きくしないと近似の精度が高まらない.

```
In [16]: 1 plot(LogNormal(), -0.1, 10; label="LogNormal(0,1)")
```

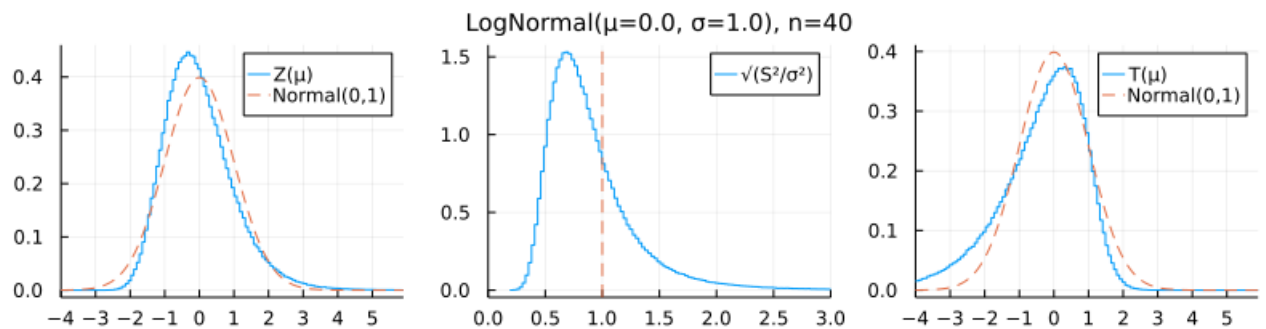
Out[16]:



```
In [17]: 1 plot_T(LogNormal(), 40)
```

skewness = 6.184877138632554, kurtosis = 110.9363921763115

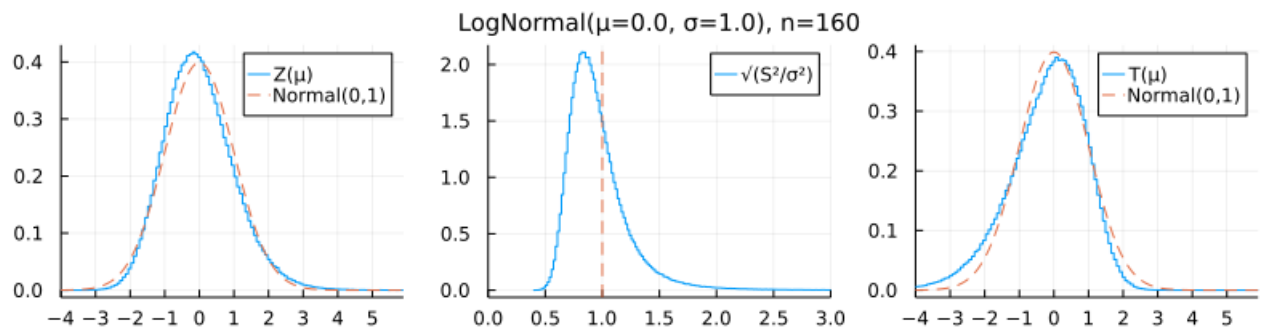
Out[17]:



```
In [18]: 1 plot_T(LogNormal(), 160)
```

skewness = 6.184877138632554, kurtosis = 110.9363921763115

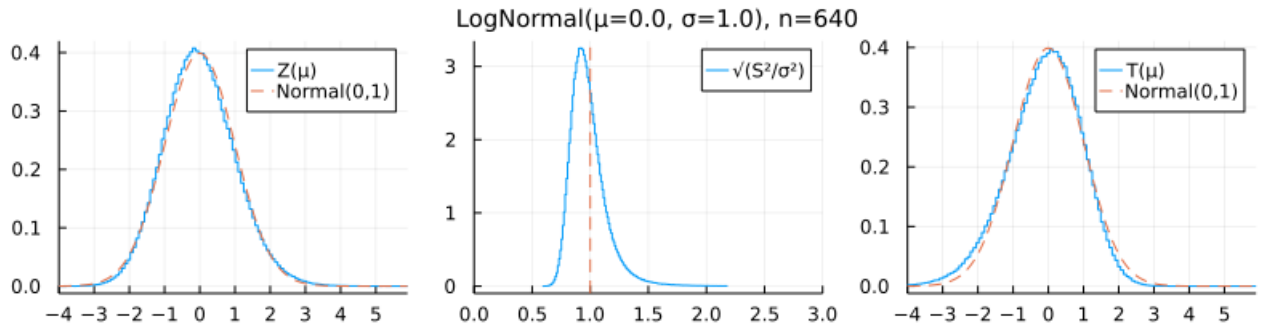
Out[18]:



```
In [19]: 1 plot_T(LogNormal(), 640)
```

skewness = 6.184877138632554, kurtosis = 110.9363921763115

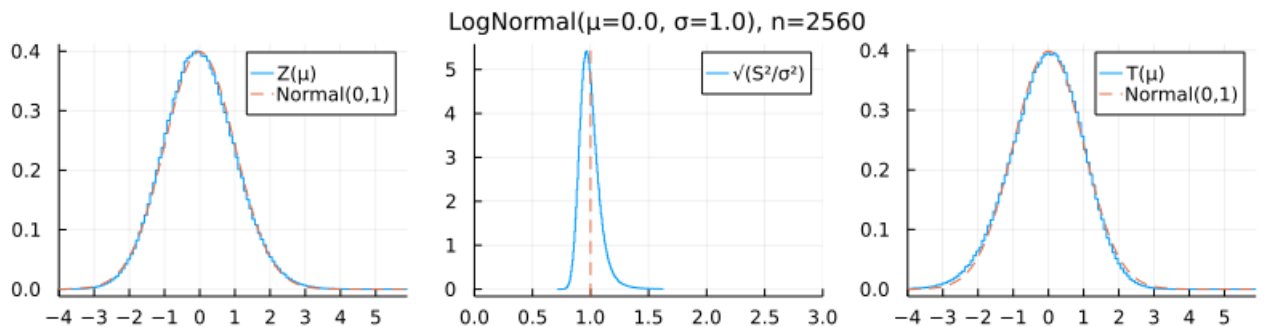
Out[19]:



```
In [20]: 1 plot_T(LogNormal(), 2560)
```

skewness = 6.184877138632554, kurtosis = 110.9363921763115

Out[20]:



この例から、左右非対称で外れ値を含む母集団分布からの無作為抽出で得たデータを扱う場合には、中心極限定理を使った近似を用いることは要注意であることがわかる。

注意: $X \sim D$ と仮定する。一般に、分布 D の歪度(わいど, skewness)と尖度(せんど, kurtosis, 過剰尖度, excessive kurtosis)をそれぞれ

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

と書く。(D が正規分布の場合には $\bar{\kappa}_3 = \bar{\kappa}_4 = 0$ となる。)

このとき、 $\bar{\kappa}_4 \geq \bar{\kappa}_3^2 - 2$ および以下が成立することを示せる:

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

そして、多変量分布の中心極限定理(これはこのノート群では証明していないが、1変量の場合と同じ方法で証明される)によって、 n が大きくなると、 \bar{X}, S^2 の同時分布は、それらを平均と分散共分散行列に持つ2変量正規分布で近似される。 $n \rightarrow \infty$ で分散共分散行列が 0 に収束することから、 (\bar{X}, S^2) の分布が (μ, σ^2) に集中して行くこともわかる(大数の法則)。

この一般的な結果の視覚化については「[標本分布について](#)」のノート

(<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/04%20Distribution%20of%20samples.ipynb>)を参照せよ。そこで詳しく解説されている結果と視覚化は「平均の検定や信頼区間」のような統計学における基本を理解するためにかなり重要なことなのだが、入門的な教科書ではほとんど説明されていない。理解を深めるために貴重な解説になっていると思われるので是非とも繰り返し参照して欲しい。

1.3 問題: 他の様々な分布について $T(\mu)$ の分布を確認

上の例以外の分布についても $T(\mu)$ の分布がどうなっているかを確認せよ。

ヒント: 「[標本分布について](#)」のノート

(<https://nbviewer.org/github/genkuroki/Statistics/blob/master/2022/04%20Distribution%20of%20samples.ipynb>)を参照せよ。

次節で使う $z_{\alpha/2}$ の値の例: $\alpha = 5\%, 1\%, 0.1\%$ のそれぞれについて,

```
In [21]: 1 @. round(quantile(Normal(), 1 - (0.05, 0.01, 0.001) / 2); digits=4)
```

```
Out[21]: (1.96, 2.5758, 3.2905)
```

1.4 P値の定義(1)の標準正規分布を使う場合に対応する平均の信頼区間

有意水準を $0 \leq \alpha \leq 1$ と書き, 標準正規分布において $z_{\alpha/2}$ 以上になる確率は $\alpha/2$ になると仮定する:

$$z_{\alpha/2} = \text{quantile}(\text{Normal}(0, 1), 1 - \alpha/2).$$

例えば,

$$z_{5\%/2} \approx 1.9600, \quad z_{1\%/2} \approx 2.5758, \quad z_{0.1\%/2} \approx 3.2905.$$

P値関数 $\text{pvalue}_{\text{Normal}}(\bar{x}, s^2 | n, \mu = \mu_0) = 2(1 - \text{cdf}(\text{Normal}(0, 1), |t(\mu_0)|))$ に対応する信頼度 $1 - \alpha$ の信頼区間は次のようになる:

$$\text{confint}_{\text{Normal}}(\bar{x}, s^2 | n, \mu = \mu_0) = \left[\bar{x} - z_{\alpha/2} \sqrt{s^2/n}, \bar{x} + z_{\alpha/2} \sqrt{s^2/n} \right].$$

証明: 検定と信頼区間の表裏一体性より, P値関数 $\text{pvalue}_{\text{Normal}}(\bar{x}, s^2 | n, \mu = \mu_0)$ に対応する信頼度 $1 - \alpha$ の信頼区間は次のように定義されるのであった:

$$\text{confint}_{\text{Normal}}(\bar{x}, s^2 | n, \mu = \mu_0) = \{ \mu_0 \in \mathbb{R} \mid \text{pvalue}_{\text{Normal}}(\bar{x}, s^2 | n, \mu = \mu_0) \geq \alpha \}.$$

そして,

$$\begin{aligned} \text{pvalue}_{\text{Normal}}(\bar{x}, s^2 | n, \mu = \mu_0) &\geq \alpha \\ \iff 1 - \text{cdf}(\text{Normal}(0, 1), |t(\mu_0)|) &\geq \alpha/2 \\ \iff |t(\mu_0)| = \frac{|\bar{x} - \mu_0|}{\sqrt{s^2/n}} &\leq z_{\alpha/2} \\ \iff \bar{x} - z_{\alpha/2} \sqrt{s^2/n} \leq \mu_0 &\leq \bar{x} + z_{\alpha/2} \sqrt{s^2/n}. \end{aligned}$$

これより, P値の定義(1)に対応する平均の信頼区間が上のようになることがわかった.

証明終

1.5 平均の検定で使用されるP値の定義(2) t 分布を使う場合

分布 D が左右対称の分布ならば $n = 10$ のようなかなり小さな n で中心極限定理による近似の誤差は非常に小さくなる場合がある. しかし, そういう場合であっても, 大数の法則を使った不偏分散 S^2 による σ^2 の近似の精度は低いままの場合がある.

そういう場合の補正を t 分布を使って行う処方箋を採用しよう.

統計モデルとして正規分布の標本分布を仮定: 以下では, 統計モデルとして, 平均 μ , 分散 σ^2 を持つ正規分布のサイズ n の標本分布を採用する. その確率密度関数は次のように表される:

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

X_1, \dots, X_n はこの統計モデルに従う確率変数達であるとし, それらの標本平均と不偏分散をそれぞれ \bar{X} , S^2 と表す.

\bar{X} と S^2 の同時分布: 標本平均と不偏分散

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

を使うと,

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n ((x_i - \bar{x})(\bar{x} - \mu))^2 = (n-1)s^2 + n(\bar{x} - \mu)^2.$$

なので,

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

このことから、 \bar{X}, S^2 は独立な確率変数であり、 \bar{X} は平均 μ 、分散 σ^2/n の正規分布に従い、 $(n-1)S^2/\sigma^2$ は自由度 $n-1$ の χ^2 分布に従うことを示せる:

$$\bar{X} \sim \text{Normal}\left(\mu, \sqrt{\sigma^2/n}\right), \quad \frac{(n-1)S^2}{\sigma^2} \sim \text{Chisq}(n-1).$$

詳しくは「標本分布について」のノートの「正規分布の標本分布の場合」の節を参照せよ. 自由度 $n-1$ の χ^2 分布の密度関数を作るために必要な、 $u = s^2$ とおいたときの因子 $u^{(n-1)/2-1}$ は

$$dy_1 \cdots dy_{n-1} \propto (\sqrt{u})^{n-2} d\sqrt{u} d\theta \propto u^{(n-1)/2-1} du d\theta$$

のような計算で出て来る. ここで $d\theta$ は $n-2$ 次元単位球面の微小面積要素である.

T統計量が従う分布: ゆえに、 T 統計量

$$T(\mu) = \frac{\bar{X} - \mu}{\sqrt{S^2/n}}$$

は自由度 $n-1$ の t 分布に従う:

$$T(\mu) \sim \text{TDist}(n-1).$$

詳しくは、「正規分布の標本分布から自然に t 分布に従う確率変数が得られること」を参照せよ.

注意: 自由度が大きな t 分布は標準正規分布とほぼ同じになるので、この結果はP値の定義(1)で使った結果と整合的であり、この結果が意味を持つのは n が大きくない場合にのみ意味を持つ. 実践的には n が十分に大きな場合には $\text{TDist}(n-1)$ を標準正規分布で置き換えてよい.

P値の定義: 仮説 $\mu = \mu_0$ の下で $|T(\mu_0)| \geq |t(\mu_0)|$ となる確率を、自由度 $n-1$ の t 分布に従ってランダムに生成される値の絶対値が $|t(\mu_0)|$ 以上になる確率として正確に求めて、その値をデータ x_1, \dots, x_n に関する仮説 $\mu = \mu_0$ のP値として採用する. そのP値を次のように書く:

$$\text{pvalue}_{\text{TDist}}(\bar{x}, s^2 | n, \mu = \mu_0) = 2(1 - \text{cdf}(\text{TDist}(n-1), |t(\mu_0)|)).$$

このP値は n が大きな場合には定義(1)のP値 $\text{pvalue}_{\text{Normal}}(\bar{x}, s^2 | n, \mu = \mu_0)$ と近似的に一致する. ゆえに、P値の定義(1)を使うことを止めて、こちらの定義(2)の方だけを使うことにしても害がないと考えられる.

我々はこれ以後こちらのP値の定義のみを使用する.

注意: ただし、こちらの定義(2)のP値は先の定義(1)のP値よりも少し大きくなり、定義(2)の場合の信頼区間は定義(1)の場合の信頼区間よりも少し広くなる.

```
In [22]: 1 df = 10:10:100
2 α = [0.05, 0.01, 0.001]
3 z = @. round(quantile(Normal(), 1-α'/2); digits=4)
4 t = @. round(quantile(TDist(df), 1-α'/2); digits=4)
5 table_of_t = DataFrame(
6     df = Any[df; Inf],
7     var"α = 5%" = [t[:,1]; z[1]],
8     var"α = 1%" = [t[:,2]; z[2]],
9     var"α = 0.1%" = [t[:,3]; z[3]],
10 )
```

Out[22]: 11 rows × 4 columns

	df	α = 5%	α = 1%	α = 0.1%
	Any	Float64	Float64	Float64
1	10	2.2281	3.1693	4.5869
2	20	2.086	2.8453	3.8495
3	30	2.0423	2.75	3.646
4	40	2.0211	2.7045	3.551
5	50	2.0086	2.6778	3.496
6	60	2.0003	2.6603	3.4602
7	70	1.9944	2.6479	3.435
8	80	1.9901	2.6387	3.4163
9	90	1.9867	2.6316	3.4019
10	100	1.984	2.6259	3.3905
11	Inf	1.96	2.5758	3.2905

```
In [23]: 1 quantile(TDist(30), 1-0.05/2) / quantile(Normal(), 1-0.05/2)
```

Out[23]: 1.0419948899114568

1.6 標本分布におけるT統計量の分布の視覚化

前節では、 t 分布を使用することによって、分布 D が左右対称でおとなしめもしくは正規分布に非常に近い分布でかつ n が小さな場合に精度が高まることを期待してP値を再定義した。

その期待が実際に正しそうなことを確認するために、以下では再度 $T(\mu)$ の分布の視覚化を行う。

そのとき、非対称な分布の場合に n を大きくしたとき、 t 分布の使用に害がないことも確認しておこう。

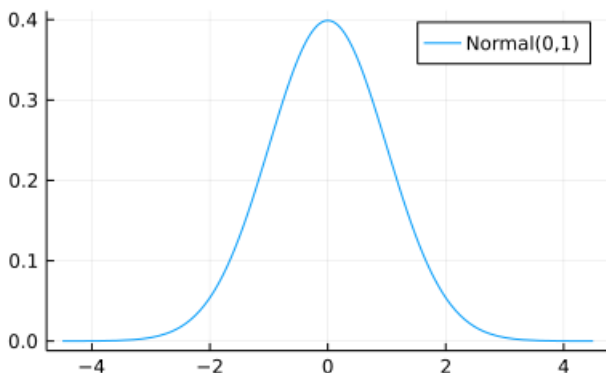
さらに、 $\sqrt{S^2/\sigma^2}$ の分布と $\sqrt{\text{Chisq}(n-1)/(n-1)}$ の比較も同時に行うことにしよう。($X \sim \text{Chisq}(\nu)$ のとき \sqrt{X} が従う分布を $\text{Chi}(\nu)$ と書き **χ 分布** と呼ぶ。)

正規分布の場合

これは理想的(過ぎる)場合である。

```
In [24]: 1 plot(Normal(), -4.5, 4.5; label="Normal(0,1)")
```

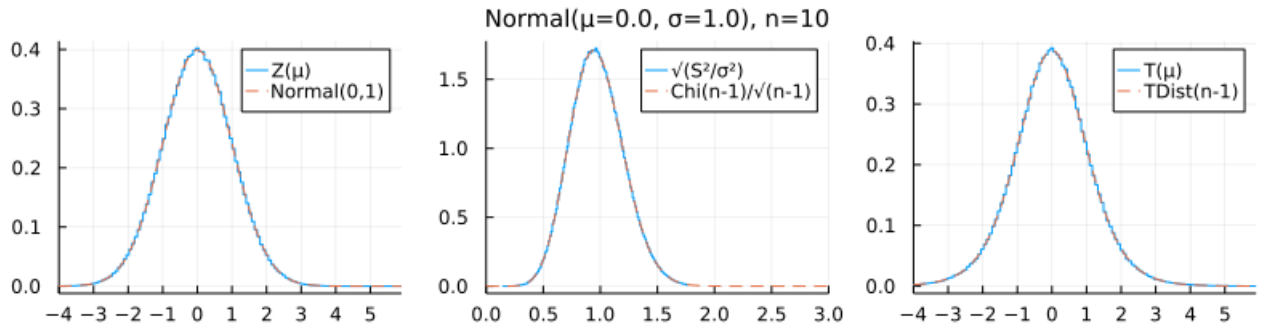
Out[24]:



```
In [25]: 1 plot_T(Normal(), 10, plot_chi=true, plot_t=true)
```

skewness = 0.0, kurtosis = 0.0

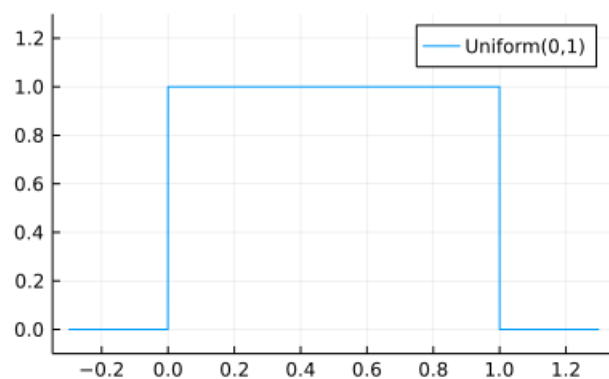
Out[25]:



一様分布の場合

```
In [26]: 1 plot(Uniform(), -0.3, 1.3; label="Uniform(0,1)", ylim=(-0.1, 1.3))
```

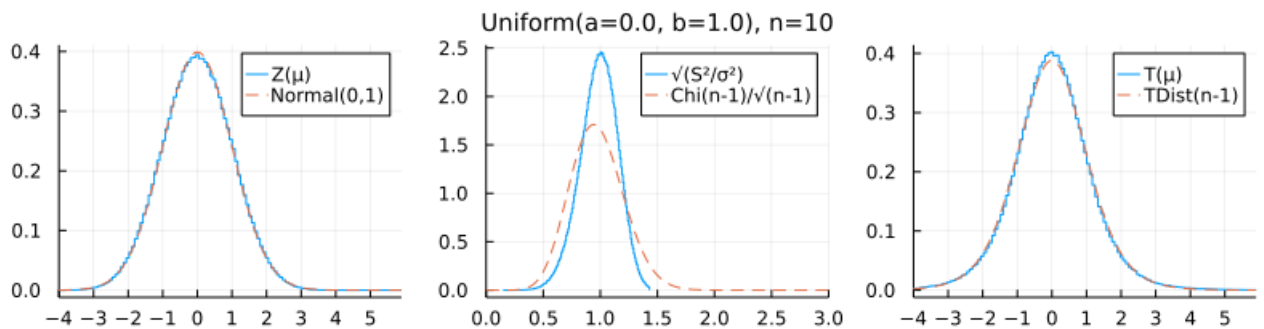
Out[26]:



```
In [27]: 1 plot_T(Uniform(), 10, plot_chi=true, plot_t=true)
```

skewness = 0.0, kurtosis = -1.2

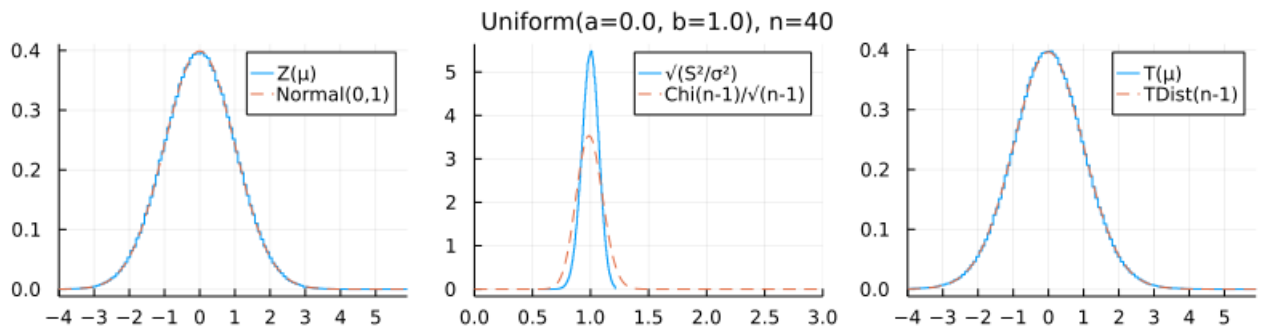
Out[27]:



```
In [28]: 1 plot_T(Uniform(), 40, plot_chi=true, plot_t=true)
```

skewness = 0.0, kurtosis = -1.2

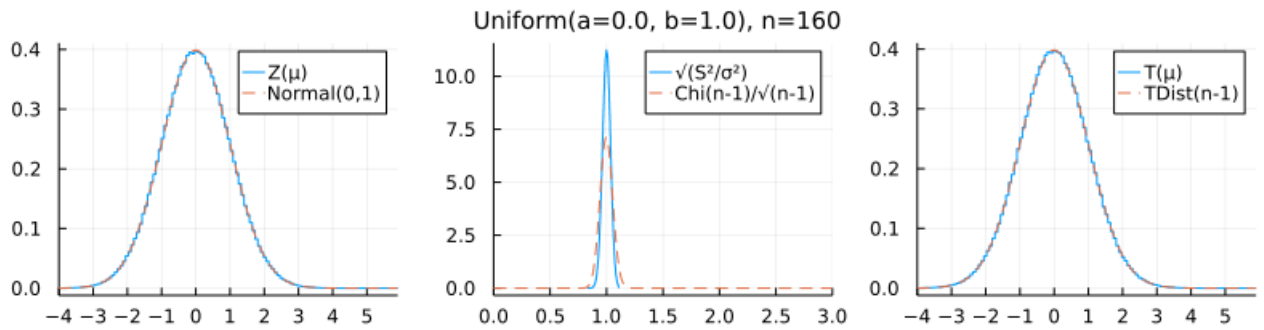
Out[28]:



```
In [29]: 1 plot_T(Uniform(), 160, plot_chi=true, plot_t=true)
```

skewness = 0.0, kurtosis = -1.2

Out[29]:



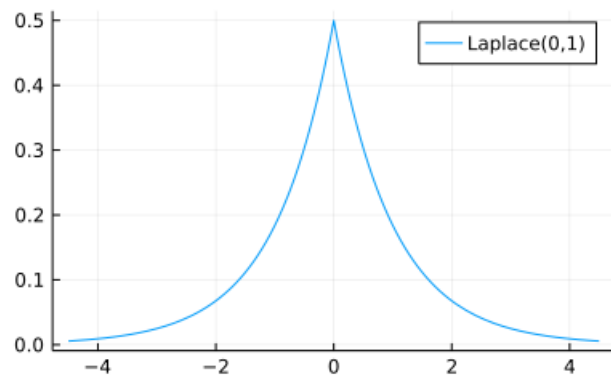
上の例を見れば、 $T(\mu)$ の分布が t 分布でよく近似されている状況であっても、 $(n-1)S^2/\sigma^2$ の分布が χ^2 分布で近似されていると考えるとはいけないことがわかる。この点は以下でも同様である。

だから、正規分布の標本分布モデルにおいて成り立っている $(n-1)S^2/\sigma^2 \sim \text{Chisq}(n-1)$ を用いて、分散 σ^2 の区間推定を行うことは大きなリスクが伴う。平均の区間推定と違って、分散の区間推定は非常に難しい。

Laplace分布の場合

```
In [30]: 1 plot(Laplace(), -4.5, 4.5; label="Laplace(0,1)")
```

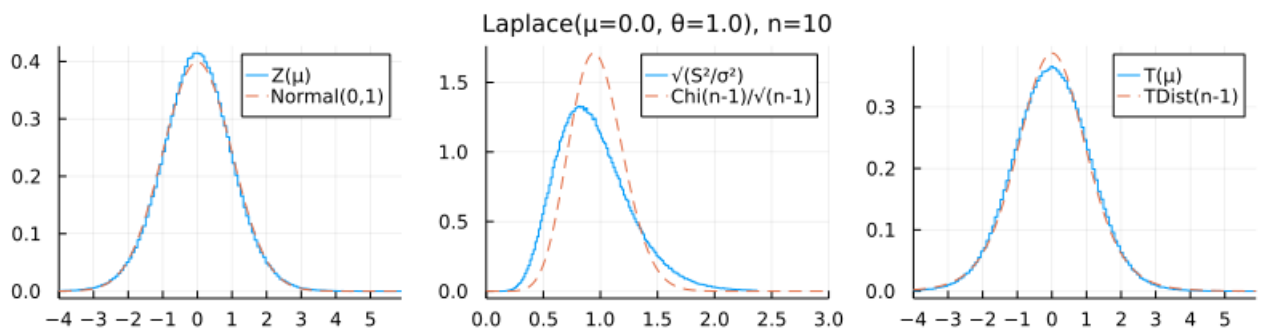
Out[30]:



```
In [31]: 1 plot_T(Laplace(), 10, plot_chi=true, plot_t=true)
```

skewness = 0.0, kurtosis = 3.0

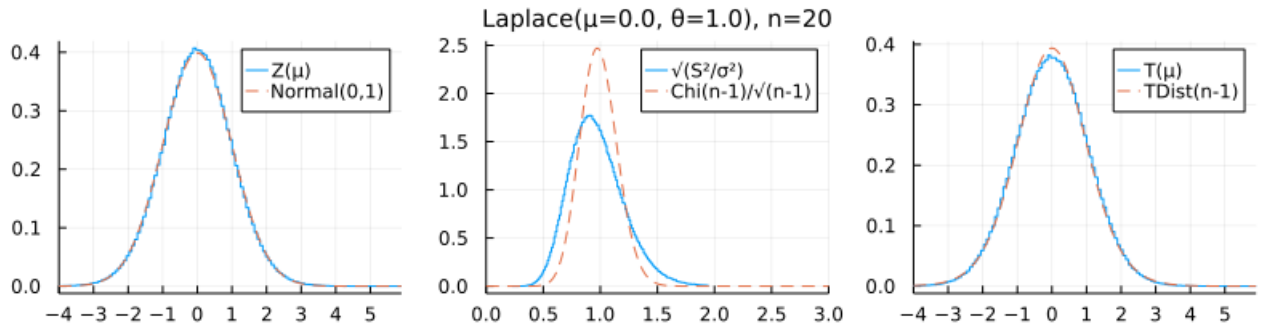
Out[31]:



```
In [32]: 1 plot_T(Laplace(), 20, plot_chi=true, plot_t=true)
```

skewness = 0.0, kurtosis = 3.0

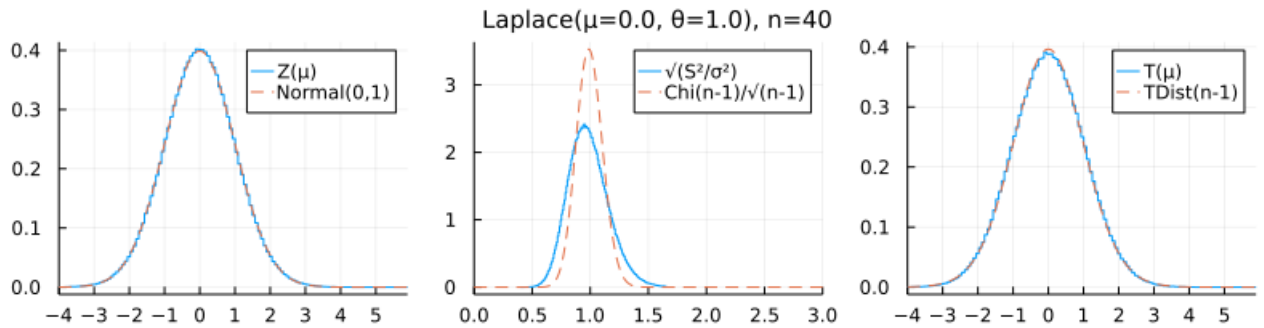
Out[32]:



```
In [33]: 1 plot_T(Laplace(), 40, plot_chi=true, plot_t=true)
```

skewness = 0.0, kurtosis = 3.0

Out[33]:

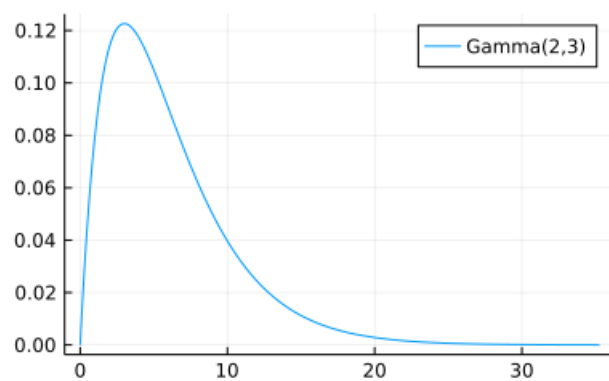


ガンマ分布の場合

ガンマ分布は左右非対称な分布の典型例である。

```
In [34]: 1 plot(Gamma(2,3); label="Gamma(2,3)")
```

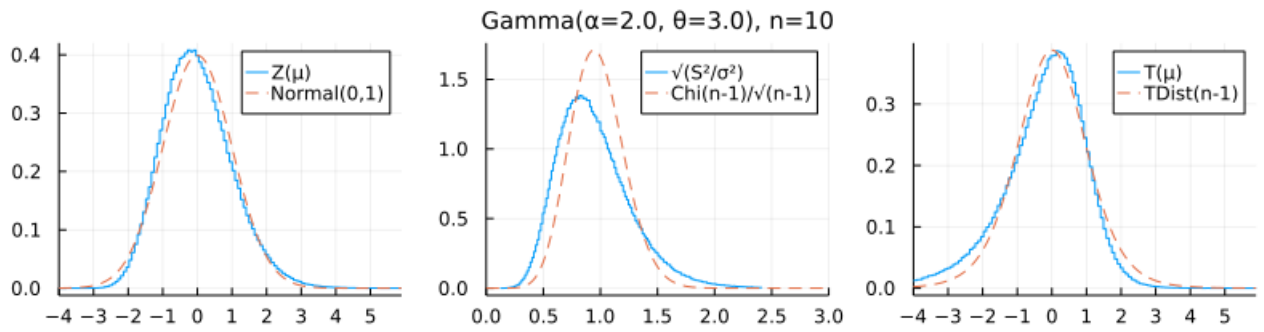
Out[34]:




```
In [35]: 1 plot_T(Gamma(2,3), 10, plot_chi=true, plot_t=true)
```

skewness = 1.414213562373095, kurtosis = 3.0

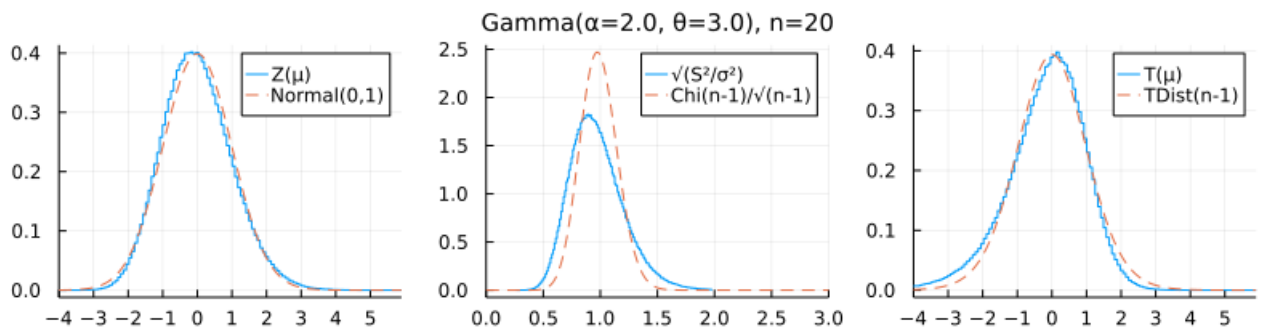
Out[35]:



```
In [36]: 1 plot_T(Gamma(2,3), 20, plot_chi=true, plot_t=true)
```

skewness = 1.414213562373095, kurtosis = 3.0

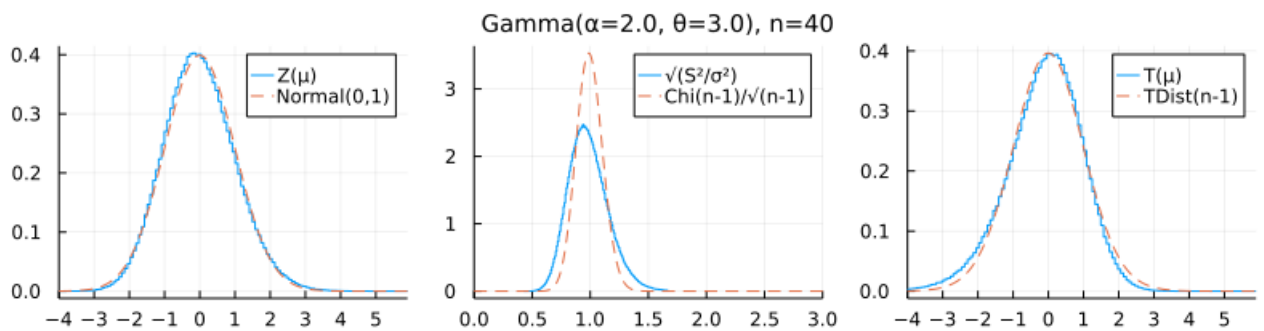
Out[36]:



```
In [37]: 1 plot_T(Gamma(2,3), 40, plot_chi=true, plot_t=true)
```

skewness = 1.414213562373095, kurtosis = 3.0

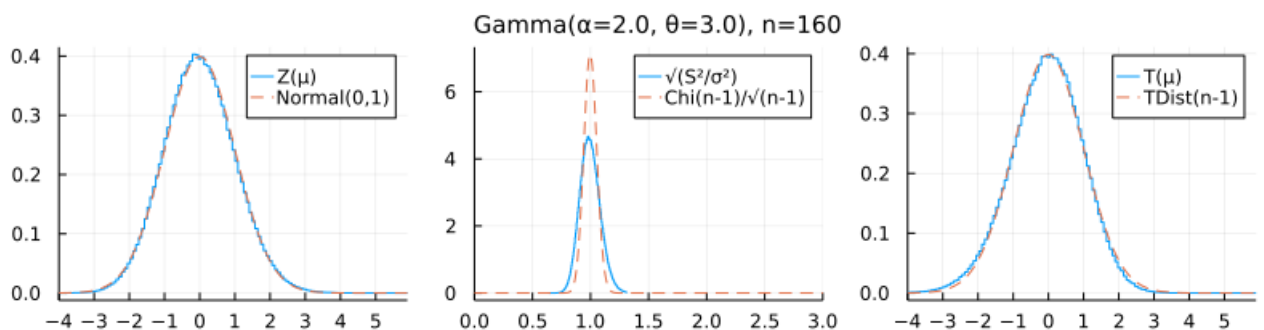
Out[37]:



```
In [38]: 1 plot_T(Gamma(2,3), 160, plot_chi=true, plot_t=true)
```

skewness = 1.414213562373095, kurtosis = 3.0

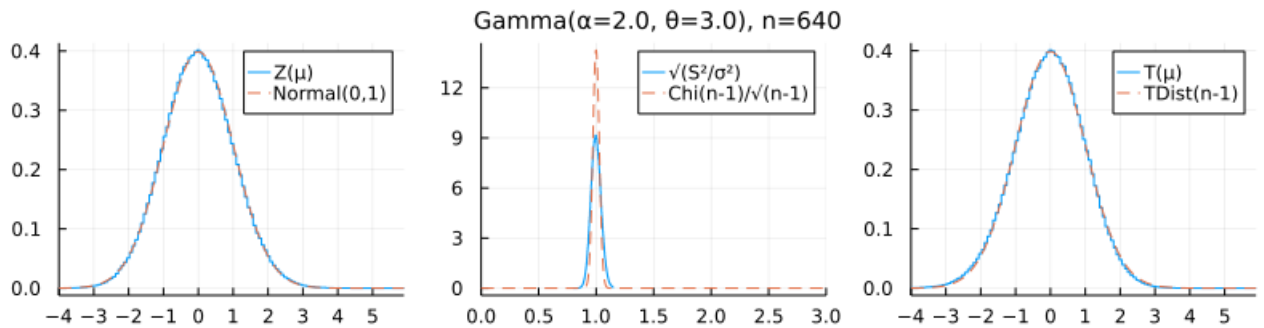
Out[38]:



```
In [39]: 1 plot_T(Gamma(2,3), 640, plot_chi=true, plot_t=true)
```

skewness = 1.414213562373095, kurtosis = 3.0

Out[39]:



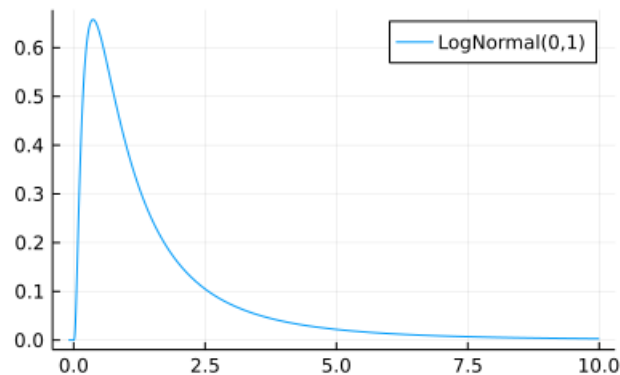
対数正規分布の場合

対数正規分布は左右非対称でかつ **外れ値が出易い** 分布の典型例である。

このような場合は n を非常に大きくしないと近似の精度が高まらない。

```
In [40]: 1 plot(LogNormal(), -0.1, 10; label="LogNormal(0,1)")
```

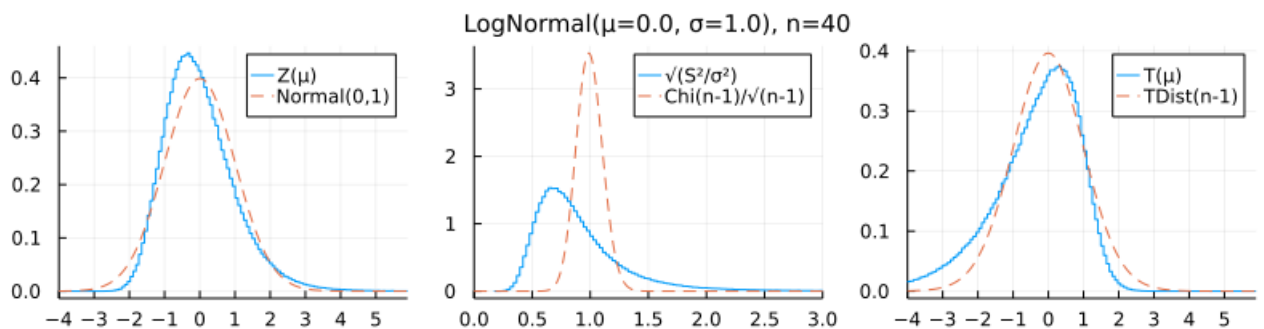
Out[40]:



```
In [41]: 1 plot_T(LogNormal(), 40, plot_chi=true, plot_t=true)
```

skewness = 6.184877138632554, kurtosis = 110.9363921763115

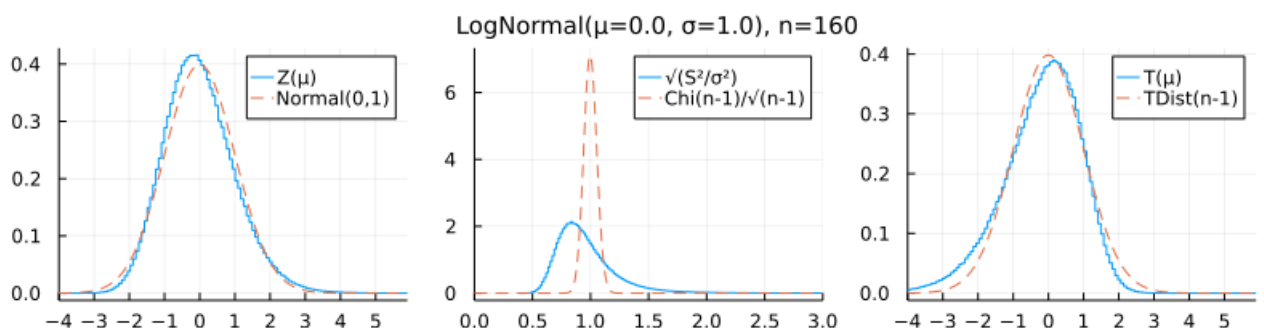
Out[41]:



```
In [42]: 1 plot_T(LogNormal(), 160, plot_chi=true, plot_t=true)
```

skewness = 6.184877138632554, kurtosis = 110.9363921763115

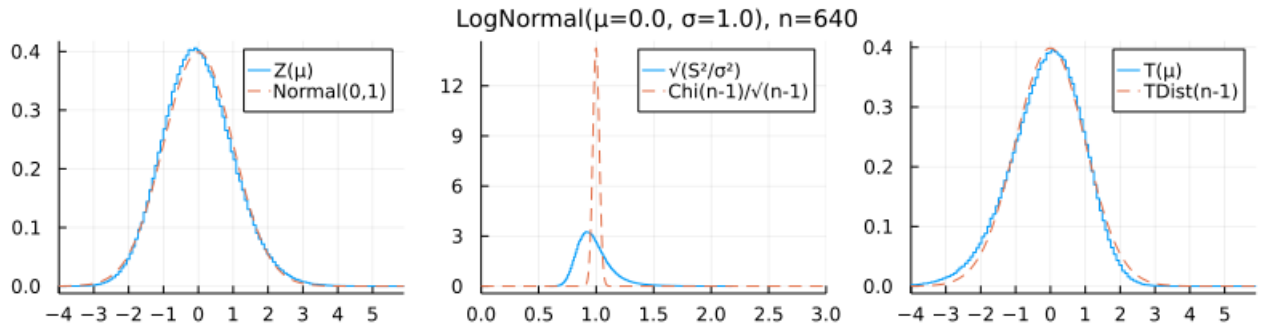
Out[42]:



```
In [43]: 1 plot_T(LogNormal(), 640, plot_chi=true, plot_t=true)
```

skewness = 6.184877138632554, kurtosis = 110.9363921763115

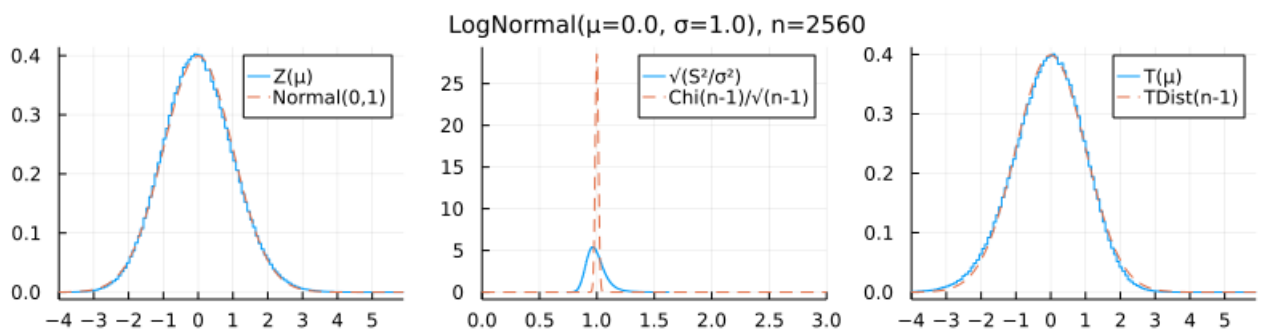
Out[43]:



```
In [44]: 1 plot_T(LogNormal(), 2560, plot_chi=true, plot_t=true)
```

skewness = 6.184877138632554, kurtosis = 110.9363921763115

Out[44]:



この例から、左右非対称で外れ値を含む母集団分布からの無作為抽出で得たデータを扱う場合には、中心極限定理を使った近似を用いることは要注意であることがわかる。

1.7 分布の対数の標本分布のT統計量の視覚化

データの数値 x_i がすべて正の値の場合には、 $y_i = \log x_i$ と対数を取った後に、「分布の対数の平均が λ_0 である」という仮説を扱った方が誤差が減る場合がある。

より一般に対数に限らず、データを適切な座標に変換してから統計分析にかけた方が誤差が小さくなる場合がある。

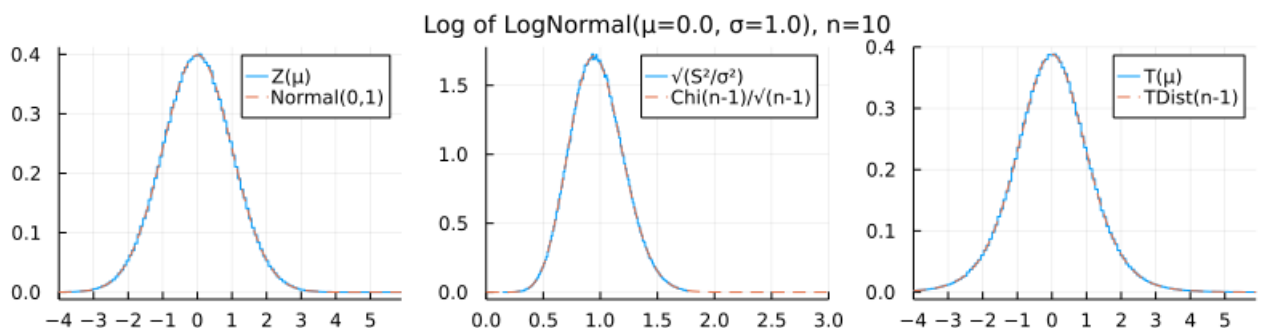
対数正規分布の場合

対数正規分布に従う確率変数の対数は定義より正規分布に従う。この場合には正値のデータの対数を取ってから計算することの効果は自明に絶大になる。(この場合は自明過ぎてあまり意味がない。)

```
In [45]: 1 plot_T(LogNormal(), 10, plot_chi=true, plot_t=true, logsample=true)
```

skewness = 6.184877138632554, kurtosis = 110.9363921763115

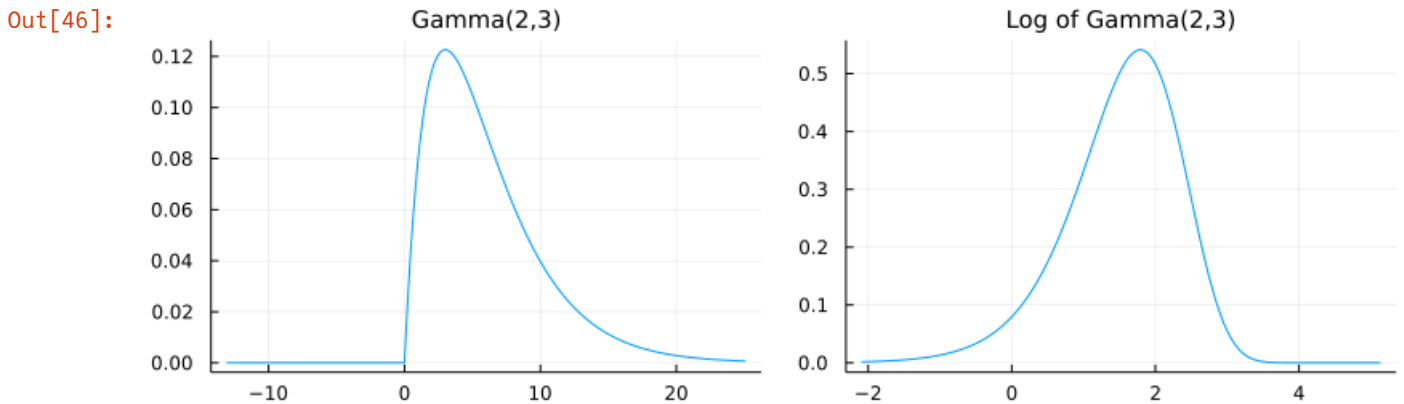
Out[45]:



ガンマ分布の場合

ガンマ分布とその対数の密度関数達

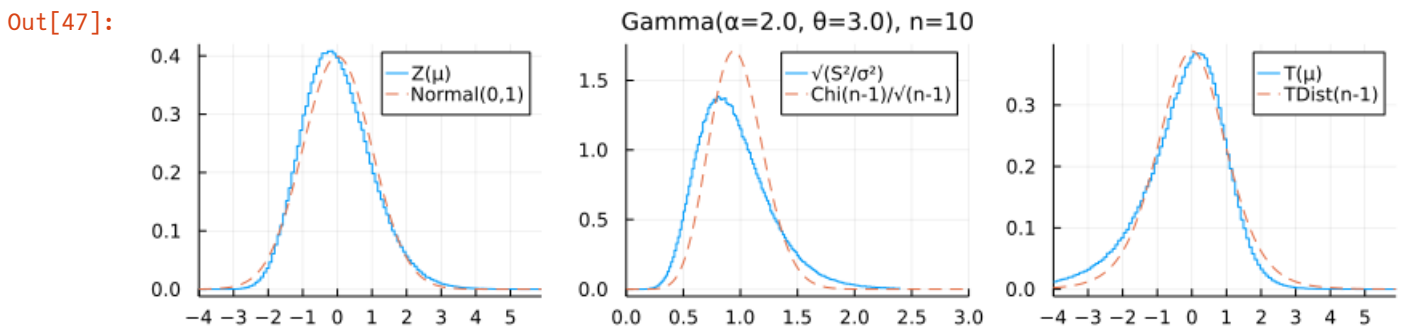
```
In [46]: 1 μ, σ = mean(Gamma(2, 3)), std(Gamma(2, 3))
2 P = plot(x → pdf(Gamma(2, 3), x), μ-4.5σ, μ+4.5σ; label="", title="Gamma(2,3)")
3 μ, σ = logmean(Gamma(2, 3)), logstd(Gamma(2, 3))
4 Q = plot(x → pdflog(Gamma(2, 3), x), μ-4.5σ, μ+4.5σ; label="", title="Log of Gamma(2,3)")
5 plot(P, Q; size=(800, 250))
```



対数を取る前

```
In [47]: 1 plot_T(Gamma(2,3), 10, plot_chi=true, plot_t=true)
```

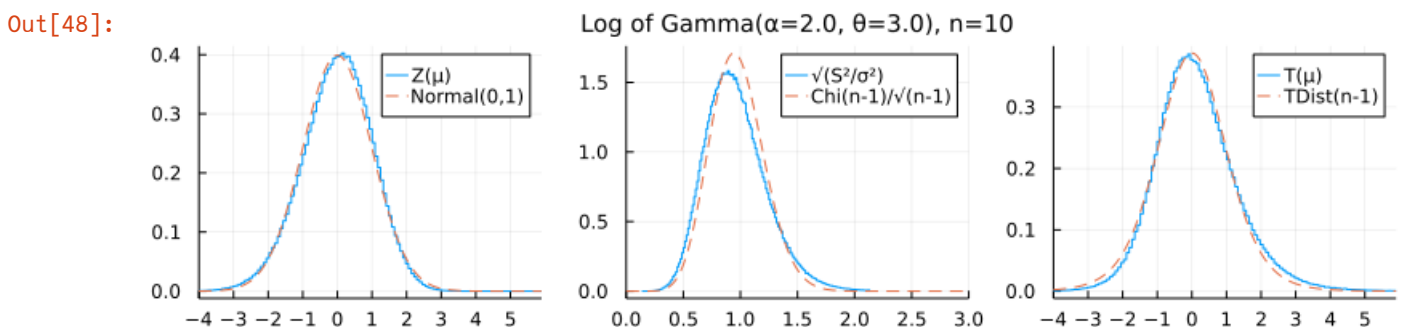
skewness = 1.414213562373095, kurtosis = 3.0



対数を取った後

```
In [48]: 1 plot_T(Gamma(2,3), 10, plot_chi=true, plot_t=true, logsample=true)
```

skewness = 1.414213562373095, kurtosis = 3.0



よく見ないと分からないかもしれないが、対数を取った後の方が近似の精度が上がっていることが分かる。

以上では対数を取ることによって誤差を減らすことができる場合があることを示したが、より一般に、データを適切な座標に変換してから統計分析にかけた方が誤差が小さくなる場合がある。

データの数値の生成のされ方に関してなにがしかの知識がある場合には、その知識をうまく統計分析の枠組みに取り入れた方が誤差が小さな分析が可能になる可能性が増える。

1.8 平均に関するP値の使い方

母集団からの無作為抽出で得たサイズ n のデータ x_1, \dots, x_n の標本平均と不偏分散をそれぞれ \bar{x} , s^2 と書くとき、そのデータの数値に関する仮説「平均は $\mu = \mu_0$ である」のP値は

データの数値と仮説「平均は $\mu = \mu_0$ である」の整合性の指標

として使われる。

ただし、P値を計算するときに用いた統計モデルや近似がうまく行っているというような仮定が不適切なことが原因で整合性の指標をおかしな値になってしまうリスクにも十分に配慮しなければいけない。さらに、データの取得法に問題があったことを疑わなければいけない場合は珍しくない(むしろ普通である)。統計分析の妥当性を確保するためには、様々な可能性を疑って、問題を一つひとつ丁寧に潰して行く必要がある。

有意水準と呼ばれる閾値 $0 \leq \alpha \leq 1$ を最初に決めておいて(目的に合わせて小さな正の値に取る)、P値が有意水準 α 未満になったときに、

データの数値と仮説「平均は $\mu = \mu_0$ である」のあいだに整合性が無さすぎる

と判定し、

データの数値によって仮説「平均は $\mu = \mu_0$ である」が **棄却** された

とすることがある。このような手続きを **検定**(testing, test)と呼ぶのであった。

1.9 P値の定義(2)のt分布を使う場合に対応する平均の信頼区間

有意水準を $0 \leq \alpha \leq 1$ と書き、自由度 ν の t 分布において $t_{\alpha/2}(\nu)$ 以上になる確率は $\alpha/2$ になると仮定する:

$$t_{\alpha/2, \nu} = \text{quantile}(\text{TDist}(\nu), 1 - \alpha/2).$$

例えば、

$$t_{5\%/2, 10} \approx 2.2281, \quad t_{5\%/2, 20} \approx 2.0860, \quad t_{5\%/2, 30} \approx 2.0423.$$

自由度を大きくする極限では、 $t_{5\%/2, \infty} = z_{5\%/2} \approx 1.9600$ となる。 $t_{5\%/2, 30} \approx 2.0423$ はその値よりも 4.2% 程度大きい。(信頼区間もその割合で広がる。)

P値 $\text{pvalue}_{\text{TDist}}(\bar{x}, s^2 | n, \mu = \mu_0) = 2(1 - \text{cdf}(\text{TDist}(n-2), |t(\mu_0)|))$ に対応する信頼度 $1 - \alpha$ の信頼区間は次のようになる:

$$\text{confint}_{\text{TDist}}(\bar{x}, s^2 | n, \mu = \mu_0) = \left[\bar{x} - t_{\alpha/2, n-1} \sqrt{s^2/n}, \bar{x} + t_{\alpha/2, n-1} \sqrt{s^2/n} \right].$$

これと $\text{confint}_{\text{Normal}}(\bar{x}, s^2 | n, \mu = \mu_0)$ の違いは $z_{\alpha/2}$ を $t_{\alpha/2, n-1}$ に置き換える分の違いしかない。

証明もP値の定義(1)の場合と同様である。標準正規分布を t 分布で置き換えるだけでよい。

証明: 検定と信頼区間の表裏一体性より、P値関数 $\text{pvalue}_{\text{TDist}}(\bar{x}, s^2 | n, \mu = \mu_0)$ に対応する信頼度 $1 - \alpha$ の信頼区間は次のように定義されるのであった:

$$\text{confint}_{\text{TDist}}(\bar{x}, s^2 | n, \mu = \mu_0) = \{ \mu_0 \in \mathbb{R} \mid \text{pvalue}_{\text{TDist}}(\bar{x}, s^2 | n, \mu = \mu_0) \geq \alpha \}.$$

そして、

$$\begin{aligned} \text{pvalue}_{\text{TDist}}(\bar{x}, s^2 | n, \mu = \mu_0) &\geq \alpha \\ \iff 1 - \text{cdf}(\text{TDist}(n-1), |t(\mu_0)|) &\geq \alpha/2 \\ \iff |t(\mu_0)| = \frac{|\bar{x} - \mu_0|}{\sqrt{s^2/n}} &\leq t_{\alpha/2, n-1} \\ \iff \bar{x} - t_{\alpha/2, n-1} \sqrt{s^2/n} &\leq \mu_0 \leq \bar{x} + t_{\alpha/2, n-1} \sqrt{s^2/n}. \end{aligned}$$

これより、P値の定義(2)に対応する平均の信頼区間が上のようになることがわかった。

証明終

2 平均に関するP値と信頼区間の計算例

2.1 平均に関するP値と信頼区間の計算の仕方

(0) 前提

有意水準: $0 < \alpha < 1$.

検定したい仮説: 平均は $\mu = \mu_0$ である.

データの数値: x_1, \dots, x_n .

母集団からの無作為抽出(または同分布の独立試行)で得られたサイズ n のデータの数値を x_1, \dots, x_n とする.

(1) データの数値の標本平均 \bar{x} と不偏分散 s^2 を計算する:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

(2) データの数値の t 値 $t(\mu_0)$ を計算する:

$$t = t(\mu_0) = \frac{\bar{x} - \mu_0}{\sqrt{s^2/n}}.$$

(3) 仮説「平均は $\mu = \mu_0$ である」のP値を自由度 $n-1$ の t 分布に従ってランダムに生成される値の絶対値がデータの数値の t 値の絶対値以上になる確率として求める:

$$(\text{仮説「平均は } \mu = \mu_0 \text{ である」のP値}) = 2(1 - \text{cdf}(\text{TDist}(n-1), |t(\mu_0)|)).$$

(4) 定数 $t_{\alpha/2, n-1}$ を, 自由度 $n-1$ の t 分布に従ってランダムに生成される値が $t_{\alpha/2, n-1}$ 以上になる確率は $\alpha/2$ であるという条件によって定める:

$$t_{\alpha/2, n-1} = \text{quantile}(\text{TDist}(n-1), 1 - \alpha/2).$$

(5) 次の公式によって, 平均 μ に関する信頼度 $1 - \alpha$ の信頼区間を求める:

$$(\text{平均 } \mu \text{ に関する信頼度 } 1 - \alpha \text{ の信頼区間}) = \left[\bar{x} - t_{\alpha/2, n-1} \sqrt{s^2/n}, \bar{x} + t_{\alpha/2, n-1} \sqrt{s^2/n} \right].$$

注意・警告: 中心極限定理による近似が十分にうまく行っているという仮定が成立していない場合には以上で求めたP値や信頼区間の誤差は大きくなる. 特にデータを取得した母集団の分布が左右非対称で外れ値を含んでいる場合には相当にサンプルサイズ n を大きくしないと誤差が小さくならない. そのような場合には何らかの工夫や別の方法を検討した方がよいかもしれない.

2.2 平均に関するP値と信頼区間の計算例

サイズ $n = 11$ のデータの数値: 13.7, 12.9, 4.4, 5.2, 3.1, 2.9, 7.2, 10.3, 4.7, 4.6, 3.6

標本平均: $\text{mean}(13.7, 12.9, 4.4, 5.2, 3.1, 2.9, 7.2, 10.3, 4.7, 4.6, 3.6) = 6.6$

不偏分散: $\text{var}(13.7, 12.9, 4.4, 5.2, 3.1, 2.9, 7.2, 10.3, 4.7, 4.6, 3.6) = 15.35$

検定したい仮説: 平均は $\mu = 9.0$ である.

t値: $(6.6 - 9.0)/\sqrt{15.35/11} \approx -2.03167106$

P値: $2(1 - \text{cdf}(\text{TDistribution}(10), 2.03167106)) \approx 0.0696099$

有意水準: $\alpha = 5\% = 0.05$

t分布の分位点: $c = t_{\alpha/2, n-1} = \text{quantile}(\text{TDistribution}(10), 0.975) \approx 2.22814$

信頼区間: $6.6 - 2.22814 \sqrt{15.35/11}, 6.6 + 2.22814 \sqrt{15.35/11} = 3.96791, 9.23209$

以上の計算法は[WolframAlpha \(https://www.wolframalpha.com/\)](https://www.wolframalpha.com/)でそのまま使える.

2.2.1 WolframAlphaでの平均に関するP値と信頼区間の計算例

$\text{mean}(13.7, 12.9, 4.4, 5.2, 3.1, 2.9, 7.2, 10.3, 4.7, 4.6, 3.6) \rightarrow$ [実行 \(https://www.wolframalpha.com/input?i=mean%2813.7%2C+12.9%2C+4.4%2C+5.2%2C+3.1%2C+2.9%2C+7.2%2C+10.3%2C+4.7%2C+4.6%2C+3.6%29\)](https://www.wolframalpha.com/input?i=mean%2813.7%2C+12.9%2C+4.4%2C+5.2%2C+3.1%2C+2.9%2C+7.2%2C+10.3%2C+4.7%2C+4.6%2C+3.6%29)

$\text{var}(13.7, 12.9, 4.4, 5.2, 3.1, 2.9, 7.2, 10.3, 4.7, 4.6, 3.6) \rightarrow$ [実行 \(https://www.wolframalpha.com/input?i=var%2813.7%2C+12.9%2C+4.4%2C+5.2%2C+3.1%2C+2.9%2C+7.2%2C+10.3%2C+4.7%2C+4.6%2C+3.6%29\)](https://www.wolframalpha.com/input?i=var%2813.7%2C+12.9%2C+4.4%2C+5.2%2C+3.1%2C+2.9%2C+7.2%2C+10.3%2C+4.7%2C+4.6%2C+3.6%29)

$(6.6 - 9.0)/\sqrt{15.35/11} \rightarrow$ [実行 \(https://www.wolframalpha.com/input?i=%286.6+-9.0%29%2Fsqrt%2815.35%2F11%29\)](https://www.wolframalpha.com/input?i=%286.6+-9.0%29%2Fsqrt%2815.35%2F11%29)

$2(1 - \text{cdf}(\text{TDistribution}(10), 2.03167106)) \rightarrow$ [実行 \(https://www.wolframalpha.com/input?i=2%281+-cdf%28TDistribution%2810%29%2C+2.03167106%29%29\)](https://www.wolframalpha.com/input?i=2%281+-cdf%28TDistribution%2810%29%2C+2.03167106%29%29)

quantile(TDistribution(10), 0.975) → 実行 (<https://www.wolframalpha.com/input?i=quantile%28TDistribution%2810%29%2C0.975%29>)

6.6 - 2.22814 sqrt(15.35/11), 6.6 + 2.22814 sqrt(15.35/11) → 実行 (<https://www.wolframalpha.com/input?i=6.6+-+2.22814+sqrt%2815.35%2F11%29%2C6.6+-+2.22814+sqrt%2815.35%2F11%29>)

2.2.2 Julia言語での平均に関するP値と信頼区間の計算例

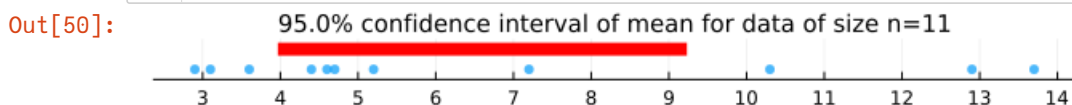
```
In [49]: 1 x = [13.7, 12.9, 4.4, 5.2, 3.1, 2.9, 7.2, 10.3, 4.7, 4.6, 3.6]
2 n = length(x)
3 x̄ = mean(x)
4 s² = var(x)
5
6 μ = 9.0
7 t_value = (x̄ - μ) / √(s²/n)
8 P_value = 2ccdf(TDist(n-1), abs(t_value))
9
10 α = 0.05
11 c = quantile(TDist(n-1), 1 - α/2)
12 confidence_interval = [x̄ - c*√(s²/n), x̄ + c*√(s²/n)]
13
14 @show x n x̄ s²
15 println()
16 @show t_value P_value
17 println()
18 @show α c confidence_interval;
```

```
x = [13.7, 12.9, 4.4, 5.2, 3.1, 2.9, 7.2, 10.3, 4.7, 4.6, 3.6]
n = 11
x̄ = 6.6
s² = 15.35
```

```
t_value = -2.031671060092391
P_value = 0.06960994092988786
```

```
α = 0.05
c = 2.228138851986274
confidence_interval = [3.967913856820957, 9.232086143179043]
```

```
In [50]: 1 plot_confint_of_mean(x; α)
```



この図の区間を見ると、一瞬

データを示す青丸達のかかなりの部分が区間をはみ出している!

と思ってしまいがちだが、この区間はデータの分布の範囲を表す区間ではなく、モデルの分布の平均パラメータ μ に関する区間であることを思い出せばこれで問題ないことがわかる。

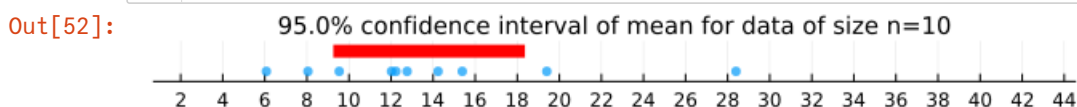
以下の図のようにサンプルサイズ n を大きくすれば、平均の信頼区間の幅は狭くなる。

サンプルサイズ n を a 倍すると信頼区間の幅はおおよそ \sqrt{a} 分の 1 になる。

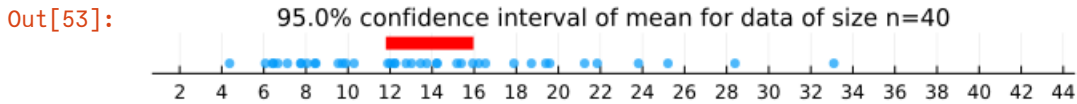
```
In [51]: 1 Random.seed!(4649373)
2 dist = Gamma(30, 1)
3 x = rand(Gamma(5,3), 640)
4 a, b = extrema(x)
5 xlim = (a - 0.05(b - a), b + 0.05(b - a))
```

Out[51]: (0.6804657464956572, 44.56206635799026)

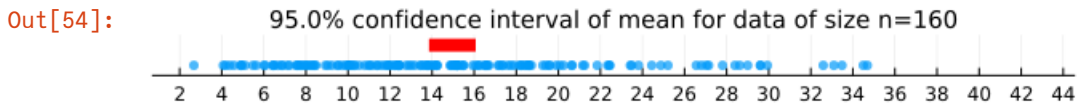
```
In [52]: 1 plot_confint_of_mean(x[1:10]; xlim, xtick=0:2:100)
```



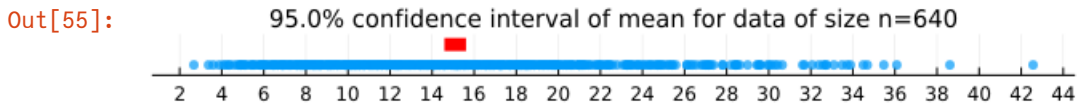
```
In [53]: 1 plot_confint_of_mean(x[1:40]; xlim, xtick=0:2:100)
```



```
In [54]: 1 plot_confint_of_mean(x[1:160]; xlim, xtick=0:2:100)
```



```
In [55]: 1 plot_confint_of_mean(x; xlim, xtick=0:2:100)
```



2.2.3 R言語での平均に関するP値と信頼区間の計算例

```
x = c(13.7, 12.9, 4.4, 5.2, 3.1, 2.9, 7.2, 10.3, 4.7, 4.6, 3.6)
n = length(x)
xbar = mean(x)
s2 = var(x)
mu = 9
t.value = (xbar - mu)/sqrt(s2/n)
p.value = 2*(1 - pt(abs(t.value), n-1))
alpha = 0.05
c = qt(1 - alpha/2, n-1)
conf.int = c(xbar - c*sqrt(s2/n), xbar + c*sqrt(s2/n))

cat("data: x = ", x, "\n")
cat("data size n = ", n, "\n")
cat("sample mean of data = ", xbar, "\n")
cat("unbiased variance of data = ", s2, "\n")
cat("\n")
cat("t.value = ", t.value, "\n")
cat("p.value = ", p.value, "\n")
cat("\n")
cat("alpha = ", alpha, "\n")
cat("t_{alpha/2, n-1} = ", c, "\n")
cat("conf.int = ", conf.int, "\n")
```

以下のようにすれば一発で計算できる:

```
t.test(c(13.7, 12.9, 4.4, 5.2, 3.1, 2.9, 7.2, 10.3, 4.7, 4.6, 3.6), mu=9)
```


In [56]:

```
1 R""  
2 x = c(13.7, 12.9, 4.4, 5.2, 3.1, 2.9, 7.2, 10.3, 4.7, 4.6, 3.6)  
3 n = length(x)  
4 xbar = mean(x)  
5 s2 = var(x)  
6 mu = 9  
7 t.value = (xbar - mu)/sqrt(s2/n)  
8 p.value = 2*(1 - pt(abs(t.value), n-1))  
9 alpha = 0.05  
10 c = qt(1 - alpha/2, n-1)  
11 conf.int = c(xbar - c*sqrt(s2/n), xbar + c*sqrt(s2/n))  
12  
13 cat("data: x = ", x, "\n")  
14 cat("data size n = ", n, "\n")  
15 cat("sample mean of data = ", xbar, "\n")  
16 cat("unbiased variance of data = ", s2, "\n")  
17 cat("\n")  
18 cat("t.value = ", t.value, "\n")  
19 cat("p.value = ", p.value, "\n")  
20 cat("\n")  
21 cat("alpha = ", alpha, "\n")  
22 cat("t_{alpha/2, n-1} = ", c, "\n")  
23 cat("conf.int = ", conf.int, "\n")  
24 "";
```

```
data: x = 13.7 12.9 4.4 5.2 3.1 2.9 7.2 10.3 4.7 4.6 3.6  
data size n = 11  
sample mean of data = 6.6  
unbiased variance of data = 15.35  
  
t.value = -2.031671  
p.value = 0.06960994  
  
alpha = 0.05  
t_{alpha/2, n-1} = 2.228139  
conf.int = 3.967914 9.232086
```

In [57]:

```
1 R""  
2 t.test(c(13.7, 12.9, 4.4, 5.2, 3.1, 2.9, 7.2, 10.3, 4.7, 4.6, 3.6), mu=9)  
3 ""
```

Out[57]: RObject{VecSxp}

One Sample t-test

```
data: c(13.7, 12.9, 4.4, 5.2, 3.1, 2.9, 7.2, 10.3, 4.7, 4.6, 3.6)  
t = -2.0317, df = 10, p-value = 0.06961  
alternative hypothesis: true mean is not equal to 9  
95 percent confidence interval:  
 3.967914 9.232086  
sample estimates:  
mean of x  
 6.6
```

2.3 必修練習問題: 平均に関するP値と信頼区間の計算

次のサイズ 20 データとして次が与えられているとする:

```
38.1, 28.9, 30.4, 27.6, 38.0, 35.3, 30.0, 29.3, 32.9, 34.8,  
39.4, 38.4, 28.4, 27.0, 35.9, 30.7, 28.2, 33.2, 33.0, 36.0
```

以上で説明した方法が以下を求めよ.

(1) 「平均は $\mu = 30.0$ である」という仮説のP値.

(2) 平均の 99% 信頼区間.

まずは, この問題よりも上の方のみを見て(これより下を見ない), 計算してみよ.

がんばって計算した結果が正解とぴったり合っていると非常に気分のよい思いをすることだろう.

mean(38.1, 28.9, ... 36.0) → 実行 ([https://www.wolframalpha.com/input?](https://www.wolframalpha.com/input?i=mean%2838.1%2C+28.9%2C+30.4%2C+27.6%2C+38.0%2C+35.3%2C+30.0%2C+29.3%2C+32.9%2C+34.8%2C+39.4%2C+38.4%2C+36.0%2C%29)
[i=mean%2838.1%2C+28.9%2C+30.4%2C+27.6%2C+38.0%2C+35.3%2C+30.0%2C+29.3%2C+32.9%2C+34.8%2C+39.4%2C+38.4%2C+36.0%2C%29](https://www.wolframalpha.com/input?i=mean%2838.1%2C+28.9%2C+30.4%2C+27.6%2C+38.0%2C+35.3%2C+30.0%2C+29.3%2C+32.9%2C+34.8%2C+39.4%2C+38.4%2C+36.0%2C%29))
→ 32.775

var(38.1, 28.9, ... 36.0) → 実行 ([https://www.wolframalpha.com/input?](https://www.wolframalpha.com/input?i=var%2838.1%2C+28.9%2C+30.4%2C+27.6%2C+38.0%2C+35.3%2C+30.0%2C+29.3%2C+32.9%2C+34.8%2C+39.4%2C+38.4%2C+36.0%2C%29)
[i=var%2838.1%2C+28.9%2C+30.4%2C+27.6%2C+38.0%2C+35.3%2C+30.0%2C+29.3%2C+32.9%2C+34.8%2C+39.4%2C+38.4%2C+36.0%2C%29](https://www.wolframalpha.com/input?i=var%2838.1%2C+28.9%2C+30.4%2C+27.6%2C+38.0%2C+35.3%2C+30.0%2C+29.3%2C+32.9%2C+34.8%2C+39.4%2C+38.4%2C+36.0%2C%29))
→ 16.1588

$\sqrt{s^2/n}$ の計算: sqrt(16.1588/20) → 実行 (<https://www.wolframalpha.com/input?i=sqrt%2816.1588%2F20%29>) → 0.898855

t値の計算: (32.775 - 30)/0.898855 → 実行 (<https://www.wolframalpha.com/input?i=%2832.775+-30%29%2F0.898855%29>) → 3.087261

$\alpha = 0.01, n - 1 = 19$ のときの $t_{\alpha/2, n-1}$ の計算: `quantile(TDistribution(19), 0.995) → 実行`
<https://www.wolframalpha.com/input?i=quantile%28TDistribution%2819%29%2C0.995%29> → 2.86093

```
In [58]: 1 x = [38.1, 28.9, 30.4, 27.6, 38.0, 35.3, 30.0, 29.3, 32.9, 34.8,
2         39.4, 38.4, 28.4, 27.0, 35.9, 30.7, 28.2, 33.2, 33.0, 36.0]
3 show_confint_of_mean(x;  $\mu=30$ ,  $\alpha=0.01$ )
4 plot_confint_of_mean(x;  $\alpha=0.01$ )
```

Out[58]:

99.0% confidence interval of mean for data of size n=20

The plot shows a distribution of data points (blue dots) and a red horizontal bar representing the 99.0% confidence interval of the mean. The x-axis is labeled from 27 to 40. The data points are clustered between 27 and 39, with a peak around 33-34. The red bar is centered around 30.5, spanning from approximately 30.1 to 31.1.

```
In [59]: 1 R""
2 x = c(38.1, 28.9, 30.4, 27.6, 38.0, 35.3, 30.0, 29.3, 32.9, 34.8,
3         39.4, 38.4, 28.4, 27.0, 35.9, 30.7, 28.2, 33.2, 33.0, 36.0)
4 t.test(x, mu=30, conf.level=0.99)
5 """
```

(2) 99%信頼区間 $\approx [30.20, 35.35]$

In []: 1