Bernoulli試行と関連確率分布

- 黒木玄

- 2022-04-09〜2022-05-21

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)

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 WolframAlphaの使い方
2 一様乱数の生成の繰り返し
3 Bernoulli試行
4 離散分布
4.1 離散分布と確率質量函数
4.2 離散確率変数
4.3 離散確率変数の期待値
4.4 離散確率変数の期待値を取る操作の基本性質
4.5 確率変数の一般的な定義
4.6 離散確率変数に関する確率
4.7 離散確率変数の分散と標準偏差
4.8 問題: 期待値と分散の特徴付け
4.9 問題: 平均二乗誤差のバイアス・バリアンス分解
4.10 カテゴリカル分布の定義
4.11 問題: ルーレットの例の賞金の期待値と分散
4.12 Bernoulli分布の定義
4.13 問題: Bernoulli分布の期待値と分散
4.14 オッズの定義と例
4.15 オッズとロジットおよびロジスティック変換
4.16 Bernoulli試行の確率分布
4.17 問題: n=3, p=0.3 の場合のBernoulli試行の確率分布
4.18 Bernoulli試行から応用上基本的かつ重要な確率分布が大量に得られること
4.19 二項定理
4.20 二項分布の定義
4.21 問題: 二項分布の期待値と分散
4.22 問題: 当たりが30%の確率で出るくじを100回引いた場合
4.23 幾何分布の定義
4.24 幾何分布の期待値と分散
4.25 問題: 当たりが出る確率が1%のガチャを当たりが出るまで回す場合
4.26 負の二項展開
4.27 負の二項分布の定義
4.28 負の二項分布の期待値と分散
4.29 問題: 当たりが出る確率が5%のガチャを当たりが5回出るまで回す場合
4.30 多項分布の定義
5 連続分布など
5.1 連続分布と確率密度函数
5.2 一様分布の定義
5.3 正規分布の定義
5.4 連続確率変数
5.5 連続確率変数に関する確率とその期待値
5.6 確率変数の確率密度函数
5.7 確率変数の確率質量函数
5.8 連続分布における確率変数の期待値を取る操作の基本性質

5.9 問題: Markovの不等式とChebyshevの不等式
5.10 問題: Jensenの不等式
5.11 問題: 相加相乗平均の不等式
5.12 問題: Kullback-Leibler情報量とGibbsの情報不等式
5.13 連続分布における確率変数の分散
5.14 問題: 正規分布モデルによる他の分布の最良近似
5.15 確率変数の一次函数による変換(アフィン変換)の期待値と分散
5.16 分布のアフィン変換, スケール変換, 平行移動, 逆数の定義
5.17 確率変数の累積分布函数
5.18 分位点函数 (quantile function)
5.19 Bernoulli分布の累積分布函数
5.20 カテゴリカル分布の累積分布函数
5.21 二項分布の累積分布函数のグラフ
5.22 負の二項分布の累積分布函数のグラフ
5.23 一様分布の累積分布函数
5.24 標準正規分布の累積分布函数と分位点函数
5.25 問題: 標準正規分布の四分位数
5.26 問題: 連続分布の累積分布函数は一様分布に従う.
5.27 分位点函数への一様乱数の代入で元の分布の乱数を作れること
5.28 一様分布の期待値と分散
5.29 標準正規分布の期待値と分散
5.30 一様乱数の生成の繰り返しの分布

1 WolframAlphaの使い方

WolframAlpha (https://www.wolframalpha.com/)を利用すれば以下のような計算が可能である. 例えば,

$$(10000-2020)^2\cdot 0.1+(3000-2020)^2\cdot 0.2+(1000-2020)^2\cdot 0.3+(300-2020)^2\cdot 0.4$$

は次のように入力すれば計算できる:

- (10000 - 2020)^2 0.1 + (3000 - 2020)^2 0.2 + (1000 - 2020)^2 0.3 + (300 - 2020)^2 0.4 (https://www.wolframalpha.com/input?i=%2810000+-+2020%29%5E2+0.1+%2B+%283000+-+2020%29%5E2+0.2+%2B+%281000+-+2020%29%5E2+0.3+%2B+%28300+-+2020%29%5E2+0.4)

二項係数を含む式

$$\sum_{k=5}^{300}\binom{300}{k}0.01^k(1-0.01)^{300-k}$$

は次のように入力すれば計算できる:

- sum_{k=5}^{300} binom(300, k) 0.01^k (1 - 0.01)^(300-k) (https://www.wolframalpha.com/input?i=sum_%7Bk%3D5%7D%5E%7B300%7D+binom%28300%2C+k%29+0.01%5Ek+%281+-+0.01%29%5E%28300-k%29)

これは当たりが出る確率が1%のガチャを300回まわしたときに幸運にも当たりが5回以上出る確率になっている.

積分

$$2\int_{1.96}^{\infty}\frac{\exp(-x^2/2)}{\sqrt{2\pi}}\,dx$$

は次のように入力すれば計算できる:

- 2 int_{1.96}^{infty} exp(-x^2/2)/sqrt(2π) dx (https://www.wolframalpha.com/input?i=2+int_%7B1.96%7D%5E%7Binfty%7D+exp%28-x%5E2%2F2%29%2Fsqrt%282%CF%80%29+dx)

これはほぼ $5\% = 0.05$ になる.

統計学での練習問題を解くためにはこの手の計算を何らかの手段でできるようになっておく必要がある.

In [1]:
```julia
ENV["LINES"], ENV["COLUMNS"] = 100, 100
using BenchmarkTools
using Distributions
using Printf
using QuadGK
using Random
Random.seed!(4649373)
using Roots
using SpecialFunctions
using StaticArrays
using StatsBase
using StatsFuns
using StatsPlots
default(fmt = :png, titlefontsize = 10, size = (400, 250))
using SymPy
```

In [2]:
```julia
# Override the Base.show definition of SymPy.jl:
# https://github.com/JuliaPy/SymPy.jl/blob/29c5bfd1d10ac53014fa7fef468bc8deccadc2fc/src/types.

@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::SymbolicObject)
    print(io, as_markdown("\\displaystyle " * sympy.latex(x, mode="plain", fold_short_frac=fal
end
@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::AbstractArray{Sym})
    function toeqnarray(x::Vector{Sym})
        a = join(["\\displaystyle " * sympy.latex(x[i]) for i in 1:length(x)], "\\\\")
        """\\left[ \\begin{array}{r}$a\\end{array} \\right]"""
    end
    function toeqnarray(x::AbstractArray{Sym,2})
        sz = size(x)
        a = join([join("\\displaystyle " .* map(sympy.latex, x[i,:]), "&") for i in 1:sz[1]],
        "\\left[ \\begin{array}{" * repeat("r",sz[2]) * "}" * a * "\\end{array}\\right]"
    end
    print(io, as_markdown(toeqnarray(x)))
end
```

2 一様乱数の生成の繰り返し

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

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

In [3]:
```julia
[rand() for _ in 1:10]
```

Out[3]:
```
10-element Vector{Float64}:
 0.5105030210947423
 0.13229885279256504
 0.05873542170272006
 0.742812017444488
 0.8417141593154808
 0.44238748406950934
 0.5002487759410339
 0.7049642022880609
 0.6687158029467306
 0.7672724745322447
```

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

上と同じ結果を与えるコードは次のようにもっとシンプルに書ける.

In [4]:
```
rand(10)
```

Out[4]:
```
10-element Vector{Float64}:
 0.7702439301833676
 0.6263555234233755
 0.4983977148383013
 0.6011722501239075
 0.18013689329541094
 0.37632266071643605
 0.13414398433355734
 0.9652043335626921
 0.7352525630207614
 0.23381276022709152
```

3 Bernoulli試行

確率 p で 1 を, 確率 $1-p$ で 0 をランダムに毎回独立に生成することの繰り返しを **Bernoulli試行** (ベルヌーイ試行, ベルヌイ試行)と呼ぶ.

Bernoulli試行は上で説明した `rand()` の繰り返しに似ており, 乱数の値が 1 と 0 しかない場合の乱数列生成だとみなされる.

In [5]:
```
@show rand(Bernoulli(0.3), 20);
```

```
rand(Bernoulli(0.3), 20) = Bool[0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1]
```

これは 1 が生成される確率が $p=0.3$ で試行回数が $n=20$ のBernoulli試行の例であり, この場合には生成された 1 の個数の方が 0 の個数より少なくなっている.

乱数の出目が大幅に偏っていれば 1 の個数の方が多くなることもありえる. 1 が生成される確率が $p=0.3$ で試行回数が $n=20$ のBernoulli試行で生成された 1 の個数が 10 より大きくなる確率は約 1.7% になる.

In [6]:
```
ccdf(Binomial(20, 0.3), 10)
```

Out[6]:
```
0.017144816431258425
```

0以上1未満の乱数生成を繰り返して, 乱数の値が p 未満(以下でもよい)ならば 1 を生成し, p 以上(より大きいでもよい)のとき 0 を生成することにしてもBernoulli試行が得られる.

In [7]:
```
T = rand(20)
for i in eachindex(T) @printf "%.16f\n" T[i] end
```

```
0.7086835170539888
0.3129878077458778
0.4432299124211534
0.6380100590067396
0.7457553759007464
0.0042479988608554
0.4282692167397828
0.2591080281694401
0.6056891396222264
0.4718831513039459
0.2359750577106341
0.3658417371594386
0.4169159761488047
0.2234702295437706
0.2267949185649013
0.1054052960574101
0.5401955636162634
0.3620455865151407
0.9332297268604005
0.1009962782059489
```

以下の計算では T の成分が $p=0.3$ 未満のときに 1 に, それ以外のときに 0 に変換している. これによって, 一様乱数列からBernoulli試行が得られる.

In [8]:
```
X = [t < 0.3 ? 1 : 0 for t in T]
for i in eachindex(T) @printf "%.16f → %d\n" T[i] X[i] end
```

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

Bernoulli試行に限らず, 何らかの確率分布に従う乱数の生成の独立試行は, 0以上1未満の一様乱数の生成を経由して作ることができる. その意味でコンピュータでのプログラミングでおなじみの `rand()` 函数は基本的である.

毎回同確率で当たりが出るくじを繰り返し引くことは, 当たりのときに 1, 外れのときに0とすれば, Bernoulli試行モデルで記述できることの例になっている.

In [9]:
```
AH = [x == 1 ? "当たり" : "外れ" for x in X]
for i in eachindex(T) @printf "%.16f → %d → %s\n" T[i] X[i] AH[i] end
```

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

ビジネス分野では, 「広告を見てくれる」「リンクをクリックしてくれる」「商品を購入してくれる」などが「当たり」の例として扱われたりする.

現実の統計分析でBernoulli試行モデルを使う場合には, 「当たり」がよいことばかりとは限らない. たとえば, 「病気になる」「死ぬ」「事故が起きる」「機器が壊れる」などもBernoulli試行モデルで扱われる場合がある.

現実への応用において, Bernoulli試行モデルはある処置によって生じる「当たり」が出る確率の変化を推定するために使われることが多い.

例えば, 「薬を飲むと病気になる確率が減るか？」とか「ウェブサイトの構成を変えると商品を買ってもらえる確率が上がるか？」などが問題になる.

しかし, Bernoulli試行モデルにおける「すべての場合に同じ確率で独立に当たりまたは外れが生成される」という条件は相当に強い仮定になっており, 多くの場合には非現実的であり, 現実への応用では様々な工夫が必要になる.

統計分析のために用いられる **統計モデルは現実にはぴったり一致していない** と考えた方がよい. 以下ではこの点について誤解を招きかねない単純化を行う場合が多いので注意して欲しい. 多くの入門的教科書では **統計モデルは現実にはぴったり一致していない** という実践的に非常に重要なことが強調されていないことが多いので注意が必要である.

4 離散分布

4.1 離散分布と確率質量函数

以下, a_i 達は互いに異なると仮定する.

有限集合 $\{a_1, a_2, \ldots, a_r\}$ もしくは可算集合 $\{a_1, a_2, a_3, \ldots\}$ 上の **離散確率分布** (discrete probability distribution)もしくは **離散分布** (discrete distribution)とは, 0以上の有限個の実数達 $p_1, p_2, \ldots, p_r$ もしくは0以上の可算個の実数達 $p_1, p_2, p_3, \ldots$ で総和が 1 になるものであると定める:

$$p_i \geq 0, \quad \sum_i p_i = 1.$$

確率分布を単に分布と呼ぶことがよくある.

そのとき, a_i を p_i に対応させる函数 $P(a)$ をその離散分布の **確率質量函数** (probability mass function, pmf)と呼ぶ:

$$P(a_1) = p_1,\ P(a_2) = p_3,\ P(a_3) = p_3,\ \ldots$$

このとき,

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

離散分布を与えることは確率質量函数を与えることに等しい. 以下では主に確率質量函数を与えることによって離散分布を構成する.

直観的には各 $P(a_i) = p_i$ はこの離散分布に従う乱数において値 a_i が確率 $P(a_i) = p_i$ で生成されることを意味すると考える.

しかし, 実際の数学的取り扱いでは乱数の「ランダム性」には触れずに単なる「確率の値の表」として

x	a_1	a_2	a_3	$\cdots$
$P(x)$	$P(a_1)$	$P(a_2)$	$P(a_3)$	$\cdots$

を扱うことになる.

以下で使いまわすルーレットの例: 例えば, $1, 2, 3, 4$ の値が確率それぞれ $0.1, 0.2, 0.3, 0.4$ で出るルーレットは,

x	1	2	3	4
$P(x)$	0.1	0.2	0.3	0.4

という確率の値の表で定まる離散分布で自然にモデル化される.

4.2 離散確率変数

離散確率変数 (discrete random variable)とは, 確率質量函数 $P(x)$ が与えられた有限集合 $\{a_1, a_2, \ldots, a_r\}$ または可算集合 $\{a_1, a_2, a_3, \ldots\}$ 上の **実数値函数** のことであると定める. (後で, 有限集合や可算集合上の函数でなくても, 取り得る値全体の集合が有限集合や離散的な可算集合になるような確率変数も離散確率変数と呼ぶことになるが, 現時点ではそういう細かいことを気にする必要はない.)

実際には, 複素数値の確率変数やベクトルや行列に値を持つ確率変数も考えることがある. そのような場合には, **複素数値確率変数**, **ベクトル値確率変数**, **行列値確率変数** などと呼ぶことにする.

函数のことを「変数」と呼ぶことに違和感を覚える人がいるかもしれないが, 直観的には確率変数を「ランダムに値が決まる変数」とみなしたい場合も多いので「変数」という呼び方もそう悪くない.

以下では, 有限集合 $\{a_1, a_2, \ldots, a_r\}$ または可算集合 $\{a_1, a_2, a_3, \ldots\}$ は $\mathbb{R}$ の部分集合である場合について考える.

確率質量函数 $P(x)$ が与えられていると仮定する. このとき, $x = a_i$ に実数の $x = a_i$ 自身を対応させる函数として定義される確率変数を X と書き, **確率質量函数 P(x) を持つ離散確率変数** と呼ぶ:

$$X : x = a_i \mapsto x = a_i.$$

このとき,

$$X \sim P(x)$$

という書き方をすることがある. さらに, その確率質量函数で与えられる離散分布に Dist という名前が付けられているとき, X は **分布** Dist **に従う確率変数** であるともいい,

$$X \sim \mathrm{Dist}$$

という書き方をすることがある. 「分布〜に従う確率変数」は "random variable following distribution 〜" の翻訳になっている. $X \sim \mathrm{Dist}$ のような表記は確率分布による統計モデルをコンピュータで実装するために使われている確率プログラミング言語での採用されている記号法でもある.

確率変数は大文字で表すことが多く, 大文字で書かれた確率変数に対応する通常の変数を対応する小文字で表すことが多い. この対応は絶対ではないが, ある程度守っておくと, 記号を読み易くなる場合がある.

例えば, $1, 2, 3, 4$ の値が確率それぞれ $0.1, 0.2, 0.3, 0.4$ で出るルーレットのモデル化の離散分布において, 集合 $\{1, 2, 3, 4\}$ 上の任意の実数値函数は離散確率変数になる.

$x = 1, 2, 3, 4$ のそれぞれにその値自身 $x \in \mathbb{R}$ を対応させる函数が定める確率変数を X と書くことにする. この X は確率質量函数 $P(x) = 0.1x$ $(x = 1, 2, 3, 4)$ を持つ離散確率変数である.

直観的には, この確率変数 X の値はランダムに決まり, その値は確率 $0.1, 0.2, 0.3, 0.4$ でそれぞれ $1, 2, 3, 4$ になると考える. (これはあくまでも直観的な解釈であり, 確率変数は数学的には単なる函数に過ぎない.)

$x \in \{1, 2, 3, 4\}$ を実数 $f(x) \in \mathbb{R}$ に対応させる函数も離散確率変数になる. その確率変数を $f(X)$ と表す.

直観的には, 確率変数 $f(X)$ の値はランダムに決まり, その値は確率 $0.1, 0.2, 0.3, 0.4$ でそれぞれ $f(1), f(2), f(3), f(4)$ になると考える.

一般に確率変数 X の函数 $f(X)$ も確率変数になると考えてよい. (変数の値がランダムに決まるとき, その変数の函数の値もランダムに決まる. 値がランダムに決まると考えたくなるものを確率変数と呼ぶ習慣になっている.)

4.3 離散確率変数の期待値

確率変数の最も重要な性質はその期待値が定義されることである. **確率変数の定義を期待値が定義されている変数(のようなもの)のことだと思ってもよい.** それくらい期待値の概念は確率変数の理解において基本的である.

確率質量函数 $P(x)$ が定められた有限集合 $\{a_1, a_2, \ldots, a_r\}$ または可算集合 $\{a_1, a_2, a_3, \ldots\}$ 上の函数 $f(x)$ が定める離散確率変数を $f(X)$ と書くのであった. このとき, 離散確率変数 $f(X)$ の **期待値**(expectation value) $E[f(X)]$ が次のように定義される:

$$E[f(X)] = \sum_i f(a_i)P(a_i).$$

この和が無限和になる場合には絶対値の期待値 $E[|f(X)|] = \sum_i |f(a_i)|P(a_i)$ が有限の値に収束するとき(すなわち絶対収束するとき)にのみ, 期待値 $E[f(X)] = \sum_i f(a_i)P(a_i)$ が定義されていると考える. (これ以後, こういう細かい話は略す.)

X の従う分布の名前が Dist のとき $E[X]$ を **分布** Dist **の期待値** と呼ぶ.

確率変数の期待値を **平均** または **平均値** と呼ぶこともあるが, あとで定義する標本の平均値とは区別する必要がある. 混乱しないためには常に「確率変数の平均(期待値)」「標本の平均」のように「〜の平均」という言い方をすればよいだろう.

例えば, 上のルーレットの例では,

$$\begin{aligned} E[X] &= 1 \cdot 0.1 + 2 \cdot 0.2 + 3 \cdot 0.3 + 4 \cdot 0.4 = 3, \\ E[f(X)] &= f(1) \cdot 0.1 + f(2) \cdot 0.2 + f(3) \cdot 0.3 + f(4) \cdot 0.4 \end{aligned}$$

上のルーレットの例で, $1, 2, 3, 4$ の値が出たときに得られる賞金の額がそれぞれ10000円, 3000円, 1000円, 300円のとき, 賞金は次の確率変数 $g(X)$ でモデル化される:

$$g(1) = 10000, \quad g(2) = 3000, \quad g(3) = 1000, \quad g(4) = 300.$$

このとき, 賞金 $g(X)$ の期待値は次のように2020円だと計算される:

$$\begin{aligned} E[g(X)] &= g(1)P(1) + g(2)P(2) + g(3)P(3) + g(4)P(4) \\ &= 10000 \cdot 0.1 + 3000 \cdot 0.2 + 1000 \cdot 0.3 + 300 \cdot 0.4 \\ &= 1000 + 600 + 300 + 120 = 2020. \end{aligned}$$

4.4 離散確率変数の期待値を取る操作の基本性質

確率変数の期待値を取る操作(**期待値汎函数**) $f(X) \mapsto E[f(X)]$ は以下の性質を満たしている.

- 線形性: $E[\alpha f(X) + \beta g(X)] = \alpha E[f(X)] + \beta E[g(X)]$.
- 単調性: 任意の $x = a_i$ について $f(x) \leq g(x)$ が成立しているならば, $E[f(X)] \leq E[g(X)]$.
- 規格化条件: $E[\alpha] = \alpha$. (定数 α の期待値は α. 特に $E[1] = 1$).

証明: 確率変数 $f(X), g(X)$ と定数 α, β について,

$$
\begin{aligned}
E[\alpha f(X) + \beta g(X)] &= \sum_i (\alpha f(a_i) + \beta g(a_i)) P(a_i) \\
&= \alpha \sum_i f(a_i) P(a_i) + \beta \sum_i g(a_i) P(a_i) = \alpha E[f(X)] + \beta E[g(X)]
\end{aligned}
$$

任意の $x = a_i$ について $f(x) \leq g(x)$ が成立していると仮定すると, $P(a_i) \geq 0$ なので,

$$
E[f(X)] = \sum_i f(a_i) P(a_i) \leq \sum_i g(a_i) P(a_i) = E[g(X)].
$$

定数 α を定数確率変数とみなすと, $\sum_i P(a_i) = 1$ なので,

$$
E[\alpha] = \sum_i \alpha P(a_i) = \alpha \sum_i P(a_i) = \alpha.
$$

証明終.

重要: 確率変数に関する多くの計算は上で説明した期待値を取る操作の基本性質のみを使って実行可能なことが多い. 統計学がらみの理論的計算に習熟したいならば, $E[\ ]$ を使った計算の習熟を最初の目標にして勉強するとよいだろう. 以下で説明するように確率変数に関する確率も $E[\ ]$ を使って計算可能である.

4.5 確率変数の一般的な定義

ここで離散とは限らない一般の確率変数を少々大雑把なスタイルで定義しておこう.

変数 X が確率変数であるとは, X の函数 $f(X)$ の期待値 $E[f(X)]$ を取る操作が定義されていて, 期待値を取る操作が前節の基本性質を満たしていることだと定める.

ただし, $E[|f(X)|] = \infty$ になる場合には $f(X)$ の期待値はうまく定義されないと考える.

確率変数には他にも様々な解釈(定義)の仕方があるが, 統計学で必要な様々な理論的な計算をするときには確率変数とは「その函数の期待値が定義されている変数」のことであると了解しておくと, 様々な議論を楽に理解できることが多い.

4.6 離散確率変数に関する確率

$f(X)$ は確率質量函数 $P(x)$ が定められた有限集合 $\{a_1, a_2, \ldots, a_r\}$ または可算集合 $\{a_1, a_2, a_3, \ldots\}$ 上の確率変数であるとする.

このとき, 例えば, $c \leq f(X) \leq d$ となる確率 $P(c \leq f(X) \leq d)$ を次のように定める:

$$
P(c \leq f(X) \leq d) = \sum_{c \leq f(a_i) \leq d} P(a_i).
$$

確率変数に関する $c \leq f(X) \leq d$ 以外の条件(例えば $f(X) \geq c$ や $f(X) < d$ など)の確率も同様に定める.

このような確率は, 確率変数 $1_{c \leq f(X) \leq d}(X)$ を, 条件 $c \leq f(x) \leq d$ が成立するときに 1 になり, 成立していないとき 0 になる函数として,

$$
1_{c \leq f(X) \leq d}(x) = \begin{cases} 1 & (c \leq f(x) \leq d) \\ 0 & (\text{otherwise}) \end{cases}
$$

と定めると,

$$
P(c \leq f(X) \leq d) = \sum_{c \leq f(a_i) \leq d} P(a_i) = \sum_i 1_{c \leq f(X) \leq d}(a_i) P(a_i) = E[1_{c \leq f(X) \leq d}(X)]
$$

と期待値で書けるので, 期待値の概念は確率の概念を含んでいるといえる.

例えば, 上のルーレットの例において, 賞金が500円以上5000円以下になる確率は $P(500 \leq g(X) \leq 5000)$ と書け,

$$
P(500 \leq g(X) \leq 5000) = 0.2 + 0.3 = 0.5
$$

となる. さらに,

$$
1_{500 \leq g(x) \leq 5000}(x) = \begin{cases} 1 & (x = 2, 3) \\ 0 & (\text{otherwise}) \end{cases}
$$

となるので,

$$
E[1_{500 \leq g(X) \leq 5000}(X)] = \sum_{i=1}^{4} 1_{500 \leq g(X) \leq 5000}(i) P(i) = 0 \cdot 0.1 + 1 \cdot 0.2 + 1 \cdot 0.3 + 0 \cdot 0.4 = 0.5
$$

となって, 上の結果と一致する.

4.7 離散確率変数の分散と標準偏差

離散確率変数 $Y = f(X)$ の期待値を

$$\mu_Y = E[Y] = E[f(X)] = \sum_i f(a_i)P(a_i)$$

と書くとき, $Y = f(X)$ の **分散** (variance) $\mathrm{var}(Y) = \mathrm{var}(f(X))$ が次のように定義される:

$$\sigma_Y^2 = \mathrm{var}(Y) = E\left[(Y - \mu_Y)^2\right] = \sum_i (f(a_i) - \mu_Y)^2\, P(a_i).$$

分散の平方根を **標準偏差** (standard deviation) と呼ぶ. 確率変数 Y の標準偏差を次のように表すことにする:

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

$a > 0$ のとき, Y を a 倍するとその分散は a^2 倍になるが, 標準偏差ならば Y と同じく a 倍になる. こういう性質を持っているので Y と直接比較する場合には Y の分散ではなく標準偏差を使った方が便利である.

X の従う分布の名前が Dist のとき $\mathrm{var}(X)$ を **分布** Dist **の分散** と呼ぶ. (標準偏差についも同様とする.)

確率変数の分散や標準偏差は確率変数が従う分布の **広がりの大きさ**, **確率的揺らぎの大きさ** の指標の1つになっている. そのような指標の中で **分散と標準偏差は期待値と相性がよい**.

注意: 確率変数の期待値(平均)以外の分布の中心を表す値として, 中央値の確率変数版を定義することもでき, 分布の広がりの大きさの指標として, 四分位数の確率変数版を定義することもできる. 期待値(平均)と分散(と標準偏差)の組み合わせと中央値と四分位数の組み合わせは目的と状況に合わせて使い分けるとよいだろう.

確率変数の分散については

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

という公式もよく使われる. この公式は以下のようにして証明される:

$$\begin{aligned}\mathrm{var}(Y) &= E\left[(Y - \mu_Y)^2\right] = E\left[Y^2 - 2\mu_Y Y + \mu_Y^2\right] = E[Y^2] - 2\mu_Y E[Y] + \mu_Y^2 \\ &= E[Y^2] - 2\mu_Y\mu_Y + \mu_Y^2 = E[Y^2] - \mu_Y^2 = E[Y^2] - E[Y]^2.\end{aligned}$$

この証明では期待値を取る操作の線形性と規格化条件しか使われていないので, この結果は非常に一般的に成立しており, 後で説明する連続確率変数の分散についても成立している.

注意: コンピュータによる浮動小数点演算で分散を実装するときには, $\mathrm{var}(Y) = E[Y^2] - E[Y]^2$ という公式を使わずに, 分散の定義式 $E\left[(Y - \mu_Y)^2\right]$ を使って実装した方が良い場合がある. なぜならば, 浮動小数点演算では大きな数値の差が小さな数値になるとき桁落ちが生じるからである. 同じ程度の大きさの数値の差の計算になり易い $Y - \mu_Y$ と違って, $E[Y^2] - E[Y]^2$ は大きな数のあいだの差で計算結果が小さな数になる可能性が高いので注意が必要である.

4.8 問題: 期待値と分散の特徴付け

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

解答例: 期待値を取る操作の基本性質のみを使って示せる. $\mu = E[X]$ とおく. このとき,

$$(X - t)^2 = ((X - \mu) - (t - \mu))^2 = (X - \mu)^2 - 2(t - \mu)(X - \mu) + (t - \mu)^2$$

なので, これに $E[\ ]$ を作用させ, $E[\ ]$ の線形性(足し算と引き算と定数倍を $E[\ ]$ の外に出せること)と規格化条件(定数の期待値はその定数そのものの値になること)および $E[X] = \mu$, $E[(X - \mu)^2] = \mathrm{var}(X)$ を使うと,

$$E[(X - t)^2] = E[(X - \mu)^2] - 2(t - \mu)(E[X] - \mu) + (t - \mu)^2 = \mathrm{var}(X) + (t - \mu)^2.$$

これより, $f(t)$ は $t = \mu = E[X]$ で最小値 $\mathrm{var}(X)$ になることがわかる.

解答終

注意: 上の証明では期待値を取る操作の基本性質しか使っておらず, X が有限離散分布に従う確率変数であることを使っていない. 上の問題の結果は X が有限離散分布に限らないもっと一般の分布に従っている場合にも成立している.

注意: $g(t) = E[|X - t|]$ を最小化する t として確率変数 X の中央値を特徴付けることができる. ただし, 確率変数の中央値は一般には一意(ただ一つ)には決まらない. この結果は数学的にそれなりに非自明なので, 数学に自身がある人は確率変数の中央値を自分で適切に定義して, 中央値であることと $g(t)$ を最小化することが同値であることを証明してみよ. (おそらく初学者にとってこれは難問である.)

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

注意: 標本(サンプル, データ)の平均や分散や中央値と確率変数(もしくはそれが従う分布)の平均や分散や中央値の概念を統計学では厳密に区別する必要がある. 混乱しないように注意して欲しい. 混乱を防ぐためには単に「平均」「分散」などと言うことを少なめにして, できるだけ「確率変数の平均や分散」「標本の平均や分散」のように「○○の平均や分散」と言うようにすればよいと思う.

4.9 問題: 平均二乗誤差のバイアス・バリアンス分解

X は確率変数であるとし, その期待値 $\mu = E[X]$ と分散 $E[(X-\mu)^2]$ がうまく定義されているとする. 定数 θ に対して次が成立することを示せ:

$$E[(X-\theta)^2] = (E[X]-\theta)^2 + E[(X-\mu)^2].$$

注意: 確率変数 X の実現値を定数 θ の推定値とみなす状況において, 左辺の $E[(X-\theta)^2]$ を推定の **平均二乗誤差** と呼び, $E[X]-\theta$ を推定の **バイアス**(bias)と呼び, $E[(X-\mu)^2]$ を **バリアンス** (variance, 分散)と呼び, この公式を **平均二乗誤差のバイアス・バリアンス分解** と呼ぶ.

解答例: $(X-\theta)^2$ の括弧の内側に $0 = -\mu + E[X]$ を足すと,

$$\begin{aligned}(X-\theta)^2 &= ((X-\mu)-(E[X]-\theta))^2 \\ &= (X-\mu)^2 + 2(E[X]-\theta)(X-\mu) + (E[X]+\theta)^2.\end{aligned}$$

$E[X]-\theta$ と $(E[X]+\theta)^2$ が確率変数ではなく, 定数であることに注意しながら, 期待値を取る操作の線形性と規格化条件を使うと,

$$\begin{aligned}(X-\theta)^2 &= E[(X-\mu)^2] + 2(E[X]-\theta)(E[X]-\mu) + (E[X]-\theta)^2 \\ &= E[(X-\mu)^2] + (E[X]-\theta)^2.\end{aligned}$$

たし算の順序を交換すれば欲しい公式が得られる.

解答終

注意: 上の解答例で用いた式の変形は統計学の文脈で数え切れないくらい登場する. 例えば, 前節の問題でもこの方法を用いた.

4.10 カテゴリカル分布の定義

番号の集合 $\{1, 2, \dots, r\}$ 上の確率質量函数

$$P(i) = p_i \quad (p_1, \dots, p_r \geq 0,\ p_1 + \cdots + p_r = 1)$$

が定める有限離散分布を **カテゴリカル分布** と呼び, 次のように表す:

$$\mathrm{Categorical}(p_1, p_2, \dots, p_r).$$

確率の表は次のようになる:

x	1	2	$\cdots$	r
$P(x)$	p_1	p_2	$\cdots$	p_r

多項分布の確率質量函数は次のようにも書ける:

$$P(i) = p_1^{\delta_{i1}} \cdots p_r^{\delta_{ir}}.$$

ここで δ_{ij} は $i = j$ のとき 1 になり, それ以外のとき 0 になるKroneckerのデルタである.

`categ = Categorical(0.1, 0.2, 0.3, 0.4)` のとき `rand(categ)` は値が $1, 2, 3, 4$ にそれぞれ確率 $0.1, 0.2, 0.3, 0.4$ でなる確率変数だとみなされる. (値がランダムに決まる量も確率変数と呼ぶ習慣がある.) `categ` は上の方で繰り返し使ったルーレットの分布に等しい.

In [10]:
```
categ = Categorical(0.1, 0.2, 0.3, 0.4)
```

Out[10]:
```
Categorical{Float64, Vector{Float64}}(support=Base.OneTo(4), p=[0.1, 0.2, 0.3, 0.4])
```

In [11]:
```
@show rand(categ, 20);
```

```
rand(categ, 20) = [4, 4, 4, 3, 4, 2, 3, 3, 4, 3, 2, 2, 1, 4, 2, 4, 4, 4, 4, 4]
```

In [12]:
```
bar(support(categ), a -> pdf(categ, a); label="", title="Categorical(0.1, 0.2, 0.3, 0.4)", alp
```

Out[12]:

4.11 問題: ルーレットの例の賞金の期待値と分散

ルーレットの例の分布は Categorical(0.1, 0.2, 0.3, 0.4) に等しい. この分布に従う確率変数を X と書く. 賞金の確率変数 $g(X)$ は

$$g(1) = 10000, \quad g(2) = 3000, \quad g(3) = 1000, \quad g(4) = 300,$$

で定義される. X と $g(X)$ の期待値と分散と標準偏差を求めよ. 整数にならない標準偏差は小数点以下第1桁目まで求めよ.

解答例: X の期待値, 分散, 標準偏差はそれぞれ

$$\mu_X = 1, \quad \sigma_X^2 = 1, \quad \sigma_X = 1.$$

$g(X)$ の期待値, 分散, 標準偏差はそれぞれ

$$\mu_{g(X)} = 2020, \quad \sigma_{g(X)}^2 = 8055600, \quad \sigma_{g(X)} \approx 2838.2.$$

詳しい計算法については以下を見よ. 以下と同じ数値が得られれば正解である. 浮動小数点計算では小数点以下3桁まで正しければ正解であるとする(四捨五入すること).

解答終

X の期待値と分散と標準偏差:

In [13]:
```
μ_X = 1 * 0.1 + 2 * 0.2 + 3 * 0.3 + 4 * 0.4
@show μ_X;
```

```
μ_X = 3.0
```

X の分散と標準偏差:

In [14]:
```
σ²_X = (1 - 3)^2 * 0.1 + (2 - 3)^2 * 0.2 + (3 - 3)^2 * 0.3 + (4 - 3)^2 * 0.4
σ_X = √σ²_X
@show σ²_X σ_X;
```

```
σ²_X = 1.0
σ_X = 1.0
```

$g(X)$ の期待値:

In [15]:
```
μ_g = 10000 * 0.1 + 3000 * 0.2 + 1000 * 0.3 + 300 * 0.4 ▷ Int
@show μ_g;
```

```
μ_g = 2020
```

$g(X)$ の分散と標準偏差:

In [16]:
```
σ²_g = (10000 - 2020)^2 * 0.1 + (3000 - 2020)^2 * 0.2 + (1000 - 2020)^2 * 0.3 + (300 - 2020)^2
σ_g = √σ²_g
@show σ²_g round(σ_g; digits=1);
```

```
σ²_g = 8055600
round(σ_g; digits = 1) = 2838.2
```

以下はJulia言語のより高級な機能を使った計算法である.

In [17]:
```
categ = Categorical(1//10, 2//10, 3//10, 4//10)
μ_X = mean(categ) ▷ Int
σ²_X = var(categ) ▷ Int
σ_X = std(categ) ▷ Int
@show μ_X σ²_X σ_X;
```

```
μ_X = 3
σ²_X = 1
σ_X = 1
```

期待値を取る操作は分布 dist と確率変数 f に期待値を対応させる函数 $E(f, \mathrm{dist})$ として定義可能である.

In [18]:
```
E(f, dist::DiscreteUnivariateDistribution; imax = 10^6) =
    sum(x -> f(x) * pdf(dist, x), support(dist)[1:min(end, imax)])
```

Out[18]:
```
E (generic function with 1 method)
```

$E(f, \mathrm{dist})$ を使って X の期待値と分散を計算すると以下のようになる.

In [19]:
```
categ = Categorical(1//10, 2//10, 3//10, 4//10)
μ_X = E(x -> x, categ) ▷ Int
σ²_X = E(x -> (x - μ_X)^2, categ) ▷ Int
σ_X = √σ²_X ▷ Int
@show μ_X σ²_X σ_X;
```

```
μ_X = 3
σ²_X = 1
σ_X = 1
```

In [20]:
```
g(x) = get((10000, 3000, 1000, 300), x, 0) # x = 1,2,3,4 のそれぞれを 10000, 3000, 1000, 300
μ_g = E(g, categ) ▷ Int
σ²_g = E(x -> (g(x) - μ_g)^2, categ) ▷ Int
σ_g = √σ²_g
@show μ_g σ²_g round(σ_g; digits=1);
```

```
μ_g = 2020
σ²_g = 8055600
round(σ_g; digits = 1) = 2838.2
```

4.12 Bernoulli分布の定義

$0 \leq p \leq 1$ と仮定する. 値が $1, 0$ になる確率をそれぞれ $p, 1-p$ とすることによって定まる離散分布を成功確率 p の **Bernoulli 分布** (ベルヌーイ分布, ベルヌイ分布)と呼び, 次のように表す:

$$\mathrm{Bernoulli}(p).$$

この分布の確率質量函数は次のようになる:

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

4.13 問題: Bernoulli分布の期待値と分散

Bernoulli分布 $\mathrm{Bernoulli}(p)$ に従う確率変数 X の期待値と分散を求めよ. さらに分散の最大値とそれを与える p を求めよ.

解答例: X の期待値 $E[X]$ と分散 $\mathrm{var}(X)$ は以下のように計算される:

$$
\begin{aligned}
E[X] &= \sum_{x=1,0} xp(x) = 1p + 0(1-p) = p, \\
E[X^2] &= \sum_{x=1,0} x^2 p(x) = 1^2 p + 0^2(1-p) = p, \\
\mathrm{var}(X) &= E[X^2] - E[X]^2 = p - p^2 = p(1-p).
\end{aligned}
$$

$\mathrm{var}(X) = -(p-1/2)^2 + 1/4$ なので分散 $\mathrm{var}(X)$ は $p = 1/2$ のとき最大値 $1/4$ になる.

解答終

In [21]:
```
p = 0.3
bern = Bernoulli(p)
```

Out[21]:
```
Bernoulli{Float64}(p=0.3)
```

In [22]:
```
@show mean(bern) var(bern);
```

```
mean(bern) = 0.3
var(bern) = 0.21
```

In [23]:
```
@show rand(bern, 20);
```

```
rand(bern, 20) = Bool[0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 1]
```

In [24]:
```
bar(0:1, x -> pdf(bern, 1 - x); label="", title="Bernoulli(0.3)", alpha=0.5)
plot!(; xtick=(0:1, 1 .- (0:1)), ytick=0:0.1:1)
```

Out[24]:

4.14 オッズの定義と例

成功確率 p に対して, 失敗確率分の成功確率

$$
\frac{p}{1-p} = \frac{1}{1-p} - 1
$$

を **オッズ** (odds)と呼ぶ. オッズは $0 \le p < 1$ について単調増加函数になり, $p = 1$ のときオッズは ∞ になると考えられる.

オッズは賭け事の世界では普通によく使われる用語である. オッズは勝利確率 p の単調増加函数になっているので, 勝利確率とは別の勝ち易さの指標になっている.

例えば, 勝利確率が p のギャンブルで勝利したときに賭けた金額の何倍の賞金をもらえることにすると公平になるだろうか? 勝利確率が p の賭け事に1万円賭けたとき, 勝利したら賭けた1万円はそのままで賞金として k 万円もらえ, 敗北したら賭けた1万円が全額没収されるとする. そのときの期待値は $pk - (1-p)$ 万円なので, 期待値が 0 の公平なギャンブルにするためには k をオッズの逆数にしなければいけない. オッズの低い勝ち難いギャンブルではその分だけ勝利したときに沢山の賞金を獲得できないと損になる.

例えば, 勝利確率が 1/2 のギャンブルではオッズは $(1/2)/(1/2) = 1$ になる.

例えば, 勝利確率が 1/3 のギャンブルではオッズは $(1/3)/(2/3) = 1/2$ になり, 勝利したときに賭けた金額の2倍の賞金が得られるなら公平なギャンブルになる. 実際, そのとき, 1 万円を賭けると, 1/3 の確率で 2 万円もらえ, 2/3 の確率で 1 万円を失うので, 確かに期待値は 0 になる. (これはBernoulli分布モデルにおける確率変数の期待値の例にもなっている.)

4.15 オッズとロジットおよびロジスティック変換

確率 p についてそのオッズを $u = p/(1-p) = 1/(1-p) - 1$ と書くと, 逆に確率はオッズで $p = u/(1+u) = 1 - 1/(1+u)$ と書ける. これによって, $0 < p < 1$ と $0 < u < \infty$ のあいだの一対一の対応が得られる. さらに $x = \log(u)$, $u = \exp(x)$ とおくと, $0 < u < \infty$ と $-\infty < x < \infty$ のあいだの一対一に対応も得られる. p と x の関係を直接書き下すと,
$p = \exp(x)/(1+\exp(x)) = 1/(1+\exp(-x))$, $x = \log(p/(1-p))$ となる. 以上によって,

$$p = \frac{u}{1+u} = \frac{e^x}{1+e^x} = \frac{1}{1+e^{-x}}, \quad u = \frac{p}{1-p} = e^x, \quad x = \log(u) = \log\frac{p}{1-p}$$

によって,

$$0 < p < 1, \quad 0 < u < \infty, \quad -\infty < x < \infty$$

のあいだの一対一対応が得られたことになる. これらは賭け事とは無関係に非常によく使われる変数変換であり, 賭け事由来のオッズの概念はこのような形でも数学的に非常に役に立つことが知られている.

以下の変数変換はそれぞれ **ロジット変換** (logit変換), **ロジスティック変換** (logistic変換)と呼ばれている:

$$x = \mathrm{logit}(p) = \log\frac{p}{1-p}, \quad p = \mathrm{logistic}(x) = \frac{1}{1+e^{-x}}.$$

確率を意味するパラメータの p を $-\infty$ から ∞ のあいだを自由に動ける変数 x に変換してから様々な計算を行った方が都合が良い場合があるので, これらの変換はよく使われる. (既存の統計ソフトが行ってくれる計算においてユーザーから見えない所でこのような変換が使われている場合もある.)

4.16 Bernoulli試行の確率分布

Bernoulli分布 $\mathrm{Bernoulli}(p)$ における $x = 1, 0$ の確率は

$$P(x) = p^x(1-p)^{1-x}$$

と書ける. 実際, $P(1) = p^1(1-p)^0 = p$ となり, $P(0) = p^x(1-p)^{1-x} = p^0(1-p)^1 = 1-p$ となる. これをBernoulli試行の場合に拡張しよう.

試行回数 n のBernoulli試行の結果は長さ n の 1 と 0 からなる数列 $(x_1, \dots, x_n)$ になる. そのような列全体の個数は, 各々の x_k が2通りでそれらが n 個あるから, 2^n 個になる. ($p = 0.5$ の場合を除いて, すべての場合の確率が等しくなるわけではない.)

Bernoulli試行が毎回確率 p で 1 が出て確率 $1-p$ で 0 が出る独立な試行の繰り返しなので, 与えられた $x_1, \dots, x_n = 1, 0$ について, n 回のBernoulli試行の結果が $(x_1, \dots, x_n)$ になる確率 $P(x_1, \dots, x_n)$ は, Bernoulli分布における確率

$$P(x_i) = p^{x_i}(1-p)^{1-x_i}$$

の積

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

になると考えられる. (数学的にはこれによってBernoulli試行の確率分布を定める.) これは, $x_1, \dots, x_n$ の中の 1 (=「当たり」)の個数を k と書くと,

$$p^k(1-p)^{n-k}$$

とシンプルに書来直される.

このようにして定まる集合 $\{\, (x_1, \dots, x_n) \mid x_1, \dots, x_n = 1, 0 \,\}$ 上の離散分布を試行回数 n, 成功確率 p の **Bernoulli試行の確率分布** と呼ぶことにする.

4.17 問題: n=3, p=0.3 の場合のBernoulli試行の確率分布

$n = 3$, $p = 0.3$ のBernoulli試行の確率分布における確率質量函数を具体的に書き下し, 確率の総和がぴったり 1 になることを確認せよ.

解答例: $n = 3$, $p = 0.3$ のBernoulli試行の確率分布を考える. (x_1, x_2, x_3) を単に並べて $x_1x_2x_3$ と書くと,

$$\{\, (x_1, x_2, x_3) \mid x_1, x_2, x_3 = 1, 0 \,\} = \{111, 011, 101, 110, 100, 010, 001, 000\}.$$

確率質量函数の全体は以下のように記述される:

$$
\begin{aligned}
P(111) &= 0.3^3 = 0.027, \\
P(011) &= P(101) = P(110) = 0.3^2 \cdot 0.7 = 0.063, \\
P(100) &= P(010) = P(001) = 0.3 \cdot 0.7^2 = 0.147, \\
P(000) &= 0.7^3 = 0.343.
\end{aligned}
$$

これら8通りの確率の和は

$$0.027 + 3 \cdot 0.063 + 3 \cdot 0.147 + 0.343 = 0.027 + 0.189 + 0.441 + 0.343 = 1$$

と確かにぴったり 1 になっている.

解答終

In [25]:
```
27//1000 + 3*63//1000 + 3*147//1000 + 343//1000 ▷ Int
```

Out[25]: 1

4.18 Bernoulli試行から応用上基本的かつ重要な確率分布が大量に得られること

Bernoulli試行から以下の確率分布が得られる:

- Bernoulli分布 (これは $n = 1$ のBernoulli試行の確率分布と $n = 1$ の二項分布に等しい)
- 二項分布 (n 回のBernoulli試行で出た 1 の個数の分布)
- 幾何分布 (これは $k = 1$ の負の二項分布に等しい)
- 負の二項分布 (Bernoulli試行を k 回 1 が出るまで続けたときに出た 0 の個数の分布)

これらの連続極限として,

- 正規分布 (二項分布の中心極限定理)
- Poisson分布 (二項分布の時間連続極限)
- ガンマ分布 (負の二項分布の時間連続極限)

も得られる. 正規分布, ガンマ分布, Poisson分布は極めて基本的な確率分布であり, 非常に頻繁に使われる.

Bernoulli試行について十分に理解すれば, 他の重要かつ基本的な確率分布についても楽に理解できるようになるだろう!

4.19 二項定理

n 個から k 個を選ぶ組み合わせの数を

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}$$

と書き, **二項係数** と呼ぶ.

互いに可換な文字 x, y の和の n 乗の展開したときの $x^k y^{n-k}$ の係数を a_k と書く:

$$(x+y)^n = \sum_{k=0}^{n} a_k x^k y^{n-k}.$$

このとき, a_k は掛けられてる n 個の $x + y$ の中から x をピックアップする k 個を選ぶ組み合わせの数に等しいので, $a_k = \binom{n}{k}$ が成立する. ゆえに

$$(x+y) = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

特に $y = 1$ のとき,

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k.$$

これらの結果を **二項定理** と呼ぶ.

注意: 高校の教科書のように二項係数を ${}_nC_k$ と書いてもよいが, このノートではこの書き方をしない. 後で二項係数 $\binom{n}{k}$ の n は整数でない場合にも拡張されることになる.

4.20 二項分布の定義

二項分布 は試行回数 n, 成功確率 p のBernoulli試行で生じる 1 と 0 からなる長さ n の列の中の 1 の個数の分布として得られる.

試行回数 n, 成功確率 p のBernoulli試行において 1 と 0 からなる列 $(x_1, x_2, \dots, x_n)$ が生じる確率は

$$p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)}$$

と書けるのであった. 1 と 0 からなる列 $(x_1, x_2, \dots, x_n)$ でそれが含む 1 の個数 $x_1 + \cdots + x_n$ が k になるものの個数は, n 個の x_i 達から k 個の 1 になるものを選ぶ組み合わせの数

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

になるので, 試行回数 n, 成功確率 p のBernoulli試行において, 1 の個数 $x_1 + \cdots + x_n$ が k になる確率は

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

になる. この確率質量函数で定義される集合 $\{0, 1, \dots, n\}$ 上の有限離散分布を **二項分布** と呼び, 次のように表す:

$$\mathrm{Binomial}(n, p).$$

二項分布における確率の総和が 1 になることは二項定理を使えばただちに確認できる:

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

4.21 問題: 二項分布の期待値と分散

K は二項分布に従う確率変数であるとする:

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

K の期待値と分散を求めよ.

解答例: 二項定理を使う. K の期待値は次のようにして得られる:

$$\begin{aligned} E[K] &= \sum_{k=0}^{n} kP(k) = \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^{n} n \frac{(n-1)!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = np(p + (1-p))^{n-1} = np, \end{aligned}$$

上の計算の5番目の統合で二項定理を使っている. この計算法と同様のアイデアで以下のように計算することができる:

$$\begin{aligned} E[K(K-1)] &= \sum_{k=0}^{n} k(k-1)P(k) = \sum_{k=2}^{n} k(k-1) \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^{n} n(n-1) \frac{(n-2)!}{(k-2)!(n-k)!} p^k (1-p)^{n-k} = n(n-1)p^2 \sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} (1-p)^{n-k} \\ &= n(n-1)p^2 (p + (1-p))^{n-2} = n(n-1)p^2, \end{aligned}$$

この計算の5番目の統合でも上と同様に二項定理を使っている. $\mathrm{var}(K) = E[K^2] - E[K]^2$ を使って K の分散を計算しよう:

$$\begin{aligned} &E[K(K-1)] = E[K^2] - E[K], \\ &E[K^2] = E[K(K-1)] + E[K] = n(n-1)p^2 + np = n^2p^2 - np^2 + np = np(1-p) + n^2p^2, \\ &\mathrm{var}(K) = E[K^2] - E[K]^2 = np(1-p) + n^2p^2 - (np)^2 = np(1-p). \end{aligned}$$

解答終

注意: 二項分布 $\mathrm{Binomial}(n, p)$ に従う確率変数の期待値 np と分散 $np(1-p)$ は $n = 1$ の場合のBernoulli分布 $\mathrm{Bernoulli}(p)$ の期待値と p と分散 $p(1-p)$ のちょうど n 倍になっている.

期待値についてこの事実は次のように説明される. Bernoulli試行における 1 と 0 の列 $(x_1, \dots, x_n)$ に x_i を対応させる函数によって定まる確率変数を X_i と書くと, 1 と 0 からなる列の中の 1 の個数を意味する確率変数は $K = X_1 + \cdots + X_n$ になる. K の期待値が二項分布の期待値になる. 各々の X_i は成功確率 p のBernoulii分布に従うので, その期待値は $E[X_i] = p$ になり, 期待値を取る操作の線形性(加法性)より,

$$E[K] = E[X_1] + \cdots + E[X_n] = p + \cdots + p = np.$$

分散についても

$$\mathrm{var}(K) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n) = p(1-p) + \cdots + p(1-p) = np(1-p)$$

が成立することは偶然ではなく, 必然的にこうなる理由がある. この点については後で確率変数の独立性について説明してから詳しく説明する. 答えを先走って言うと, 独立な確率変数の和の分散はそれぞれの確率変数の分散の和に等しくなる.

二項分布における確率質量函数のグラフは以下のようになる.

In [26]:

```
bin = Binomial(16, 0.3)
x = support(bin)
bar(x, x -> pdf(bin, x); label="", title="Binomial(20, 0.3)", alpha=0.5)
plot!(; xtick=x)
```

Out[26]:

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

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

In [27]:

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

Out[27]:

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

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

In [28]:

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

Out[28]:

In [29]:

```
λ, n = 3, 1000
p = λ/n
bin = Binomial(n, p)
μ, σ = mean(bin), std(bin)
x = 0:5λ
groupedbar(x, [pdf.(bin, x) pdf.(Poisson(n*p), x)];
    label=["Binomial(300, 0.01)" "Poisson($(n*p))"], alpha=0.5)
plot!(; xtick=x, xlim=(-2, 16.5))
plot!(Normal(μ, σ); label="normal approximation", lw=1, c=:red, ls=:dash)
```

Out[29]:

4.22 問題: 当たりが30%の確率で出るくじを100回引いた場合

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

(1) 当たりの回数 K の期待値と分散と標準偏差を求めよ.

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

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

有効桁3桁で求めよ.

解答例: $p = 0.3$, $n = 100$ とおく.

(1) K の期待値, 分散, 標準偏差:

$$E[K] = np = 30, \quad \mathrm{var}(K) = np(1-p) = 21, \quad \mathrm{std}(K) = \sqrt{\mathrm{var}(K)} \approx 4.58.$$

(2) $K \leq 25$ となる確率:

$$P(K \leq 25) = \sum_{k=0}^{25} \binom{n}{k} p^k (1-p)^{n-k} \approx 0.163.$$

$K \leq 30$ となる確率:

$$P(K \leq 20) = \sum_{k=0}^{20} \binom{n}{k} p^k (1-p)^{n-k} \approx 0.0165.$$

(2) $K \geq 35$ となる確率:

$$P(K \geq 35) = \sum_{k=35}^{100} \binom{n}{k} p^k (1-p)^{n-k} \approx 0.163.$$

$K \geq 40$ となる確率:

$$P(K \geq 40) = \sum_{k=40}^{100} \binom{n}{k} p^k (1-p)^{n-k} \approx 0.0210.$$

解答終

注意: WolframAlpha (https://www.wolframalpha.com/)では確率 $P(K \leq 25)$ を

- sum_{k=0}^{25} binom(100, k) 0.3^k (1 - 0.3)^{100-k} (https://www.wolframalpha.com/input?i=sum_%7Bk%3D0%7D%5E%7B25%7D+binom%28100%2C+k%29+0.3%5Ek+%281+-+0.3%29%5E%7B100-k%7D)

と入力して求めることができる. 他の場合も試してみよ. (もちろんJulia言語を導入して使ってもよい.)

In [30]:

```
logP(n, p, k) = logabsbinomial(n, k)[1] + k*log(p) + (n-k)*log(1-p)

p = 0.3
n = 100
@show n*p
@show n*p*(1-p)
@show √(n*p*(1-p))
for Kmax in (25, 20)
    @eval @show exp(logsumexp(logP(n, p, k) for k in 0:$Kmax))
end
for Kmin in (35, 40)
    @eval @show exp(logsumexp(logP(n, p, k) for k in $Kmin:n))
end
```

```
n * p = 30.0
n * p * (1 - p) = 21.0
√(n * p * (1 - p)) = 4.58257569495584
exp(logsumexp((logP(n, p, k) for k = 0:25))) = 0.16313010446635048
exp(logsumexp((logP(n, p, k) for k = 0:20))) = 0.016462853241869316
exp(logsumexp((logP(n, p, k) for k = 35:n))) = 0.16285828837178473
exp(logsumexp((logP(n, p, k) for k = 40:n))) = 0.020988576003924307
```

In [31]:

```
p = 0.3
n = 100
@eval @show bin = Binomial($n, $p)
@show mean(bin)
@show var(bin)
@show std(bin)
for Kmax in (25, 20)
    @eval @show cdf(bin, $Kmax)
end
for Kmin in (35, 40)
    @eval @show ccdf(bin, $Kmin - 1)
end
```

```
bin = Binomial(100, 0.3) = Binomial{Float64}(n=100, p=0.3)
mean(bin) = 30.0
var(bin) = 21.0
std(bin) = 4.58257569495584
cdf(bin, 25) = 0.16313010446635237
cdf(bin, 20) = 0.0164628532418695
ccdf(bin, 35 - 1) = 0.16285828837178704
ccdf(bin, 40 - 1) = 0.020988576003924644
```

4.23 幾何分布の定義

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

確率 p で 1 が確率 $1 - p$ で 0 が生じるようなBernoulli試行を続けたとき, ちょうど $m + 1$ 回目で 1 が初めて出る確率は, m 回 0 が続いて $m + 1$ 回目に 1 が出る確率

$$(1 - p)^m \times p$$

に等しいので,

$$P(m) = p(1 - p)^m$$

になる. この確率質量函数で定義される可算集合 $\{0, 1, 2, \ldots\}$ 上の無限離散分布を **幾何分布** (geometric distribution)と呼び, 次のように表す:

$$\mathrm{Geometric}(p).$$

幾何分布における確率の総和が 1 になることは, 等比級数の和を使って, 次のように証明される:

$$\sum_{m=0}^{\infty} P(m) = p \sum_{m=0}^{\infty} (1 - p)^m = \frac{p}{1 - (1 - p)} = 1.$$

$1 - (1 - p)$ が p になることがこれ以後頻繁に現れる.

幾何分布の確率質量函数の形は次のようになる.

In [32]:

```
geom = Geometric(0.3)
x = 0:16
bar(x, x -> pdf(geom, x); label="", title="Geometric(0.3)", alpha=0.5)
plot!(; xtick = x)
```

Out[32]:

p が小さい場合の幾何分布は平均 $\lambda = 1/p$ の指数分布で近似される: λ が大きなとき,

$$p(1-p)^m = \frac{1}{\lambda}\left(1-\frac{1}{\lambda}\right)^m \approx \frac{1}{\lambda}\exp\left(-\frac{m}{\lambda}\right).$$

In [33]:

```
geom = Geometric(0.03)
p = succprob(geom)
x = 0:200
plot(x, x -> pdf(geom, x); label="Geometric($p)")
plot!(Exponential(1/p), 0, 200; ls=:dash, label="Exponential(1/$p)")
```

Out[33]:

4.24 幾何分布の期待値と分散

幾何分布に従う確率変数 M の期待値と分散は次のようになる:

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

詳しくは以下のように計算される. 上で経験した等比級数の和の計算から,

$$\sum_{m=0}^{\infty}(1-p)^m = \frac{1}{p}$$

となることもわかる. 両辺を p で微分して -1 倍する操作を2回施すと,

$$\sum_{m=1}^{\infty} m(1-p)^{m-1} = \frac{1}{p^2}, \quad \sum_{m=2}^{\infty} m(m-1)(1-p)^{m-2} = \frac{2}{p^3}.$$

これらの公式を使うと,

$$\begin{aligned}
E[M] &= \sum_{m=0}^{\infty} mP(m) = \sum_{m=1}^{\infty} mp(1-p)^m = p(1-p)\sum_{m=1}^{\infty} m(1-p)^{m-1} = p(1-p)\frac{1}{p^2} = \frac{1-p}{p}, \\
E[M(M-1)] &= \sum_{m=0}^{\infty} m(m-1)P(m) = \sum_{m=2}^{\infty} m(m-1)p(1-p)^m \\
&= p(1-p)^2\sum_{m=2}^{\infty} m(m-1)(1-p)^{m-2} = p(1-p)^2\frac{2}{p^3} = \frac{2(1-p)^2}{p^2} \\
E[M(M-1)] &= E[M^2] - E[M], \\
E[M^2] &= E[M(M-1)] + E[M] = \frac{2(1-p)^2}{p^2} + \frac{1-p}{p}, \\
\mathrm{var}(M) &= E[M^2] - E[M]^2 = \frac{2(1-p)^2}{p^2} + \frac{1-p}{p} - \frac{(1-p)^2}{p^2} = \frac{(1-p)^2}{p^2} + \frac{p(1-p)}{p^2} = \frac{1-p}{p^2}.
\end{aligned}$$

$N = M + 1$ とおく. N は始めて 1 が出るまでの試行回数を意味する確率変数になる. N の確率質量函数は

$$P(n-1) = p(1-p)^{n-1}$$

になる. N の期待値と分散は上の結果より, 次のようになる:

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

$E[N] = 1/p$ の意味: 当たりが確率 p で出るガチャを回し続けるとき, 当たりが初めて出るまでに回す回数の期待値は $1/p$ になる.

注意・警告: 例えば, 当たりが $1\% = 0.01$ の確率で出るガチャを回し続けたときに, 当たりが初めて出るまでに回す回数の期待値は $1/0.01 = 100$ 回になる. これは「だいたい100回程度ガチャを回せばほぼ確実に当たりが出ること」を意味 **しない**. 以下の問題を解いてみよ.

4.25 問題: 当たりが出る確率が1%のガチャを当たりが出るまで回す場合

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

(1) N の期待値と分散と標準偏差を求めよ.

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

確率や標準偏差の値は有効桁3桁以上が合っていれば正解とする.

解答例: $p = 0.01$ とおく.

(1) N の期待値と分散と標準偏差:

$$E[N] = \frac{1}{p} = 100, \quad \mathrm{var}(N) = \frac{1-p}{p^2} = 9900, \quad \mathrm{std}(N) = \sqrt{\mathrm{var}(N)} \approx 99.5.$$

(2) N が150回以上になる確率 = 最初の149回分がすべて外れの確率:

$$(1-p)^{149} = 0.99^{149} \approx 0.224.$$

N が200回以上になる確率 = 最初の199回分がすべて外れの確率:

$$(1-p)^{199} = 0.99^{199} \approx 0.135.$$

N が300回以上になる確率 = 最初の199回分がすべて外れの確率:

$$(1-p)^{299} = 0.99^{299} \approx 0.0495.$$

解答終

注意・警告続き: 当たりが出る確率が1%のガチャを当たりが出るまで回し続ける人達が沢山いると, 300回以上回してしまう人が5%程度すなわち20人に1人程度生じるということになる. このようなガチャを当たりが出るまで回し続ける場合にあなたは, その20人の1人になってしまうことを覚悟しておくべきだと思われる.

注意: p が小さな幾何分布の標準偏差は

$$\sqrt{\frac{1-p}{p^2}} \approx \sqrt{\frac{1}{p^2}} = \frac{1}{p}$$

となる. 実際, 上の場合も $\mathrm{var}(N)$ はほぼ $1/p = 100$ になっている.

In [34]:

```
p = 0.01
@show 1/p
@show (1 - p)/p^2
@show √((1 - p)/p^2)
for Nmin in (150, 200, 300)
    @eval @show (1 - p)^($Nmin - 1)
end
```

```
1 / p = 100.0
(1 - p) / p ^ 2 = 9900.0
√((1 - p) / p ^ 2) = 99.498743710662
(1 - p) ^ (150 - 1) = 0.2236886739786474
(1 - p) ^ (200 - 1) = 0.13533300490703204
(1 - p) ^ (300 - 1) = 0.04953625663766235
```

In [35]:
```
@show mean(Geometric(0.01)) + 1
@show var(Geometric(0.01))
@show std(Geometric(0.01))
for Nmin in (150, 200, 300)
    @eval @show ccdf(Geometric(0.01), $(Nmin - 1) - 1)
end
```

```
mean(Geometric(0.01)) + 1 = 100.0
var(Geometric(0.01)) = 9900.0
std(Geometric(0.01)) = 99.498743710662
ccdf(Geometric(0.01), 149 - 1) = 0.22368867397864767
ccdf(Geometric(0.01), 199 - 1) = 0.13533300490703226
ccdf(Geometric(0.01), 299 - 1) = 0.049536256637662465
```

Monte Carlo 法による近似計算:

In [36]:
```
N = rand(Geometric(0.01), 10^8) .+ 1
@show mean(N)
@show var(N)
@show std(N)
@show mean(≥(150), N)
@show mean(≥(200), N)
@show mean(≥(300), N);
```

```
mean(N) = 99.99065754
var(N) = 9898.741484625854
std(N) = 99.4924192319488
mean((≥)(150), N) = 0.22367281
mean((≥)(200), N) = 0.1352877
mean((≥)(300), N) = 0.04951668
```

4.26 負の二項展開

任意の数 a と0以上の整数 m について, **二項係数** (binomial coefficients)を

$$\binom{a}{m} = \frac{a(a-1)(a-2)\cdots(a-m+1)}{m!}$$

と定める. たとえば, $m = 0, 1, 2, 3$ の場合には,

$$\binom{a}{0} = 1, \quad \binom{a}{1} = a, \quad \binom{a}{2} = \frac{a(a-1)}{2}, \quad \binom{a}{3} = \frac{a(a-1)(a-2)}{3!}.$$

$a = n$ も0以上の整数のとき, これは組み合わせの数と一致する:

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

$f(x) = (1+x)^a$ について,

$$f'(x) = a(1+x)^{a-1}, \quad f''(x) = a(a-1)(1+x)^{a-2}, \quad \dots \quad f^{(m)}(x) = a(a-1)\cdots(a-m+1)(1+x)^{a-m}$$

なので,

$$f^{(m)}(0) = a(a-1)\cdots(a-m+1), \qquad \frac{1}{m!}f^{(m)}(0) = \binom{a}{m}.$$

ゆえに, $f(x) = (1+x)^a$ は $x = 0$ で次のようにTaylor展開される:

$$(1+x)^a = f(x) = \sum_{m=0}^{\infty} \frac{1}{m!} f^{(m)}(0) x^m = \sum_{m=0}^{\infty} \binom{a}{m} x^m = 1 + ax + \frac{a(a-1)}{2}x^2 + \frac{a(a-1)(a-2)}{3!}x^3 + \cdots$$

このTaylor展開は $|x| < 1$ で常に収束している. $a = n$ が0以上の整数の場合には $m = 0, 1, \dots, n$ に関する有限和になり, 通常の二項定理が得られる. このTaylor展開を **二項展開** (binomial expansion)と呼ぶ.

次の形の二項係数を **負の二項係数** と呼ぶことにする:

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

上の二項展開の x, a を $-x, -a$ で置き換えることによって, $|x| < 1$ のとき,

$$(1-x)^{-a} = \sum_{m=0}^{\infty}(-1)^m\binom{-a}{m}x^m = \sum_{m=0}^{\infty}\binom{a+m-1}{m}x^m.$$

この公式を **負の二項展開** と呼ぶことにする.

4.27 負の二項分布の定義

$0 < p \leq 1$ であると仮定し, k は正の整数であるとし, $m = 0, 1, 2, \ldots$ であるとする.

確率 p で 1 が確率 $1-p$ で 0 が生じるようなBernoulli試行を続けたとき, 1 が合計でちょうど k 個出たときの試行回数が $k+m$ になる確率は, 最初の $k+m-1$ 回中の 0 の個数が m 個で最後の $m+k$ 回目に 1 が出る確率

$$\binom{k+m-1}{m}p^{k-1}(1-p)^m \times p$$

に等しいので,

$$P(m) = \binom{k+m-1}{m}p^k(1-p)^m = (-1)^m\binom{-k}{m}p^k(1-p)^m$$

になる. この確率質量函数で定義される可算集合 $\{0, 1, 2, 3, \ldots\}$ 上の無限離散分布を **負の二項分布** (negative binomial distribution)と呼び, 次のように表す:

$$\mathrm{NegativeBinomial}(k, p).$$

負の二項分布における確率の総和が 1 になることは, 前節の負の二項展開からただちに得られる:

$$\sum_{m=0}^{\infty}P(m) = p^k\sum_{m=0}^{\infty}\binom{k+m-1}{m}(1-p)^m = p^k(1-(1-p))^{-k} = p^kp^{-k} = 1.$$

注意: 以上および以下の計算は k が正の整数でなくても,

$$\binom{k+m-1}{m} = \frac{k(k+1)\cdots(k+m-1)}{m!} > 0 \quad (k > 0,\ m = 0, 1, 2, 3, \ldots)$$

なので, $k > 0$ ならば確率質量函数

$$P(m) = \binom{k+m-1}{m}p^k(1-p)^m = (-1)^m\binom{-k}{m}p^k(1-p)^m \quad (m = 0, 1, 2, 3, \ldots)$$

は定義されている. 期待値と分散の計算も上で計算した結果と同じになる. このように拡張された離散分布も負の二項分布と呼ばれる.

4.28 負の二項分布の期待値と分散

M は負の二項分布 $\mathrm{NegativeBinomial}(k, p)$ に従う確率変数であるとし, $N = M + k$ とおく.

M は成功確率 p のBernoulli試行をちょうど k 回 1 が出るまで繰り返すときに出た 0 の個数を意味する確率変数にになる.

N はちょうど k 回 1 が出たときの試行回数を意味する確率変数であり, その確率質量函数は

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

になる(この式の形は二項分布の確率質量函数に近いが少しだけ違っている).

それらの期待値と分散は次のようになる:

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

これらは以下のようにして示される. 二項係数について

$$
\begin{aligned}
m\binom{a+m-1}{m} &= m\frac{a(a+1)\cdots(a+m-1)}{m!} \\
&= a\frac{(a+1)\cdots(a-m+1)}{(m-1)!} = a\binom{a+m-1}{m-1}, \\
m(m-1)\binom{a+m-1}{m} &= m(m-1)\frac{a(a+1)(a+2)\cdots(a+m-1)}{m!} \\
&= a(a+1)\frac{(a+2)\cdots(a+m-1)}{(m-2)!} = a(a+1)\binom{a+m-1}{m-2}
\end{aligned}
$$

が成立することとと負の二項展開を使うと, 負の二項分布に従う確率変数 M の期待値と分散は以下のように計算される:

$$
\begin{aligned}
E[M] &= \sum_{m=0}^{\infty} mP(m) = p^k \sum_{m=1}^{\infty} m\binom{k+m-1}{m}(1-p)^m = p^k \sum_{m=1}^{\infty} k\binom{k+m-1}{m-1}(1-p)^m \\
&= p^k k(1-p) \sum_{m=1}^{\infty} \binom{(k+1)+(m-1)-1}{m-1}(1-p)^{m-1} = p^k k(1-p)(1-(1-p))^{-(k+1)} \\
&= p^k k(1-p)p^{-(k+1)} = \frac{k(1-p)}{p}, \\
E[M(M-1)] &= \sum_{m=0}^{\infty} m(m-1)P(m) = p^k \sum_{m=2}^{\infty} m(m-1)\binom{k+m-1}{m}(1-p)^m = p^k \sum_{m=2}^{\infty} k(k+1)\binom{k+m-1}{m-2}(1-p)^m \\
&= p^k k(k+1)(1-p)^2 \sum_{m=1}^{\infty} \binom{(k+2)+(m-2)-1}{m-2}(1-p)^{m-2} = p^k k(k+1)(1-p)^2(1-(1-p))^{-(k+2)} \\
&= p^k k(k+1)(1-p)^2 p^{-(k+2)} = \frac{k(k+1)(1-p)^2}{p^2}, \\
E[M(M-1)] &= E[M^2] - E[M], \\
E[M^2] &= E[M(M-1)] + E[M] = \frac{k(k+1)(1-p)^2}{p^2} + \frac{k(1-p)}{p}, \\
\mathrm{var}(M) &= E[M^2] - E[M]^2 = \frac{k(k+1)(1-p)^2}{p^2} + \frac{k(1-p)}{p} - \frac{k^2(1-p)^2}{p^2} \\
&= \frac{-k(1-p)^2}{p^2} + \frac{kp(1-p)}{p^2} = \frac{k(1-p)}{p^2}.
\end{aligned}
$$

上の結果より, $N = M + k$ の期待値と分散は次のようになる:

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

注意: 負の二項分布に従う確率変数の期待値と分散は $k = 1$ の場合の幾何分布の期待値と分散のちょうど k 倍になっている. これは偶然ではなく, 必然的にこうなる理由がある. この点については後で確率変数の独立性について説明してから詳しく説明する.

In [37]:

```
@vars k p
k*(k+1)*(1-p)^2/p^2 + k*(1-p)/p - k^2*(1-p)^2/p^2 ▷ simplify
```

Out[37]: $\frac{k(1-p)}{p^2}$

In [38]:

```
k, p = 4, 0.25
negbin = NegativeBinomial(k, p)
@show mean(negbin), k*(1-p)/p
@show mean(negbin) + k, k/p
@show var(negbin), k*(1-p)/p^2;
```

```
(mean(negbin), (k * (1 - p)) / p) = (12.0, 12.0)
(mean(negbin) + k, k / p) = (16.0, 16.0)
(var(negbin), (k * (1 - p)) / p ^ 2) = (48.0, 48.0)
```

In [39]:

```
negbin = NegativeBinomial(10, 0.5)
μ, σ = mean(negbin), std(negbin)
x = 0:round(Int, μ + 4σ)
bar(x, x -> pdf(negbin, x); label="",
    title="NegativeBinomial(10, 0.5)", alpha=0.5)
plot!(; xtick=0:5:maximum(x)+5)
```

Out[39]:

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

In [40]:

```
k, θ = 5, 2
L = 10
p = 1/(L*θ)

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

Out[40]:

In [41]:

```
k, θ = 5, 2
L = 100
p = 1/(L*θ)

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

Out[41]:

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

In [42]:

```
k, p = 100, 0.4

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

Out[42]:

In [43]:

```
k, p = 1000, 0.4

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

Out[43]:

4.29 問題: 当たりが出る確率が5%のガチャを当たりが5回出るまで回す場合

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

(1) N の期待値と分散と標準偏差を求めよ.

(2) N が150回以上, 200回以上になる確率を求めよ.

確率や標準偏差の値は有効桁3桁以上が合っていれば正解とする.

解答例: $p = 0.05$, $k = 5$ とおく.

(1) N の期待値と分散と標準偏差:

$$E[N] = \frac{k}{p} = 100, \quad \mathrm{var}(N) = \frac{k(1-p)}{p^2} = 1900, \quad \mathrm{std}(N) = \sqrt{\mathrm{var}(N)} \approx 43.58.$$

(2) ($N \geq 150$ となる確率) = 1 − ($N \leq 149$ となる確率):

$$1 - \sum_{n=1}^{149} \binom{n-1}{k-1} p^k (1-p)^{n-k} \approx 0.129.$$

($N \geq 200$ となる確率) = 1 − ($N \leq 199$ となる確率):

$$1 - \sum_{n=1}^{199} \binom{n-1}{k-1} p^k (1-p)^{n-k} \approx 0.0273.$$

解答終

In [44]:

```
p = 5//100
k = 5
@show 5/p 5(1 - p)/p^2 √(5(1 - p)/p^2);
```

```
5 / p = 100//1
(5 * (1 - p)) / p ^ 2 = 1900//1
√((5 * (1 - p)) / p ^ 2) = 43.58898943540674
```

In [45]:
```
p = 0.05
k = 5
for Nmin in (150, 200)
    @show 1 - sum(binomial(n-1, k-1)*p^k*(1-p)^(n-k) for n in 1:Nmin-1)
end
```

```
1 - sum((binomial(n - 1, k - 1) * p ^ k * (1 - p) ^ (n - k) for n = 1:Nmin - 1)) = 0.12922049010
314218
1 - sum((binomial(n - 1, k - 1) * p ^ k * (1 - p) ^ (n - k) for n = 1:Nmin - 1)) = 0.02734419238
495378
```

In [46]:
```
p = 0.05
k = 5
@eval @show negbin = NegativeBinomial($k, $p)
@show mean(negbin) + k
@show var(negbin)
@show std(negbin)
for Nmin in (150, 200)
    @show ccdf(negbin, (Nmin - 1) - k)
end
```

```
negbin = NegativeBinomial(5, 0.05) = NegativeBinomial{Float64}(r=5.0, p=0.05)
mean(negbin) + k = 100.0
var(negbin) = 1899.9999999999995
std(negbin) = 43.58898943540674
ccdf(negbin, (Nmin - 1) - k) = 0.12922049010313866
ccdf(negbin, (Nmin - 1) - k) = 0.027344192384949475
```

Monte Carlo 法による近似計算:

In [47]:
```
M = rand(negbin, 10^8) .+ k
@show mean(N)
@show var(N)
@show std(N)
@show mean(≥(150), N)
@show mean(≥(200), N);
```

```
mean(N) = 99.99065754
var(N) = 9898.741484625854
std(N) = 99.4924192319488
mean((≥)(150), N) = 0.22367281
mean((≥)(200), N) = 0.1352877
```

4.30 多項分布の定義

Bernoulli試行から二項分布が得られたのと同じように, カテゴリカル分布の同分布独立試行から多項分布が得られる.

次のカテゴリカル分布を考える:

$$\mathrm{Categorical}(p_1, \dots, p_r) = \begin{array}{|c|cccc|} \hline x & 1 & 2 & \cdots & r \\ \hline P(x) & p_1 & p_2 & \cdots & p_r \\ \hline \end{array}$$

このカテゴリカル分布にしたがって, 番号の $1, \dots, r$ のどれかをランダムに生成することを独立に n 回繰り返すと, 番号 $1, \dots, r$ からなる長さ n の数列ができる. 総和が n になる0以上の整数達の組 $(k_1, \dots, k_r)$ が与えられたとき, その数列中に番号 $1, \dots, r$ がそれぞれ $k_1, \dots, k_r$ 個含まれている確率は

$$P(k_1, \dots, k_r) = \frac{n!}{k_1! \cdots k_r!} p_1^{k_1} \cdots p_r^{k_r}$$

になる. この確率質量函数によって定義された有限集合

$$\{, (k_1, \dots, k_r) \in \mathbb{Z}_{\geq 0}^r \mid k_1 + \cdots + k_r \}$$

上の多変量離散分布を多項分布と呼び,

$$\mathrm{Multinomial}(n, p_1, \dots, p_r)$$

と表す. $r = 2$ の場合の多項分布は本質的に二項分布に等しい.

5 連続分布など

5.1 連続分布と確率密度函数

$a, b \in \mathbb{R} \cup \{\pm\infty\}$, $a < b$ であるとし, 区間 $(a, b) = \{\, x \in \mathbb{R} \mid a < x < b \,\}$ を考える. ($a = -\infty$ や $b = \infty$ となっている場合があるので注意せよ.) このとき, 実数 $x \in (a, b)$ を0以上の実数に対応させる函数 $p(x)$ で

$$\int_a^b p(x)\,dx = 1$$

を満たすものを区間 (a, b) 上の **確率密度函数** (probability density function, pdf)もしくは単に **密度函数** (density function)と呼ぶ.

密度函数の定義域を実数全体に拡張する場合には区間 (a, b) の外での値は 0 にする.

区間 (a, b) 上の確率密度函数が与えられているとき, 区間 (a, b) 上の **連続確率分布** もしくは **連続分布** が与えられているいう.
(区間の両端を区間に含めても含めなくてもよいが, 連続分布の場合にはその違いを気にする必要はない. なぜならば, 連続分布において一点の確率は0だからである.)

確率分布を単に分布と呼ぶことがよくある.

1次元の区間を n 次元の $\mathbb{R}^n$ における領域に一般化して, n 変量の連続分布を定義することもできる. その場合には上の積分は n 重積分になる. 多変量連続分布については後で具体的な例を通して説明することにする.

5.2 一様分布の定義

区間 (a, b) 上の確率密度函数

$$p(x) = \frac{1}{b-a} \qquad (a < x < b)$$

によって与えられる連続分布を区間 (a, b) 上の一様分布と呼び, 次のように表す:

$$\mathrm{Uniform}(a, b).$$

半開区間 $[a, b)$, $(a, b]$ や閉区間 $[a, b]$ 上の一様分布も同様に定義されるが, 区間の両端を含めるか否かは本質的ではない.

コンピュータでの `rand()` 函数は近似的に区間 $[0, 1)$ 上の一様分布に従う確率変数とみなされる.

一様分布 $\mathrm{Uniform}(0, 1)$ の確率密度函数のグラフは以下のようになる.

In [48]:
```
plot(Uniform(0, 1), -0.4, 1.4; label="", title="Uniform(0, 1)")
```

Out[48]:

5.3 正規分布の定義

$\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}$, $\sigma > 0$ と仮定する. 実数直線 $\mathbb{R} = (-\infty, \infty)$ 上の確率密度函数

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

によって与えられる連続分布を平均 μ, 分散 σ^2 の **正規分布** (normal distribution)と呼び,

$$\mathrm{Normal}(\mu, \sigma)$$

と表す. パラメータ σ は標準偏差と呼ばれる.

特に平均 0, 分散 1 の正規分布を **標準正規分布** (standard normal distribution)と呼び,

$$\mathrm{Normal}() = \mathrm{Normal}(0, 1)$$

と表すことにする.

分布のパラメータとしての平均 μ や分散 σ^2 と後で出て来る標本の平均や分散を概念的に厳密に区別する必要がある.

互いに密接に関係はしているが異なる概念を同じ用語で呼ぶことが原因で初学者は大抵の場合混乱する.

混乱せずに済ますためには 「平均」や「分散」という用語を使う場合には必ず「○○分布の平均と分散」や「確率変数○○の平均と分散」や「標本○○の平均と分散」のように **「〜の平均と分散」という言い方をすればよいだろう.**

正規分布の確率密度函数のグラフは以下の形になる. この形は「釣鐘型」(bell-shaped)と表現されることが多い.

In [49]:
```
plot(Normal(2, 3), 2-4*3, 2+4*3; label="", title="Normal(2, 3)", xtick=-100:2:100)
```

Out[49]:

In [50]:
```
plot(Normal(), -4, 4; label="", title="Normal()", xtick=-100:100)
```

Out[50]:

標準正規分布において, 値が -1 以上 1 以下になる確率, -2 以上 2 以下になる確率, -1.96 以上 1.96 以下になる確率はそれぞれ以下のようになる.

In [51]:
```
@show cdf(Normal(), 1) - cdf(Normal(), -1)
plot(Normal(), -4, 4; label="", title="Normal()", xtick=-100:100)
plot!(Normal(), -1, 1; label="", c=1, fillrange=0, fillcolor=:red, fillalpha=0.3)
```

```
cdf(Normal(), 1) - cdf(Normal(), -1) = 0.6826894921370861
```

Out[51]:

In [52]:
```
@show cdf(Normal(), 2) - cdf(Normal(), -2)
plot(Normal(), -4, 4; label="", title="Normal()", xtick=-100:100)
plot!(Normal(), -2, 2; label="", c=1, fillrange=0, fillcolor=:red, fillalpha=0.3)
```

```
cdf(Normal(), 2) - cdf(Normal(), -2) = 0.9544997361036416
```

Out[52]:

In [53]:
```
@show cdf(Normal(), 1.96) - cdf(Normal(), -1.96)
plot(Normal(), -4, 4; label="", title="Normal()", xtick=-100:100)
plot!(Normal(), -1.96, 1.96; label="", c=1, fillrange=0, fillcolor=:red, fillalpha=0.3)
```

```
cdf(Normal(), 1.96) - cdf(Normal(), -1.96) = 0.950004209703559
```

Out[53]:

5.4 連続確率変数

区間 (a, b) 上の確率密度函数 $p(x)$ が与えられているとする.

さらに, 区間 (a, b) 上の実数値函数を **連続分布における確率変数** と呼ぶ. (積分が定義できるような良い函数に限ることにする. 複素数値, ベクトル値, 行列値の連続確率変数を考えることもある.)

たとえば, $x \in (a, b)$ をそれ自身 $x \in \mathbb{R}$ に対応させる函数は確率変数とみなされる. 以下ではそれを X と書き, 確率密度函数 $p(x)$ を持つ **連続確率変数** と呼ぶ(確率密度函数を持つ確率変数を連続確率変数と呼ぶ). そのとき,

$$X \sim p(x)$$

と書くことがある. さらに, 確率密度函数 $p(x)$ で与えられる連続分布の名前が Dist のとき, X は分布 Dist に従う確率変数であると言い(「従う」は "follow" の訳語), 以下のように書く:

$$X \sim \mathrm{Dist}.$$

5.5 連続確率変数に関する確率とその期待値

値が $a \leq c \leq d \leq b$ であると仮定する.

確率密度函数 $p(x)$ を持つ確率変数 X が c より大きく d より小さくなる確率 $P(c < X < d)$ を積分を使って

$$P(c < X < d) = \int_c^d p(x)\,dx$$

と定義する. c 以上 d 以下の確率 $P(c \leq X \leq d)$ なども同じ積分の値で定義する:

$$P(c \le X \le d) = \int_c^d p(x)\,dx.$$

特に X の値がぴったり c に等しくなる確率 $P(X = c)$ は次のように 0 になる:

$$P(X = c) = P(c \le X \le c) = \int_c^c p(x)\,dx = 0.$$

$x \in (a, b)$ を含む長さ dx の微小区間 $I = (x - t\,dx, x + u\,dx) \subset (a, b)$ $(t, u \ge 0, t + u = 1)$ に値が含まれる確率は

$$P(X \in I) = \int_{x-t\,dx}^{x+u\,dx} p(x')\,dx' \approx p(x)\,dx$$

のように近似される. これが $p(x)$ が確率の密度であることの意味である.

確率変数 $f(X)$ の期待値 $E[f(X)]$ が次のように定義される:

$$E[f(X)] = \int_a^b f(x)p(x)\,dx.$$

ただし, これは $\int_a^b |f(x)|p(x)\,dx$ が有限の値に収束する場合(積分が絶対収束する場合)にのみ定義されていると考える.

X の従う分布の名前が Dist のとき $E[X]$ を **分布** Dist **の期待値** と呼ぶ.

期待値を **平均** または **平均値** と呼ぶこともあるが, あとで定義する標本の平均値と確率変数の平均＝期待値を区別する必要があるので注意が必要である.

確率変数 $1_{c \le f(X) \le d}(X)$ を

$$1_{c \le f(X) \le d}(x) = \begin{cases} 1 & (c \le x \le d) \\ 0 & (\text{otherwise}) \end{cases}$$

と定めると,

$$E[1_{c \le f(X) \le d}(X)] = \int_a^b 1_{c \le f(X) \le d}(x)p(x)\,dx = \int_c^d p(x)\,dx = P(c \le X \le d).$$

このように期待値の概念は確率の概念を含んでいると考えてよい.

再度強調: 確率変数の定義を期待値が定義されている変数(のようなもの)のことだと思ってもよい と言ってよいくらい, 期待値の概念は確率変数について基本的である. 統計学に必要な理論的な計算の多くが期待値を取る操作 $E[\ ]$ の計算に帰着されることになる.

5.6 確率変数の確率密度函数

確率変数 $Y = f(X)$ を考える. ある y の確率密度函数 $p_Y(y)$ が存在して, Y の函数達 $g(Y)$ について

$$E[g(Y)] = \int g(y)\,p_Y(y)\,dy$$

と書けるとき, Y は **確率密度函数** $p_Y(y)$ **を持つ** という. 右辺は定積分であることに注意せよ. 積分区間は場合ごとに異なるので表記を省略した. この説明の曖昧さを除くために例を使って説明しよう.

例: X は一様分布 Uniform$(0, 2)$ に従う確率変数であるとし, $Y = X^2$ とおく. このとき, 積分変数を $x = \sqrt{y}$ によって変換すると, $dx = y^{-1/2}\,dy/2$ なので

$$E[g(Y)] = \frac{1}{2}\int_0^2 g(x^2)\,dx = \int_0^{\sqrt{2}} g(y)\,\frac{1}{4}y^{-1/2}\,dy.$$

ゆえに $Y = X^2$ は確率密度函数

$$p_Y(y) = \frac{1}{4}y^{-1/2} \qquad \left(0 < y < \sqrt{2}\right)$$

を持つ.

例: Z は標準正規分布に従う確率変数であるとし, $Y = Z^2$ とおく. このとき z^2 が z の偶函数であることと, 積分の変数変換 $z = \sqrt{y}$ によって,

$$E[g(Y)] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(z^2)\,e^{-z^2/2}\,dz = \frac{2}{\sqrt{2\pi}}\int_0^{\infty} g(z^2)\,e^{-z^2/2}\,dz = \frac{1}{\sqrt{2\pi}}\int_0^{\infty} g(y)\,e^{-y}\,y^{-1/2}\,dy.$$

ゆえに $Y = Z^2$ は確率密度函数

$$p_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y} y^{-1/2} \quad (y > 0)$$

を持つ. この確率密度函数で定義される連続分布は **自由度1のχ²分布** (カイ二乗分布)と呼ばれている.

記号法上の注意: 以上における $p_Y(y)$ を単に $p(y)$ と書いて, x と y の変数名の違いによって, X の確率密度函数 $p(x)$ と区別することもある. そういうスタイルは初学者の混乱の原因になりえるが, 手計算をするときには非常に便利であり, 無駄な添え字を付けることを無くせるので結構便利である. 実際によく使われるので, 文献を読むときには知っておかないと困る知識でもある.

5.7 確率変数の確率質量函数

連続確率変数 X の函数 $Y = f(X)$ が離散確率変数になる場合もあることにも注意せよ.

$Y = f(X)$ の取り得る値が離散的な $\{b_1, b_2, b_3, \ldots\}$ に限り, ある確率質量函数 $P_Y(y)$ が存在して,

$$E[g(Y)] = \sum_i g(b_i) P_Y(b_i)$$

が成立するとき, $Y = f(X)$ は **確率質量函数** $P_Y(y)$ **を持つ** という.

例: T は一様分布 $\mathrm{Uniform}(0, 1)$ に従う連続確率変数であるとし, $X = f(T)$ を

$$f(t) = \begin{cases} 1 & (t < 0.3) \\ 0 & (t \geq 0.3) \end{cases}$$

によって定義する. このとき,

$$E[g(X)] = \int_0^1 g(f(t))\, dt = \int_0^{0.3} g(1)\, dt + \int_{0.3}^1 g(0)\, dt = g(1) \cdot 0.3 + g(0) \cdot 0.7.$$

ゆえに $X = f(T)$ は確率質量函数

$$P_X(1) = 0.3, \quad P_X(0) = 0.7$$

を持つ. この例は一様乱数からBernoulli試行を作るときにすでに登場している.

5.8 連続分布における確率変数の期待値を取る操作の基本性質

連続確率変数 X の函数 $f(X)$ の期待値を取る操作(期待値汎函数) $E[f(X)]$ は以下の性質を満たしている.

- 線形性: $E[\alpha f(X) + \beta g(X)] = \alpha E[f(X)] + \beta E[g(X)]$,
- 単調性: 任意の $x = a_i$ について $f(x) \leq g(x)$ ならば $E[f(X)] \leq E[g(X)]$,
- 規格化: $E[\alpha] = \alpha$. (特に $E[1] = 1$).

この性質はすでに説明した離散確率変数の基本性質と完全に同じ形をしており, 離散の場合の和を積分に置き換えれば全く同じ方法で証明される.

5.9 問題: Markovの不等式とChebyshevの不等式

X は確率密度函数または確率質量函数を持つ確率変数であるとする.

$a, t > 0$ のとき, 確率変数 $Y = f(X)$ について次が成立することを示せ:

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

この不等式の $a = 1$ の場合を **Markovの不等式** と呼び, $a = 2$ の場合を **Chebyshevの不等式** と呼ぶ.

解答例: 確率変数 $1_{|Y| \geq t}(X)$ を

$$1_{|Y| \geq t}(x) = \begin{cases} 1 & (|f(x)| \geq t) \\ 0 & (\text{otherwise}) \end{cases}$$

と定めると,

$$P(|Y| \geq t) = E[1_{|Y| \geq t}(X)]$$

となるのであった. $|f(x)| \geq t$ ならば $1_{|Y| \geq t}(x) = 1$ となるので,

$$|f(x)|^a \geq t^a = t^a 1_{|Y|\geq t}(x)$$

となり, $|f(x)| < t$ ならば $1_{|Y|\geq t}(x) = 0$ となるので

$$|f(x)|^a \geq 0 = t^a 1_{|Y|\geq t}(x)$$

が自明に成立する. ゆえに期待値を取る操作の単調性と線形性より,

$$E[|Y|^a] \geq t^a E[1_{|Y|\geq t}(X)] = t^a P(|Y| \geq t).$$

両辺を t^a で割れば示したい不等式が得られる.

解答終

5.10 問題: Jensenの不等式

$y = f(x)$ は上に凸な函数であるとする. ある一次函数 $y = a(x-\mu) + f(\mu)$ が存在して(この一次函数は直観的には $y = f(x)$ の接線である), $f(x) \leq a(x-\mu) + f(\mu)$ という不等式が成立することを認めて, 確率変数 X について以下が成立することを示せ:

(1) $f(x)$ が上に凸な函数ならば $E[f(X)] \leq f(E[X])$

(2) $f(x)$ が下に凸な函数ならば $E[f(X)] \geq f(E[X])$

これらを **Jensenの不等式** (イェンセンの不等式)と呼ぶ.

解答例: $f(x)$ は上に凸な函数であるとし, $\mu = E[X]$ とおく. ある一次函数 $y = a(x-\mu) + f(\mu)$ が存在して

$$f(x) \leq a(x-\mu) + f(\mu)$$

という不等式が成立する. この両辺の x に確率変数 X を代入して期待値を取ると, 期待値を取る操作の単調性より,

$$E[f(X)] \leq E[a(X-\mu) + f(\mu)].$$

右辺に期待値を取る操作の線形性と規格化条件を適用し, $\mu = E[X]$ を代入すると,

$$E[f(X)] \leq a(E[X]-\mu) + f(\mu) = f(E[X]).$$

これで(1)が示された. $f(x)$ が下に凸な場合には $-f(x)$ が上に凸になるので, (1)から(2)が導かれることがわかる.

解答終

注意: 以上のように, Jensenの不等式は期待値を取る操作の基本性質をちょうど全部使えば証明できる. 期待値を取る操作の基本性質が十分に強力であることがこの証明から分かる.

5.11 問題: 相加相乗平均の不等式

Jensenの不等式を使って, $a_1, \ldots, a_n > 0$ のとき, 次が成立することを示せ:

$$(a_1 \cdots a_n)^{1/n} \leq \frac{a_1 + \cdots + a_n}{n}.$$

解答例: $E[f(X)]$ を

$$E[f(X)] = \frac{f(a_1) + \cdots + f(a_n)}{n}$$

と定めると, $E[\ ]$ は期待値を取る操作の基本性質をすべて満たしており, $f(x) = \log x$ のとき

$$E[\log X] = \frac{\log a_1 + \cdots + \log a_n}{n} = \log(a_1 \cdots a_n)^{1/n}, \quad E[X] = \frac{a_1 + \cdots + a_n}{n}.$$

ゆえに $f(x) = \log x$ (これは上に凸)に関するJensenの不等式 $E[\log X] \leq \log E[X]$ より,

$$\log(a_1 \cdots a_n)^{1/n} \leq \log \frac{a_1 + \cdots + a_n}{n}.$$

$\log x$ 単調増加函数なので, $(a_1 \cdots a_n)^{1/n} \leq (a_1 + \cdots + a_n)/n$ が得られる.

解答終

5.12 問題: Kullback-Leibler情報量とGibbsの情報不等式

2つの確率密度函数 $p(x)$, $q(x)$ に対して, **Kullback-Leibler情報量** $D(q||p)$ を

$$D(q||p) = \int q(x) \log \frac{q(x)}{p(x)}$$

と定める. このとき

$$D(q||p) \geq 0$$

となることを示せ. この不等式を **Gibbsの情報不等式** と呼ぶ.

解答例: $x > 0$ について, $f(x) = x \log x$ とおくと, $f'(x) = \log x + 1$, $f''(x) = 1/x > 0$ なので, $f(x)$ は下に凸な函数になる. 確率密度函数 $p(x)$ で定義される分布に従う確率変数を X と書き, $Y = q(x)/p(x)$ とおく. このとき,

$$\begin{aligned} E[Y] &= \int \frac{q(x)}{p(x)}\, p(x)\, dx = \int q(x)\, dx = 1, \\ E[f(Y)] &= \int \frac{q(x)}{p(x)} \log \frac{q(x)}{p(x)}\, p(x)\, dx = \int q(x) \log \frac{q(x)}{p(x)}\, dx = D(q||p). \end{aligned}$$

下の凸な函数 $f(x)$ に関するJensenの不等式より,

$$E[f(Y)] \geq f(E[Y]) = f(1) = 1 \log 1 = 0.$$

以上を合わせるとGibbsの情報不等式 $D(q||p) \geq 0$ が得られる.

解答終

注意: Kullback-Leibler情報量は2つの確率質量函数 $P(x)$, $Q(x)$ についても同様に

$$D(q||p) = \sum_x Q(x) \log \frac{Q(x)}{Q(x)}$$

と定義され, 上の証明における積分を和で置き換えれば, 全く同様にしてGibbsの情報不等式を示せる.

注意: Kullback-Leibler情報量は2つの確率分布の違いの大きさを測る指標として使われるので, Gibbsの情報不等式は基本的である. ($D(q||q) = 0$ となることにも注意せよ.) Kullback-Leibler情報量の **Sanovの定理** より, Kullback-Leibler情報量 $D(q||p)$ は分布 p による分布 q のMonte Carloシミュレーションの誤差の大きさの指標の1つになっていることがわかる. シミュレートする側の p とされる側の q の立場は非対称なので $D(q||p)$ が p と q の交換について不変でないことは自然である. Kullback-Leibler情報量のこの性質は以下の2つの異なる使い方をされることがある:

(1) 未知の真の分布 q がそのモデル化とみなせる分布 p でどれだけ近似されているかを測るための指標として使う. (汎化誤差の概念はこのようにして得られる. この考え方はモデル選択における情報量規準の理論の基礎になっている.)

(2) 求めるには大量の計算が必要な分布 p を計算量が少なくてすむ特殊な形の分布 q で近似したいことがある. そのときに, 特殊な形の分布 q で分布 p から(Monte Carloシミュレーションの意味で)最も出て来やすい分布を選ぶためにはKullback-Leibler情報量 $D(q||p) を最小化すればよい. (ベイズ統計における変分推論はこの典型例になっている.)

5.13 連続分布における確率変数の分散

連続確率変数 X について, 確率変数 $Y = f(X)$ の期待値を

$$\mu_Y = E[Y] = E[f(X)].$$

と書くとき, $Y = f(X)$ の **分散**(variance) $\mathrm{var}(f(X))$ は次のように定義される:

$$\mathrm{var}(Y) = E\left[(Y - \mu_Y)^2\right] = \int_a^b (f(x) - \mu_Y)^2 p(x)\, dx.$$

これが $\mathrm{var}(Y) = E[Y^2] - E[Y]^2$ を満たすことを離散の場合と完全に同様に証明できる.

連続分布における確率変数 Y の場合もその標準偏差を $\mathrm{std}(Y) = \sqrt{\mathrm{var}(Y)}$ と定める.

X の従う分布の名前が Dist のとき $\mathrm{var}(X)$ を **分布** Dist **の分散** と呼ぶ. 標準偏差についても同様とする.

5.14 問題: 正規分布モデルによる他の分布の最良近似

$p(x)$ が平均 μ, 分散 σ^2 を持つ正規分布の確率密度函数であるとする:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}.$$

確率密度函数 $q(x)$ が定める分布の平均と分散をそれぞれ μ_0, σ_0^2 と書こう:

$$\mu_0 = \int x\, q(x)\, dx, \quad \sigma_0 = \int (x-\mu_0)^2\, q(x)\, dx.$$

ここで $\int \cdots dx$ は実数直線全体での定積分を表す. このとき, 少し上の方で定義したKullback-Leibler情報量

$$D(q||p) = \int q(x) \log \frac{q(x)}{p(x)}\, dx = \int (-\log p(x)) q(x)\, dx + \int q(x) \log q(x)\, dx$$

は次の形になることを示せ:

$$D(q||p) = \frac{\sigma_0^2 + (\mu-\mu_0)^2}{2\sigma^2} + \log \sigma^2 + \log \sqrt{2\pi} + \int q(x) \log q(x)\, dx.$$

特にこの場合には, 固定された $q(x)$ に対して, $D(q||p)$ を最小にする μ, σ^2 の組み合わせは $\mu = \mu_0, \sigma^2 = \sigma_0^2$ になることを示せ.

解答例: $(x-\mu)^2 = ((x-\mu_0)-(\mu-\mu_0))^2 = (x-\mu_0)^2 - 2(\mu-\mu_0)(x-\mu_0) + (\mu-\mu_0)^2$ と正規分布の密度函数 $p(x)$ の定義より,

$$\begin{aligned} -\log p(x) &= \frac{(x-\mu)^2}{2\sigma^2} + \log \sigma^2 + \log \sqrt{2\pi} \\ &= \frac{(x-\mu_0)^2 - 2(\mu-\mu_0)(x-\mu_0) + (\mu-\mu_0)^2}{2\sigma^2} + \frac{1}{2}\log \sigma^2 + \log \sqrt{2\pi} \end{aligned}$$

ゆえに μ_0 と σ_0^2 の定義を使うと,

$$D(q||p) = \frac{\sigma_0^2 + (\mu-\mu_0)^2}{2\sigma^2} + \frac{1}{2}\log \sigma^2 + \log \sqrt{2\pi} + \int q(x) \log q(x)\, dx.$$

任意に固定された $\sigma^2 > 0$ について, これを最小化する μ は $\mu = \mu_0$ である. そのとき,

$$D(q||p) = \frac{1}{2}\left(\frac{\sigma_0^2}{\sigma^2} + \log \sigma^2\right) + (\text{constant of } \sigma^2)$$

これを σ^2 の函数とみなして増減表を書くと, これを最小化する σ^2 は $\sigma^2 = \sigma_0^2$ であることがわかる. 実際, $t = \sigma^2, t_0 = \sigma_0^2$, $f(t) = t_0/t + \log t$ とおくと, $f'(t) = (t-t_0)/t^2$ なので,

t	0	$\cdots$	t_0	$\cdots$	∞
$f'(t)$		$-$	0	$+$	
$f(t)$	∞	$\searrow$	$1+\log t_0$	$\nearrow$	∞

と増減表を書ける.

解答終

注意: 以上の結果は, 与えられた分布 $q(x)$ をKullback-Leibler情報量 $D(q||p)$ の最小化の意味で裁量近似する正規分布 $p(x)$ は $q(x)$ と同じ平均と分散を持つ正規分布になることを意味している. これ以後, 分布を定義するごとにその期待値(平均)と分散を計算して行くことになるが, そのときやっていることは本質的に「正規分布による最良近似」の計算だということになる.

5.15 確率変数の一次函数による変換(アフィン変換)の期待値と分散

離散および連続な確率変数 Y と定数 a, b について,

$$E[aY+b] = aE[Y] + b, \quad \mathrm{var}(aY+b) = a^2\, \mathrm{var}(Y)$$

となることを以下のようにして, $E[\ ]$ の基本性質の線形性と規格化条件のみを使って示せる.

$$\begin{aligned} &E[aY+b] = aE[Y] + E[b] = aE[Y] + b. \\ &\mathrm{var}(aY+b) = E[(aY+b-(aE[Y]+b))^2] \\ &\qquad = E[a^2(Y-E[Y])^2] = a^2 E[(Y-E[Y])^2] = a^2\, \mathrm{var}(Y). \end{aligned}$$

5.16 分布のアフィン変換, スケール変換, 平行移動, 逆数の定義

確率変数 Y が分布 D に従うとき, $aY+b$ $(a, b \in \mathbb{R}, a \neq 0)$ が従う分布を D の **アフィン変換** (affine transformation)と呼び,

$$aD + b$$

と書く. $a > 0$ のとき aD を D の **スケール変換** (scale transformation)と呼び, $D + b$ を分布 D の **平行移動** (translation)と呼ぶ.

さらに, $1/Y$ が従う分布を

$$1/D$$

と書き, D の **逆数** (inverse)と呼ぶことにする.

5.17 確率変数の累積分布函数

確率変数 Y に対して

$$F_Y(y) = P(Y \leq y)$$

で定義される実数 y の函数 $F_Y(y)$ を Y もしくは Y が従う確率分布の **累積分布函数** (cumulative distribution function, cdf)と呼ぶ.

Y が連続確率変数で確率密度函数 $p_Y(y)$ を持つとき,

$$F_Y(y) = \int_{-\infty}^{y} p_Y(y)\, dy.$$

ただし, 確率密度函数 $p_Y(y)$ の定義域は拡張した部分での値が 0 になるように実数直線全体に拡張しておく. 連続分布の累積分布函数の導函数は確率密度函数になる.

Y が離散確率変数で取り得る値の集合が $\{b_1, b_2, \ldots\}$ で確率質量函数 $P_Y(y)$ を持つとき,

$$F_Y(y) = \sum_{b_i \leq y} P_Y(b_i).$$

累積分布函数 $F_Y(y)$ は y について(広義)単調増加函数になり, $y \to -\infty$ で 0 に収束し, $y \to \infty$ で 1 に収束する.

さらに, $F_Y(y)$ の定義 $P(Y \leq y)$ において不等号 $\leq$ に等号が含まれていることより, 累積分布函数 $F_Y(y)$ は y について右連続になる. すなわち, y が数直線上で右から b に近付くとき, $F_Y(y)$ は $F_Y(b)$ に近付く:

$$\lim_{y \searrow b} F_Y(y) = F_Y(b).$$

5.18 分位点函数 (quantile function)

確率分布の累積分布函数 $p = F_Y(y)$ の逆函数 $y = Q_Y(p)$ を **分位点函数** (**クォンタイル函数**, quantile function)と呼ぶ. 逆函数が存在しない場合には逆函数の代わりに使える函数を適切に定義して分位点函数と呼んで利用する.

分位点函数が定義されているとき, $Q_2 = Q_Y(1/2)$ は Y の分布の中央値と呼ばれ, $Q_1 = Q_Y(1/4)$, $Q_3 = Q_Y(3/4)$ は Y の分布の第1四分位点, 第3四分位点と呼ばれる(「四分位数」と書いてもよい).

ただし, これらは確率変数もしくはそれが従う分布の中央値や四分位点であり, 標本(サンプル, データ)の中央値や四分位点とは概念的に明瞭に区別しておかなければいけない.

5.19 Bernoulli分布の累積分布函数

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

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

In [54]:
```
ber = Bernoulli(0.3)
x = vcat([-1:0.02:-0.02; NaN], [0:0.02:0.98; NaN], 1:0.02:2)
plot(x, x -> cdf(ber, x); label="")
title!("cdf of Bernoulli(0.3)")
plot!(; ytick = 0:0.1:1)
scatter!(0:1, x -> cdf(ber, x - 0.02) ; label="", c=:white, ms=3)
scatter!(0:1, x -> cdf(ber, x) ; label="", c=1, ms=3)
```

Out[54]:

5.20 カテゴリカル分布の累積分布函数

カテゴリカル分布 Categorical(0.1, 0.2, 0.3, 0.4) の累積分布函数は次になる:

$$F(x) = \begin{cases} 0 & (x < 1) \\ 0.1 & (1 \le x < 2) \\ 0.1 + 0.2 & (2 \le x < 3) \\ 0.1 + 0.2 + 0.3 & (3 \le x < 4) \\ 1 & (0.4 \le x) \end{cases}$$

In [55]:
```
categ = Categorical(0.1, 0.2, 0.3, 0.4)
x = vcat([-0.5:0.02:0.98; NaN], ([k:0.02:k+0.98; NaN] for k in 1:3)..., 4:0.02:5.5)
plot(x, x -> cdf(categ, x); label="")
title!("cdf of Categorical(0.1, 0.2, 0.3, 0.4)")
plot!(; xtick = -1:6, ytick = 0:0.1:1)
scatter!(1:4, x -> cdf(categ, x - 0.02) ; label="", c=:white, ms=3)
scatter!(1:4, x -> cdf(categ, x) ; label="", c=1, ms=3)
```

Out[55]:

5.21 二項分布の累積分布函数のグラフ

In [56]:

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

Out[56]:

5.22 負の二項分布の累積分布函数のグラフ

In [57]:

```
negbin = NegativeBinomial(4, 0.5)
x = vcat([-2:0.02:-0.02; NaN], ([k:0.02:k+0.98; NaN] for k in 0:13)...)
plot(x, x -> cdf(negbin, x); label="")
title!("cdf of NegativeBinomial(4, 0.5)")
plot!(; xtick = -2:14, ytick = 0:0.1:1)
scatter!(0:14, x -> cdf(negbin, x - 0.02) ; label="", c=:white, ms=3)
scatter!(0:14, x -> cdf(negbin, x) ; label="", c=1, ms=3)
```

Out[57]:

5.23 一様分布の累積分布函数

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

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

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

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

になる.

In [58]:
```
plot(x -> cdf(Uniform(-1, 2), x), -2.5, 3.5; label="")
title!("cdf of Uniform(-1, 2)")
yticks!(0:0.1:1)
```

Out[58]:

5.24 標準正規分布の累積分布函数と分位点函数

誤差函数が

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

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

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

In [59]:
```
plot(z -> cdf(Normal(), z), -4, 4;
    label="cdf of Normal()", legend=:topleft)
plot!(z -> (1 + erf(z/√2))/2, -4, 4;
    label="(1 + erf(z/√2))/2", ls=:dash)
plot!(; xtick=-5:5, ytick = 0:0.1:1,
    xlabel="z", ylabel="p")
plot!(; size=(300, 300))
```

Out[59]:

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

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

と書ける. 標準正規分布の分位点函数は統計学で非常によく使われる. 例えば

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

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

In [60]:
```
quantile(Normal(), 0.975)
```

Out[60]:
```
1.9599639845400576
```

In [61]:
```
plot(p -> quantile(Normal(), p), 0, 1;
    label="quantile function of Normal()", legend=:topleft)
plot!(p -> √2*erfinv(2p-1), 0, 1;
    label="z = √2 erfinv(2p-1)", ls=:dash)
plot!(; xtick = 0:0.1:1, ytick=-5:5,
    xlabel="p", ylabel="z", ylim=(-4.2, 4.2))
plot!(; size=(300, 300))
```

Out[61]:

5.25 問題: 標準正規分布の四分位数

標準正規分布の累積分布函数と分位点函数はそれぞれ

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

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

解答例: $Q(0.25) \approx -0.6745$, $Q(0.75) \approx 0.6745$. **解答終**

In [62]:
```
quantile(Normal(), 0.25), quantile(Normal(), 0.75)
```

Out[62]:
```
(-0.6744897501960818, 0.6744897501960818)
```

WolframAlpha (https://www.wolframalpha.com/) を使って計算するにはこう入力すればよい:

- √2 erfinv(2 0.25 - 1) (https://www.wolframalpha.com/input?i=%E2%88%9A2+erfinv%282+0.25+-+1%29)
- √2 erfinv(2 0.75 - 1) (https://www.wolframalpha.com/input?i=%E2%88%9A2+erfinv%282+0.75+-+1%29)

WolframAlpha にはこう聞いても答えを教えてくれる:

- quantile of standard normal distribution (https://www.wolframalpha.com/input?i=quantile+of+standard+normal+distribution)

5.26 問題: 連続分布の累積分布函数は一様分布に従う.

X は連続分布に従う確率変数であるとし, その累積分布函数を $F(x) = P(X \le x)$ と書く. このとき, 累積分布函数 $F(x)$ に確率変数 X を代入してできる確率変数 $U = F(x)$ は一様分布 $\mathrm{Uniform}(0,1)$ に従うことを示せ.

注意: このことから, T を一様分布 $\mathrm{Uniform}(0,1)$ に従う確率変数とするとき, X が従う分布の分位点函数 $Q(t) = F^{-1}(t)$ に T を代入して得られる確率変数 $Q(T) = F^{-1}(T)$ は X と同じ連続分布に従うことがわかる. (後でコンピュータで数値的にこのことを確認する.)

解答例: X は連続分布に従うと仮定していたので, X は確率密度函数 $p(x) = F'(x)$ を持つ. 累積分布函数の値は0以上1以下なので, ゆえに,

$$E[g(U)] = E[g(F(X))] = \int_{\mathbb{R}} g(F(x))p(x)\,dx = \int_{\mathbb{R}} g(F(x))\,dF(x) = \int_0^1 g(u)\,du.$$

これで $U = F(X)$ が一様分布 Uniform(0, 1) に従うことがわかった.

解答終

5.27 分位点函数への一様乱数の代入で元の分布の乱数を作れること

In [63]:
```
function plot_quantile_uniform(dist; L=10^6, tlim=(0, 1), title="", kwargs...)
    T = rand(L) # 一様乱数を大量に生成
    X = quantile.(dist, T) # 分位点函数で一様乱数を変換
    xlim = quantile.(Ref(X), tlim) .+ (-0.1, 0.1)
    X = X[first(xlim) .< X .< last(xlim)]
    if dist isa ContinuousUnivariateDistribution
        stephist(X; norm=true, label="")
        plot!(dist, xlim...; ls=:dash, label="", xlim)
    else
        bin = range(round.(extrema(X)) .+ (-0.5, 0.5)...; step=1)
        stephist(X; norm=true, bin, label="")
        plot!(x -> pdf(dist, round(x)), xlim...; ls=:dash, label="", xlim)
    end
    title!(title; kwargs...)
end
```

Out[63]:
```
plot_quantile_uniform (generic function with 1 method)
```

In [64]:
```
plot(plot_quantile_uniform(Normal(); title="Normal(0, 1)"),
    plot_quantile_uniform(Gamma(2, 3); title="Gamma(2, 3)", tlim=(0, 0.995));
    size=(800, 250))
```

Out[64]:

In [65]:
```
plot(plot_quantile_uniform(InverseGamma(2, 3); title="InverseGamma(2, 3)", tlim=(0, 0.98)),
    plot_quantile_uniform(LogNormal(); title="LogNormal(0, 1)", tlim=(0, 0.98));
    size=(800, 250))
```

Out[65]:

In [66]:
```
plot(plot_quantile_uniform(Beta(0.5, 0.5); title="Beta(0.5, 0.5)", xlim=(-0.1, 1.1), ylim=(-0.
    plot_quantile_uniform(Beta(10, 30); title="Beta(10, 30)", tlim=(0.005, 0.995));
    size=(800, 250))
```

Out[66]:

In [67]:
```
plot(plot_quantile_uniform(Binomial(10, 0.3), title="Binomial(10, 0.3)"),
    plot_quantile_uniform(Poisson(5); title="Poisson(5)");
    size=(800, 250))
```

Out[67]:

5.28 一様分布の期待値と分散

区間 (a, b) 上の確率密度函数

$$p(x) = \frac{1}{b-a} \quad (a < x < b)$$

によって与えられる一様分布

$$\mathrm{Uniform}(a, b)$$

を考え, X はこれに従う確率変数であるとする.

このとき, X の期待値(平均) μ と分散 σ^2 は以下のように計算される:

$$\mu = E[X] = \int_a^b x p(x)\,dx = \frac{1}{b-a}\int_a^b x\,dx = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

$$E[X^2] = \int_a^b x^2 p(x)\,dx = \frac{1}{b-a}\int_a^b x^2\,dx = \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}.$$

$$\mathrm{var}(X) = E[X^2] - E[X]^2 = \frac{4(a^2+ab+b^2)-3(a+b)^2}{12} = \frac{a^2-2ab+b^2}{12} = \frac{(b-a)^2}{12}.$$

In [68]:
```
function E(f, dist::ContinuousUnivariateDistribution)
    # quadgk(f, a, b)[1] は数値積分してくれる函数
    quadgk(x -> f(x) * pdf(dist, x), extrema(dist)...)[1]
end
```

Out[68]: E (generic function with 2 methods)

In [69]:
```
mean(Uniform(0, 1//0)), var(Uniform(0, 1//1))
```
Out[69]: (1//0, 1//12)

In [70]:
```
E(x -> x, Uniform(0, 1)), E(x -> x^2, Uniform(0, 1)) - E(x -> x, Uniform(0, 1))^2
```
Out[70]: (0.5, 0.08333333333333331)

In [71]:
```
mean(Uniform(0, 5//1)), var(Uniform(0, 5//1))
```
Out[71]: (5//2, 25//12)

In [72]:
```
E(x -> x, Uniform(0, 5)), E(x -> x^2, Uniform(0, 5)) - E(x -> x, Uniform(0, 5))^2
```
Out[72]: (2.5, 2.0833333333333357)

5.29 標準正規分布の期待値と分散

Z は標準正規分布に従う確率変数であるとする. このとき Z は確率密度函数

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

を持つ. このとき, $ze^{-z^2/2}$ が奇函数であり, 奇函数の $\mathbb{R}$ 全体での積分は 0 になるので,

$$E[Z] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-z^2/2}\, dz = 0.$$

$z^2 e^{-z^2/2}$ が偶函数であることと, $z = x^{1/2}$ と変数変換すると, $dz = x^{-1/2}\, dx/2$ となることを使うと,

$$\begin{aligned}\mathrm{var}(Z) = E[Z^2] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2}\, dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z^2 e^{-z^2/2}\, dz \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} xe^{-x/2} \frac{x^{-1/2}\, dx}{2} = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} x^{1/2-1} e^{-x/2}\, dx = \frac{2^{1/2}\Gamma(1/2)}{\sqrt{2\pi}} = 1.\end{aligned}$$

ここで, 5番目の等号では次の公式を使った:

$$\int_0^{\infty} x^{\alpha-1} e^{-x/\theta}\, dx = \theta^{\alpha}\Gamma(\alpha) \quad (\alpha, \theta > 0).$$

最後の等号では $\Gamma(1/2) = \sqrt{\pi}$ を使った. ガンマ函数はこれらの形で使われることが多い. これらの辺の公式については別のノートで解説することにする.

平均 μ と分散 σ^2 を持つ正規分布に従う確率変数 X は $X = \mu + \sigma Z$ と作れる.

注意: $\mathrm{var}(Z) = 1$ は以下のようにしても示される. Gauss積分の公式

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

において, $a > 0$ を任意に取って, $x = \sqrt{a}\, z$ とおくと,

$$\sqrt{a} \int_{-\infty}^{\infty} e^{-az^2}\, dz = \sqrt{\pi}, \qquad \therefore \int_{-\infty}^{\infty} e^{-az^2}\, dz = \sqrt{\pi}\, a^{-1/2}.$$

両辺を a で微分してから -1 倍して(微分と積分の順序は交換できる), $a = 1/2$ とおくと,

$$\int_{-\infty}^{\infty} z^2 e^{-az^2}\, dz = \frac{1}{2}\sqrt{\pi}\, a^{-3/2}, \qquad \int_{-\infty}^{\infty} z^2 e^{-z^2/2}\, dz = \sqrt{2\pi}.$$

ゆえに, $\mathrm{var}(Z) = E[Z^2] = (1/\sqrt{2\pi}) \int_{-\infty}^{\infty} z^2 e^{-z^2/2}\, dz = 1$.

In [73]:
```
mean(Normal(0, 1//1)) ▷ Int, var(Normal(0, 1//1)) ▷ Int
```
Out[73]: (0, 1)

In [74]:
```
E(x → x, Normal(0, 1)), E(x → x^2, Normal(0, 1)) - E(x → x, Normal(0, 1))^2
```
Out[74]:
```
(4.2013717612707503e-16, 1.0)
```

In [75]:
```
mean(Normal(2, 3//1)) ▷ Int, var(Normal(2, 3//1)) ▷ Int
```
Out[75]:
```
(2, 9)
```

In [76]:
```
E(x → x, Normal(2, 3)), E(x → x^2, Normal(2, 3)) - E(x → x, Normal(2, 3))^2
```
Out[76]:
```
(1.999999999999999, 9.0)
```

標準正規分布の歪度(わいど, skewness):

In [77]:
```
skewness(Normal(0, 1//1)) ▷ Int
```
Out[77]:
```
0
```

In [78]:
```
E(z → z^3, Normal())
```
Out[78]:
```
-5.974359421244178e-16
```

標準正規分布の尖度(せんど, kurtosis)に 3 を足した値:

In [79]:
```
kurtosis(Normal(0, 1//1)) + 3 ▷ Int
```
Out[79]:
```
3
```

In [80]:
```
E(z → z^4, Normal())
```
Out[80]:
```
2.999999999999997
```

質問: 尖度の定義でどうして 3 を引いておくか?

答え: 正規分布の尖度を 0 にするためには4次のモーメントから 3 を引いておく必要がある.

In [81]:
```
@syms t κ3 κ4 κ5 μ3 μ4 μ5
series(exp(t^2//2 + κ3 * t^3//6 + κ4 * t^4//24 + κ5 * t^5//120), t)
```
Out[81]:

$$1 + \frac{t^2}{2} + \frac{t^3\kappa_3}{6} + t^4\left(\frac{\kappa_4}{24} + \frac{1}{8}\right) + t^5\left(\frac{\kappa_3}{12} + \frac{\kappa_5}{120}\right) + O\left(t^6\right)$$

In [82]:
```
series(log(1 + t^2//2 + μ3 * t^3//6 + μ4 * t^4//24 + μ5 * t^5//120), t)
```
Out[82]:

$$\frac{t^2}{2} + \frac{t^3\mu_3}{6} + t^4\left(\frac{\mu_4}{24} - \frac{1}{8}\right) + t^5\left(-\frac{\mu_3}{12} + \frac{\mu_5}{120}\right) + O\left(t^6\right)$$

$Z \sim \mathrm{Normal}()$ (標準正規分布)のとき, $E[Z^{2k}] = 1 \cdot 3 \cdots (2k-1)$ となる.

In [83]:
```
[(prod(1:2:2k-1), E(z → z^(2k), Normal())) for k in 1:7]
```
Out[83]:
```
7-element Vector{Tuple{Int64, Float64}}:
 (1, 1.0)
 (3, 2.999999999999997)
 (15, 14.999999999999998)
 (105, 105.00000000000004)
 (945, 945.0000000000093)
 (10395, 10395.000000000617)
 (135135, 135135.0000000145)
```

5.30 一様乱数の生成の繰り返しの分布

0 と 1 のあいだにある n 個の実数達 $t_1, \ldots, t_n$ の組 $(t_1, \ldots, t_n)$ に関する確率密度函数 $p(t_1, \ldots, t_n)$ を一様分布 Uniform(0, 1) の確率密度函数

$$p(t) = \begin{cases} 1 & (0 < t < 1) \\ 0 & (\text{otherwise}) \end{cases}$$

の積で次のように定める:

$$p(t_1, \dots, t_n) = p(t_1) \cdots p(t_n) = \begin{cases} 1 & (0 < t_1, \dots, t_n < 1) \\ 0 & (\text{otherwise}) \end{cases}$$

これが実際に n 変数の確率密度函数になっていることは次のように確認できる:

$$\begin{aligned} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(t_1, \dots, t_n)\, dt_1 \cdots dt_n &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(t_1) \cdots p(t_n)\, dt_1 \cdots dt_n \\ &= \int_{-\infty}^{\infty} p(t_1)\, dt_1 \cdots \int_{-\infty}^{\infty} p(t_n)\, dt_n = 1 \cdots 1 = 1. \end{aligned}$$

この確率密度函数 $p(t_1, \dots, t_n)$ で定まる $\mathbb{R}^n$ 上の連続確率分布(n 変量の連続分布)を **一様乱数の生成の繰り返しの分布** と呼び, $\mathrm{Uniform}(0, 1)^n$ と表す

この分布はコンピュータにおける一様乱数函数 `rand()` を n 回実行して得られる長さ n の 0 から 1 のあいだの乱数の列の分布のモデル化になっている.

この分布は後で別のノートで説明するように, 一様分布 $\mathrm{Uniform}(0, 1)$ のサイズ n の **標本分布** (sample distribution)もしくは各々が一様分布 $\mathrm{Uniform}(0, 1)$ に従う n 個の **独立同分布確率変数達** (independent and identically distributed random variables)が従う同時確率分布になっている.

$(T_1, \dots, T_n)$ がこの分布に従う確率変数の組であるとき, その函数 $f(T_1, \dots, T_n)$ の期待値は

$$E[f(T_1, \dots, T_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \dots, t_n) p(t_1, \dots, t_n)\, dt_1 \cdots dt_n$$

と定義される. $f(T_1, \dots, T_n)$ が $f(T_1, \dots, T_n) = f_1(T_1) \dots f_n(T_n)$ と各々の T_i のみの函数の積になっている場合には

$$\begin{aligned} E[f_1(T_1) \cdots f_n(T_n)] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_1(t_1) \cdots f_n(t_n) p(t_1) \cdots p(t_n)\, dt_1 \cdots dt_n \\ &= \int_{-\infty}^{\infty} f_1(t_1) p(t_1)\, dt_1 \cdots \int_{-\infty}^{\infty} f_n(t_n) p(t_n)\, dt_n = E[f_1(T_1)] \cdots E[f_n(T_n)] \end{aligned}$$

が成立する. このように確率変数の積の期待値が各々の期待値の積に等しくなるとき, それらの確率変数は **独立** (independent)であるという.

以上のようなものは, $\mathrm{Uniform}(0, 1)$ に限らず, 任意の確率分布について考えることができる.

例えば, Bernoulli試行の確率分布はBernoulli分布の標本分布になっている.

In [ ]:

```
1
```

Gauss積分, ガンマ函数, ベータ函数

- 黒木玄

- 2022-04-10〜2022-05-21

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)

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 Gauss積分
  - 1.1 Gauss積分の定義
  - 1.2 問題: 正規分布での確率の総和
  - 1.3 正規分布の定義
  - 1.4 $p(x|\theta)$ という記号法について
  - 1.5 問題: Gauss積分の拡張とガンマ函数の関係
- 2 ガンマ函数
  - 2.1 ガンマ函数の定義
  - 2.2 問題: ガンマ函数はこの形でよく使われる
  - 2.3 ガンマ分布の定義
  - 2.4 問題: ガンマ分布の期待値と分散
  - 2.5 指数分布の定義
  - 2.6 Laplace分布の定義
  - 2.7 問題: Laplace分布の期待値と分散
  - 2.8 問題: 逆ガンマ分布の定義
  - 2.9 問題: 逆ガンマ分布の期待値と分散
  - 2.10 χ^2分布の定義
  - 2.11 χ^2分布と多変量正規分布の関係
  - 2.12 問題: 自由度 n のχ^2分布の n 個の標準正規分布による表示
  - 2.13 Weibull分布の定義
  - 2.14 問題: Weibull分布の期待値と分散
  - 2.15 ガンマ函数の函数等式
  - 2.16 ガンマ函数の函数等式のよく見る使い方
  - 2.17 階乗のガンマ函数表示
  - 2.18 問題: 正の半整数でのガンマ函数の値
  - 2.19 問題: 標準正規分布の偶数次のモーメント
  - 2.20 問題: 標準正規分布のモーメント母函数とキュムラント母函数
  - 2.21 問題: n-1次元単位球面の面積のガンマ函数表示
  - 2.22 $n-1$ 次元単位球面上の座標を使った面積の計算
  - 2.23 問題: n次元単位球体の面積のガンマ函数表示
  - 2.24 Stirlingの公式
  - 2.25 問題: 小さな n でもStirlingの公式による $n!$ の近似精度がそう悪くないこと
  - 2.26 Stirlingの公式とガンマ分布の正規分布近似(中心極限定理)の関係
- 3 ベータ函数
  - 3.1 ベータ函数の定義
  - 3.2 ベータ分布の定義
  - 3.3 問題: ベータ分布の期待値と分散
  - 3.4 ベータ函数は α と β を交換しても不変
  - 3.5 問題: ベータ分布の三角函数を使った表示
  - 3.6 ベータ函数の極限としてガンマ函数を表せること
  - 3.7 問題: ベータ分布の極限でガンマ分布が得られること
  - 3.8 ベータ函数の別の基本的表示

3.9 問題: t 分布の基礎になる公式
3.10 t 分布の定義
3.11 t 分布の期待値と分散
3.12 問題: t 分布の自由度無限大の極限で標準正規分布が得られること
3.13 t 分布と標準正規分布を同時プロット
4 ベータ函数のガンマ函数表示
4.1 積分 $J[f]$ を導入とベータ函数のガンマ函数表示の証明
4.2 積分 $J[f]$ に関する公式の証明
4.3 問題: ガンマ分布とベータ分布の関係 (ベータ分布のガンマ分布表示)
4.4 問題: ガンマ分布の再生成
4.5 積分 $K[f]$ を導入とベータ函数のガンマ函数表示の別証明
4.6 積分 $K[g]$ に関する公式の証明
4.7 ベータプライム分布の定義
4.8 F 分布の定義
4.9 F 分布とベータ分布やガンマ分布やχ²分布との関係
4.10 問題: ガンマ函数のGaussの乗法公式
4.11 問題: $\sin$ の無限積表示
4.12 問題: Wallisの公式
4.13 問題: ガンマ函数の相反公式(reflection formula)
4.14 ディガンマ函数, トリガンマ函数, ポリガンマ函数, Hurwitzのゼータ函数
5 Dirichlet積分=多変数ベータ函数
5.1 Dirichlet積分=多変数ベータ函数の定義とそのガンマ函数表示
5.2 Dirichlet分布の定義
5.3 多変数ベータ函数のガンマ函数表示の証明
5.4 ガンマ分布とDirichlet分布の関係
5.5 問題: Dirichlet (1839)の公式の証明
5.6 Dirichlet氏の1839年の論文との比較

In [1]:

```
ENV["LINES"], ENV["COLUMNS"] = 100, 100
using BenchmarkTools
using Distributions
using Printf
using QuadGK
using Random
Random.seed!(4649373)
using Roots
using SpecialFunctions
using StaticArrays
using StatsBase
using StatsFuns
using StatsPlots
default(fmt = :png, titlefontsize = 10, size = (400, 250))
using SymPy
```

In [2]:

```
# Override the Base.show definition of SymPy.jl:
# https://github.com/JuliaPy/SymPy.jl/blob/29c5bfd1d10ac53014fa7fef468bc8deccadc2fc/src/types.

@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::SymbolicObject)
    print(io, as_markdown("\\displaystyle " * sympy.latex(x, mode="plain", fold_short_frac=fal
end
@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::AbstractArray{Sym})
    function toeqnarray(x::Vector{Sym})
        a = join(["\\displaystyle " * sympy.latex(x[i]) for i in 1:length(x)], "\\\\")
        """\\left[ \\begin{array}{r}$a\\end{array} \\right]"""
    end
    function toeqnarray(x::AbstractArray{Sym,2})
        sz = size(x)
        a = join([join("\\displaystyle " .* map(sympy.latex, x[i,:]), "&") for i in 1:sz[1]],
        "\\left[ \\begin{array}{" * repeat("r",sz[2]) * "}" * a * "\\end{array}\\right]"
    end
    print(io, as_markdown(toeqnarray(x)))
end
```

1 Gauss積分

1.1 Gauss積分の定義

次の積分 I を **Gauss積分** と呼ぶ:

$$I = \int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

この積分の値が実際に円周率の平方根になることを以下で示そう.

Gauss積分の2乗

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

は次のグラフで示された高さ 1 の山の体積になる.

In [3]:

```
x = y = range(-3, 3, 61)
wireframe(x, y, (x, y) → exp(-(x^2+y^2)); camera=(30, 70), colorbar=false)
plot!(; zlim=(-0.02, 1.02), size=(600, 400))
plot!(; xlabel="x", ylabel="y", zlabel="z")
title!("z = exp(-(x²+y²))"; titlefontsize=12)
```

Out[3]:

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

In [4]:

```
plot(r → exp(-r^2), 0, 3; label="", title="z = exp(-r²) ⇔ r = √(-log(z))")
plot!(; xlabel="r = √(x² + y²)", ylabel="z")
plot!(; ytick=0:0.1:1)
z = 0.3
plot!([(0, z), (√(-log(z)), z)]; label="", lw=2)
plot!(; xlim=(0, 3))
annotate!(1.1, 0.2, ("√(-log(0.3)) = r", 10, :right))
annotate!(1.2, 0.3, ("← 0.3 = exp(-r²)", 10, :left))
```

Out[4]:

この図より, 高さ $0 < z \leq 1$ での山の切断面は, 半径 $\sqrt{-\log z}$ の円盤になる. その円盤の面積 $\pi(-\log z)$ を $0 < z \leq 1$ について積分すれば上の山の体積(= Gauss積分の2乗 I^2 の値)が得られる:

$$I^2 = \int_0^1 \pi(-\log z)\,dz = -\pi[z\log z - z]_0^1 = -\pi(-1) = \pi.$$

ゆえに

$$I = \int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}.$$

1.2 問題: 正規分布での確率の総和

$\mu, \sigma \in \mathbb{R}$, $\sigma > 0$ であると仮定する. 次を示せ:

$$\frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = 1.$$

解答例: 左辺の積分部分が $\sqrt{2\pi\sigma^2}$ に等しいことを示せばよい.

$$\int_{-\infty}^{\infty} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \int_{-\infty}^{\infty} \exp\left(-\frac{y^2}{2\sigma^2}\right) dy = \sqrt{2\sigma^2}\int_{-\infty}^{\infty} e^{-z^2}\,dz = \sqrt{2\sigma^2}\sqrt{\pi} = \sqrt{2\pi\sigma^2}$$

最初の等号で $x = y + \mu$ とおき, 2つめの等号で $y = \sqrt{2\sigma^2}\,z$ とおき, 3つめの等号でGauss積分を使った.

解答終

1.3 正規分布の定義

$\mu, \sigma \in \mathbb{R}$, $\sigma > 0$ であると仮定する.

次の確率密度函数によって定義される連続分布を平均 μ, 分散 σ^2 (もしくは標準偏差 σ)の **正規分布** (normal distribution)と呼び $\mathrm{Normal}(\mu, \sigma)$ と表す:

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

特に $\mu = 0$, $\sigma = 1$ の正規分布を **標準正規分布** (standard normal distribution)と呼び, $\mathrm{Normal}()$ と表すことにする:

$$p(x|0, 1) = \frac{1}{\sqrt{2\pi}} e^{-x^2}.$$

1.4 $p(x|\theta)$ という記号法について

パラメータ θ を持つ x に関する確率密度函数を $p(x|\theta)$ と書く.

この記号法によって, 縦棒 | の左側の変数 x について確率密度函数になっているが, 縦棒 | の右側の変数 θ については単なる函数で確率密度函数ではないことを表す.

パラメータ θ を持つ x に関する確率質量函数 $P(x|\theta)$ についても同様である.

1.5 問題: Gauss積分の拡張とガンマ函数の関係

$\alpha \in \mathbb{R}$, $\alpha > 0$ であると仮定する. 次を示せ:

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

解答例: $x = \sqrt{y}$ とおくと, $dx/x = d\log x = (1/2)d\log y = dy/(2y)$ なので,

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

最初の等号で被積分函数が偶函数であることを用い, 第2の等号で $x = \sqrt{y}$ とおいた.

解答終

この問題では単なる $\exp(-x^2)$ の積分であるGauss積分をそれに $|x|$ のべきをかけたものの積分に拡張した. 実はこの拡張は次の節で扱うガンマ函数の定義そのものになっている. すなわち, ガンマ函数はGauss積分の拡張の一種になっている.

正規分布の確率密度函数は釣鐘型(bell-shaped)になる.

In [5]:
```
plot(Normal(-1, 2); label="", title="Normal(-1, 2)", xtick = -100:100)
```

Out[5]:

2 ガンマ函数

2.1 ガンマ函数の定義

$\alpha > 0$ について, ガンマ函数 $\Gamma(\alpha)$ を次のように定義する:

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

前節の問題の結果より, これは

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

とGauss積分の拡張形でも表される. 特に

$$\Gamma(1/2) = \int_{-\infty}^\infty e^{-x^2}\, dx = \sqrt{\pi}.$$

このようにガンマ函数 $\Gamma(\alpha)$ はGauss積分 $\Gamma(1/2)$ の拡張になっている.

$\Gamma(1) = 1$ となることも容易に示せる:

$$\Gamma(1) = \int_0^\infty e^{-x}\, dx = [-e^{-x}]_0^\infty = -(-1) = 1.$$

注意: Gauss積分は正規分布の定義の基礎になっている. ここで扱っているようなGauss積分の拡張形は平均が0の正規分布の絶対モーメントになっており, 正規分布に関係した計算には自然にガンマ函数が出て来る.

2.2 問題: ガンマ函数はこの形でよく使われる

$\alpha, \theta > 0$ について次の公式が成立することを示せ:

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

解答例: $x = \theta y$ とおくと, $dx/x = \theta\, dy/(\theta y) = dy/y$ なので,

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

2つめの等号で $x = \theta y$ を代入して, 4つめの等号でガンマ函数の定義を使った.

解答終

ガンマ函数は主にこの形で使う と覚えておくと便利だと思う.

例えば, この問題の結果から,

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

という公式が得られる. これは「ガンマ分布において確率の総和が1になる」という意味を持つ.

2.3 ガンマ分布の定義

$\alpha, \theta > 0$ であると仮定する.

次の確率密度函数によって定義される連続分布を**形状パラメータ** α, **スケールパラメータ** θ の **ガンマ分布** (Gamma distribution)と呼び $\mathrm{Gamma}(\alpha, \theta)$ と表す:

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

ガンマ分布は正規分布から派生して出て来る分布の中で非常に基本的なものになっている.

負の二項分布の連続極限としてガンマ分布を出すことによって, ガンマ分布には「1回起こるまでにかかる時間の期待値が θ であるようなイベントが α 回起こるまでにかかる時間の分布」だと解釈できる. この点については後で「確率分布達の解釈」のノートで説明する.

スケールパラメータを変えてもガンマ分布の確率密度函数 $p(x|\alpha, \theta)$ のグラフの形状は変化せず, 単に x 軸方向のスケールが変わって, 密度の値もそれに応じてスケールされるだけになる. ゆえに, ガンマ分布の形状を知りたければ, 形状パラメータ α のみを動かして, ガンマ分布のグラフを描けばよい.

$\alpha = 1/2$ のガンマ分布の確率密度函数は次の形になる.

In [6]:
```
α, θ = 1/2, 1
plot(Gamma(α, θ), 0.01, 2; label="", title="Gamma($α, $θ)")
```

Out[6]:

$\alpha = 3$ のガンマ分布の確率密度函数は次の形になる.

In [7]:
```
α, θ = 3, 1
plot(Gamma(α, θ); label="", title="Gamma($α, $θ)")
```

Out[7]:

$\alpha = 30$ のガンマ分布の確率密度函数は次の形になる.

In [8]:
```
α, θ = 30, 1
plot(Gamma(α, θ); label="", title="Gamma($α, $θ)")
```

Out[8]:

形状パラメータ α が大きなガンマ分布は正規分布で近似される.

In [9]:
```
function plot_gamma_normal(α, θ; kwargs...)
    gam = Gamma(α, θ)
    μ, σ = mean(gam), std(gam)
    x = range(max(0, μ - 5σ), μ + 5σ, 400)
    P1 = plot(x, x -> pdf(gam, x); label="Gamma(α, θ)", kwargs...)
    α > 2 && plot!(x, x -> pdf(Normal(μ, σ), x); label="normal approx.", ls=:dash)
    title!("α=$α, θ=$θ")
end
```

Out[9]: plot_gamma_normal (generic function with 1 method)

In [10]:
```
plot(plot_gamma_normal(0.5, 1), plot_gamma_normal(1, 1); size=(800, 250))
```

Out[10]:

In [11]:
```
plot(plot_gamma_normal(1.5, 1), plot_gamma_normal(2, 1); size=(800, 250))
```

Out[11]:

In [12]:
```
plot(plot_gamma_normal(5, 1), plot_gamma_normal(15, 1); size=(800, 250))
```

Out[12]:

In [13]:
```
plot(plot_gamma_normal(50, 1), plot_gamma_normal(150, 1); size=(800, 250))
```

Out[13]:

ガンマ分布の確率密度函数は $(\alpha-1)\theta$ で最大になることを示せる. それに対して, それを近似する正規分布の期待値はもとのガンマ分布の期待値 $\alpha\theta$ に等しくなるので, $\alpha=30, \theta=1$ の場合にはその分だけずれが生じている.

In [14]:
```
plot_gamma_normal(300, 1)
```

Out[14]:

注意: 形状パラメータ α が大きなガンマ分布は正規分布でよく近似されることは本質的に階乗に関するStirlingの(近似)公式に同等である.

2.4 問題: ガンマ分布の期待値と分散

分布 $\mathrm{Gamma}(\alpha,\theta)$ の期待値と分散がそれぞれ $\alpha\theta$, $\alpha\theta^2$ となることを示せ.

解答例: $X\sim\mathrm{Gamma}(\alpha,\theta)$ と仮定する. $k=0,1,2,\ldots$ について

$$\begin{aligned}\Gamma(\alpha+k) &= (\alpha+k-1)\Gamma(\alpha+k-1) = \cdots = (\alpha+k-1)\cdots(\alpha+1)\alpha\Gamma(\alpha)\\ &= \alpha(\alpha+1)\cdots(\alpha+k-1)\Gamma(\alpha)\end{aligned}$$

なので

$$E[X^k] = \frac{1}{\theta^\alpha \Gamma(\alpha)} \int_0^\infty e^{-x/\theta} x^{\alpha+k-1}\, dx = \frac{\theta^{\alpha+k}\Gamma(\alpha+k)}{\theta^\alpha \Gamma(\alpha)} = \theta^k \alpha(\alpha+1)\cdots(\alpha+k-1).$$

ゆえに,

$$E[X] = \theta\alpha = \alpha\theta, \quad \mathrm{var}(X) = E[X^2] - E[X]^2 = \theta^2\alpha(\alpha+1) - \theta^2\alpha^2 = \theta^2\alpha = \alpha\theta^2.$$

解答終

2.5 指数分布の定義

$\alpha = 1$ のガンマ分布の密度函数は

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

の形になる. この形の確率密度函数で定義される連続確率分布を **指数分布** (exponential distribution)と呼び, $\mathrm{Exponential}(\theta)$ と表す:

$$\mathrm{Exponential}(\theta) = \mathrm{Gamma}(1,\theta).$$

ガンマ分布の期待値と分散に関する結果より, 指数分布の期待値と分散はそれぞれ θ, θ^2 になる.

指数分布は「1回起こるまでにかかる時間の期待値が θ であるようなイベントが起こるまでにかかる時間の分布」だと解釈可能である. この点については後で「確率分布達の解釈」のノートで説明する.

注意: 家計が保有する金融資産の分布は指数分布に近い形の分布になり, そのことには統計力学的な解釈がある.

指数分布の確率密度函数は次の形になる.

In [15]:
```
plot_exp(θ) = plot(Exponential(θ); label="", title="Exponential($θ)")
plot(plot_exp(1), plot_exp(3), plot_exp(10); size=(800, 200), layout=(1, 3))
```

Out[15]:

指数分布のグラフはどれも同じ形状になる. 指数分布 $\mathrm{Exponential}(\theta)$ の期待値と分散はそれぞれ θ, θ^2 になる.

2.6 Laplace分布の定義

$a \in \mathbb{R}$, $b > 0$ であると仮定する. 確率密度函数

$$p(x|a,b) = \frac{1}{2b}\exp\left(-\frac{|x-a|}{b}\right)$$

が定める $\mathbb{R}$ 上の連続確率分布を **Laplace分布** (ラプラス分布)と呼び,

$$\mathrm{Laplace}(a,b)$$

と書くことにする. Laplace分布は $x = a$ の左右両側に指数分布を並べた分布になっているので, **ダブル指数分布** (double exponential distribution)と呼ばれることもある.

$a = 0$, $b = 1$ のLaplace分布 $\mathrm{Laplace}(0,1)$ を **標準Laplace分布** と呼ぶ.

In [16]:

```
plot_laplace(a, b) = plot(Laplace(a, b); label="", title="Laplace($a,$b)")
plot(plot_laplace(0, 1), plot_laplace(0, 0.5), plot_laplace(2, 3);
    size=(800, 200), layout=(1, 3))
```

Out[16]:

2.7 問題: Laplace分布の期待値と分散

Laplace分布 $\mathrm{Laplace}(a, b)$ に従う確率変数 X について

$$E[X] = a, \quad \mathrm{var}(X) = 2b^2$$

となっていることを示せ.

解答例: Y が標準Laplace分布に従う場合には, $ye^{-|y|}$ が奇函数なので,

$$E[Y] = \frac{1}{2}\int_{-\infty}^{\infty} ye^{-|y|}\,dy = 0$$

となり, さらに,

$$\mathrm{var}(Y) = \frac{1}{2}\int_{-\infty}^{\infty} y^2 e^{-|y|}\,dy = \int_{-\infty}^{\infty} y^2 e^{-y}\,dy = \Gamma(3) = 2! = 2.$$

$X = a + bY$ とおくと, $X \sim \mathrm{Laplace}(a, b)$ となる. このとき,

$$E[X] = a + bE[Y] = a, \quad \mathrm{var}(X) = b^2\,\mathrm{var}(Y) = 2b^2.$$

解答終

2.8 問題: 逆ガンマ分布の定義

X がガンマ分布 $\Gamma(\alpha, 1/\theta)$ に従う確率変数であるとき, その逆数 $Y = 1/X$ について

$$E[f(Y)] = \int_0^{\infty} f(y)\frac{e^{-\theta/y}y^{-\alpha-1}}{(1/\theta)^{\alpha}\Gamma(\alpha)}\,dy$$

を満たすことを示せ. 確率密度函数

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

で定義される連続分布を **逆ガンマ分布** (inverse Gamma distribution)と呼び,

$$\mathrm{InverseGamma}(\alpha, \theta)$$

と表すことにする. 逆ガンマ分布はガンマ分布に従う確率変数の逆数が従う分布のことである.

解答例: 積分変数 x を $x = 1/y$ で置換すると, $dx = -dy/y^2$ より,

$$
\begin{aligned}
E[f(Y)] = E[f(1/X)] &= \int_0^\infty f(1/x)\frac{e^{-\theta x}x^{\alpha-1}}{(1/\theta)^\alpha\Gamma(\alpha)}\,dx \\
&= \int_\infty^0 f(y)\frac{e^{-\theta/y}y^{-\alpha+1}}{(1/\theta)^\alpha\Gamma(\alpha)}\frac{-dy}{y^2} = \int_0^\infty f(y)\frac{e^{-\theta/y}y^{-\alpha-1}}{(1/\theta)^\alpha\Gamma(\alpha)}\,dy.
\end{aligned}
$$

解答終

In [17]:

```
p_inversegamma(α, θ, x) = x ≥ 0 ? pdf(InverseGamma(α, θ), x) : zero(x)
function plot_inversegamma_normal(α, θ; kwargs...)
    invgam = InverseGamma(α, θ)
    if α > 2
        μ, σ = mean(invgam), std(invgam)
        x = range(max(0, μ - 5σ), μ + 5σ, 1000)
    else
        x = range(0, quantile(invgam, 0.75 + 0.1α), 1000)
    end
    P1 = plot(x, x -> p_inversegamma(α, θ, x); label="InvGam(α, θ)", kwargs...)
    α > 2 && plot!(x, x -> pdf(Normal(μ, σ), x); label="normal approx.", ls=:dash)
    title!("α=$α, θ=$θ")
end
```

Out[17]: plot_inversegamma_normal (generic function with 1 method)

In [18]:

```
plot(plot_inversegamma_normal(0.5, 2), plot_inversegamma_normal(1, 1); size=(800, 250))
```

Out[18]:

In [19]:

```
plot(plot_inversegamma_normal(1.5, 1), plot_inversegamma_normal(2, 1); size=(800, 250))
```

Out[19]:

In [20]:
```
plot(plot_inversegamma_normal(5, 1), plot_inversegamma_normal(15, 1); size=(800, 250))
```

Out[20]:

In [21]:
```
plot(plot_inversegamma_normal(50, 1), plot_inversegamma_normal(150, 1); size=(800, 250))
```

Out[21]:

2.9 問題: 逆ガンマ分布の期待値と分散

確率変数 X は $X \sim \mathrm{InverseGamma}(\alpha, \theta)$ を満たしていると仮定する. このとき, 以下が成立することを示せ:

$$\alpha > 1 \implies E[X] = \frac{\theta}{\alpha - 1}$$
$$\alpha > 2 \implies \mathrm{var}(X) = \frac{\theta^2}{(\alpha-1)^2(\alpha-2)}.$$

解答例: $\alpha > 0$ のとき, 逆ガンマ分布の確率の総和が 1 であることより,

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

$\alpha > 1$ のとき,

$$E[X] = \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-\theta/x} x^{-(\alpha-1)-1}\, dx = \frac{\theta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-1)}{\theta^{\alpha-1}} = \frac{\theta}{\alpha-1}.$$

$\alpha > 2$ のとき,

$$E[X^2] = \frac{\theta^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-\theta/x} x^{-(\alpha-2)-1}\, dx = \frac{\theta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha-2)}{\theta^{\alpha-2}} = \frac{\theta^2}{(\alpha-1)(\alpha-2)}.$$

ゆえに

$$\mathrm{var}(X) = E[X^2] - E[X]^2 = \frac{\theta^2}{(\alpha-1)^2(\alpha-2)}.$$

解答終

In [22]:
```
@vars α θ
E_X = θ/(α - 1)
E_X² = θ^2/((α - 1)*(α - 2))
var_X = E_X² - E_X^2 ▷ factor
```

Out[22]:

$$\frac{\theta^2}{(\alpha-2)\,(\alpha-1)^2}$$

2.10 χ²分布の定義

指数分布はガンマ分布 $\Gamma(\alpha, \theta)$ の形状パラメータ α を $\alpha = 1$ に固定して得られる分布であった.

χ²分布 は本質的にガンマ分布 $\Gamma(\alpha, \theta)$ のスケールパラメータ θ を $\theta = 2$ に固定して得られる分布のことである. ただし, $\alpha = \nu/2$ とおいて, ν を **自由度** (degree of freedom, df)と呼ぶ. すなわち, 分布 $\mathrm{Gamma}(\nu/2, 2)$ を自由度 ν の **χ²分布** (カイ二乗分布, chi-squared distribution)と呼び,

$$\mathrm{Chisq}(\nu) = \mathrm{Gamma}(\nu/2, 2)$$

と表す. χ²分布の確率密度函数は次の形になる:

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

ガンマ分布の期待値と分散に関する結果より, 指数分布の期待値と分散はそれぞれ ν, 2ν になる.

ガンマ分布は α が大きなとき正規分布で近似されるので, **自由度 ν が大きなχ²分布は正規分布で近似される.**

In [23]:
```
p_chisq(ν, x) = (ν > 2 || x ≥ 0 || x ≤ -0.1) ? pdf(Chisq(ν), x) : NaN
function plot_chisq_normal(ν; kwargs...)
    x = range(ν - 5√(2ν), ν + 5√(2ν), 1000)
    P1 = plot(x, x -> p_chisq(ν, x); label="Chisq(ν)", kwargs...)
    plot!(x, x -> pdf(Normal(ν, √(2ν)), x); label="Normal(ν,√(2ν))", ls=:dash)
    title!("ν = $ν")
end
```

Out[23]:
```
plot_chisq_normal (generic function with 1 method)
```

In [24]:
```
plot(plot_chisq_normal(1; ylim=(-0.05, 1.55)), plot_chisq_normal(2); size=(800, 250))
```

Out[24]:

In [25]:
```
plot(plot_chisq_normal(3), plot_chisq_normal(4); size=(800, 250))
```

Out[25]:

In [26]:
```
plot(plot_chisq_normal(10), plot_chisq_normal(30); size=(800, 250))
```

Out[26]:

In [27]:
```
plot(plot_chisq_normal(100), plot_chisq_normal(300); size=(800, 250))
```

Out[27]:

2.11 χ²分布と多変量正規分布の関係

ν が正の整数のとき, カイ二乗分布は標準正規分布に従う独立な ν 個の確率変数達の和が従う分布になっている(次の問題を見よ).

この段階では多変量正規分布を導入していないので詳しくは説明できないのだが, より一般に, $(X_1, \dots, X_\nu)$ が確率密度函数が

$$p(x_1, \dots, x_n) = \mathrm{const.}\exp\left(-\frac{1}{2}\sum_{i,j=1}^{\nu} a_{ij}x_ix_j\right)$$

の形の多変量正規分布に従うときに,

$$X^2 = \sum_{i,j=1}^{\nu} a_{ij}X_iX_j$$

の従う分布は自由度 ν のχ²分布になる. このようにχ²分布は正規分布と直接的に関係した分布になっている.

$X_1, \dots, X_\nu$ は各々が標準正規分布に従う独立な確率変数達であることは, それらの同時確率密度函数が

$$p(x_1, \dots, x_\nu) = \frac{1}{(2\pi)^{\nu/2}}\exp\left(-\frac{1}{2}(x_1^2 + \cdots + x_\nu^2)\right)$$

となることと同値なので, この場合には上で述べたように $X^2 = X_1^2 + \cdots + X_\nu^2$ が従う分布が自由度 ν のχ²分布になる. χ²分布の自由度は「独立な正規分布が何個あるか」を意味していると考えられる.

2.12 問題: 自由度 n のχ²分布の n 個の標準正規分布による表示

確率変数達 $X_1, \dots, X_n$ は同時確率密度函数

$$p(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}(x_1^2 + \cdots + x_n^2)\right)$$

を持つと仮定する. すなわち,

$$E[f(X_1,\dots,X_n)] = \int\cdots\int_{\mathbb{R}^n} f(x_1,\dots,x_n)p(x_1,\dots,x_n)\,dx_1\cdots x_n$$

が成立していると仮定する. このとき, 確率変数 $Y = X^2$ を

$$Y = X^2 = X_1^2 + \cdots + X_n^2$$

とおくと, Y は自由度 n のχ²分布に従うことを示せ. すなわち,

$$E[f(Y)] = \frac{1}{2^{n/2}\Gamma(n/2)}\int_0^\infty f(y)e^{y/2}y^{n/2-1}\,dy$$

となることを示せ. この問題の解答では, ずっと下の方で解説した n 次元球面の面積 A_{n-1} に関する

$$A_{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)}$$

という結果とこれを示すために使った方法をに自由に使ってよい.

解答例1: $Y = X_1^2 + \cdots + X_n^2$ の函数の期待値は次のように書ける:

$$E[f(Y)] = \frac{1}{(2\pi)^{n/2}}\int\cdots\int_{\mathbb{R}^n} f(x_1^2+\cdots+x_n^2)p(x_1,\dots,x_n)\exp\left(-\frac{1}{2}(x_1^2+\cdots+x_n^2)\right)dx_1\cdots x_n$$

この積分中の被積分函数は球対称なので, 半径 r に関する積分と単位球面 S^{n-1} 上の積分に分解できる:

$$\begin{aligned}E[f(Y)] &= \frac{1}{(2\pi)^{n/2}}\int_{S^{n-1}}\left(\int_0^\infty f(r^2)e^{-r^2}r^{n-1}\,dr\right)d\omega = \frac{A_{n-1}}{(2\pi)^{n/2}}\int_0^\infty f(r^2)e^{-r^2}r^{n-1}\,dr\\ &= \frac{2\pi^{n/2}}{(2\pi)^{n/2}\Gamma(n/2)}\int_0^\infty f(y)e^{-y}y^{(n-1)/2}\frac{1}{2}y^{-1/2}\,dy = \frac{1}{2^{n/2}\Gamma(n/2)}\int_0^\infty f(y)e^{-y}y^{n/2-1}\,dy.\end{aligned}$$

2つめの等号で単位球面上 S^{n-1} での定数の積分を実行し(単位球面の面積 A_{n-1} 倍になる), 3つめの等号で $r = \sqrt{y}$ とおいた.

解答終

解答例2: ずっと下の方で示すガンマ分布の再生成に帰着する方法でこの問題を解くこともできる. 以下の2つを証明できれば, $X_1, X_2, \dots, X_n$ が独立な確率変数達でかつ各 X_i が標準正規分布に従うとき, $Y = X^2 = X_1^2 + \cdots + X_n^2$ が自由度 n のχ²分布に従うことがわかる:

(1) Z が標準正規分布に従う確率変数であるとき, $Y = Z^2$ は自由度 1 のχ²分布に従う.

(2) 独立な確率変数達 Y_1, Y_2 の各々が自由度 ν_1, ν_2 のχ²分布に従うとき $Y_1 + Y_2$ は自由度 $\nu_1 + \nu_2$ のχ²分布に従う.

この(2)は $\mathrm{Chisq}(\nu) = \mathrm{Gamma}(\nu/2, 2)$ なのでガンマ分布の再生成の特別な場合に過ぎない. そして(1)は以下のように容易に証明される:

$$\begin{aligned}E[f(Y)] &= \int_{-\infty}^\infty f(z^2)\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz = 2\int_0^\infty f(z^2)\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz\\ &= \int_0^\infty f(y)\frac{e^{-y/2}}{\sqrt{2\pi}}y^{-1/2}\,dy = \int_0^\infty f(y)\frac{e^{-y/2}y^{1/2-1}}{2^{1/2}\Gamma(1/2)}\,dy\end{aligned}$$

3つめの等号で $z = \sqrt{y}$ とおき, 最後の等号でGauss積分 $\Gamma(1/2) = \sqrt{2}$ を使った. これで(1)も示された.

解答終

2.13 Weibull分布の定義

T は平均 1 の指数分布 $\mathrm{Exponential}(1)$ に従う確率変数であるとする:

$$E[f(T)] = \int_0^\infty f(t)e^{-t}\,dt.$$

$\alpha, \theta > 0$ について, 確率変数 $W = \theta T^{1/\alpha}$ が従う分布を **形状**(shape)が α で **スケール** (scale)が θ のWeilbull分布と呼び,

$$\mathrm{Weibull}(\alpha, \theta)$$

と表すことにする. このとき, $W = \theta T^{1/\alpha}$ は T について $T = (W/\theta)^\alpha$ と解けるので,

$$E[f(W)] = \int_0^i nftyf(\theta t^{1/\alpha})e^{-t}\,dt = \int_0^\infty f(w)e^{-(w/\theta)^\alpha}\frac{\alpha}{\theta}\left(\frac{w}{\theta}\right)^{\alpha-1}\,dw$$

なので, Weibull分布の確率密度函数は次の形になることがわかる:

$$p(w|\alpha,\theta) = \frac{\alpha}{\theta}\left(\frac{w}{\theta}\right)^{\alpha-1} e^{-(w/\theta)^\alpha} \quad (w > 0)$$

注意: 平均 1 の指数分布は単位時間に平均して 1 回起こるイベントが起こるまでにかかる時間の分布とみなされる(「確率分布達の解釈」のノートで説明する). Weibull分布はイベントが起こる感覚が段々減速したり加速したりするときのイベントが起こるまでにかかる時間の分布だとみなされる.

注意: Weibull分布ではガンマ分布と違って形状パラメータを大きくしても正規分布で近似されるようにならない.

In [28]:
```
function plot_weibull_normal(α, θ; kwargs...)
    weib = Weibull(α, θ)
    μ, σ = mean(weib), std(weib)
    x = range(max(0, μ - 5σ), μ + 5σ, 400)
    plot(x, x -> pdf(weib, x); label="Weibull(α, θ)", kwargs...)
    plot!(x, x -> pdf(Normal(μ, σ), x); label="normal approx.", ls=:dash)
    title!("α=$α, θ=$θ")
end
```

Out[28]: plot_weibull_normal (generic function with 1 method)

In [29]:
```
plot(plot_weibull_normal(0.5, 1), plot_weibull_normal(1, 1); size=(800, 250))
```

Out[29]:

In [30]:
```
plot(plot_weibull_normal(1.5, 1), plot_weibull_normal(2, 1); size=(800, 250))
```

Out[30]:

In [31]:
```
plot(plot_weibull_normal(5, 1), plot_weibull_normal(15, 1); size=(800, 250))
```

Out[31]:

In [32]:
```
plot(plot_weibull_normal(50, 1), plot_weibull_normal(150, 1); size=(800, 250))
```

Out[32]:

2.14 問題: Weibull分布の期待値と分散

確率変数 W が形状 α, スケール θ のWeibull分布に従っているとき, 以下が成立することを示せ:

$$E[W] = \theta\Gamma\left(1+\frac{1}{\alpha}\right), \quad \mathrm{var}(W) = \theta^2\left(\Gamma\left(1+\frac{2}{\alpha}\right) - \Gamma\left(1+\frac{1}{\alpha}\right)^2\right).$$

解答例: $W = \theta T^{1/\alpha}$, $T \sim \mathrm{Exponential}(1)$ と書ける. ゆえに,

$$E[W^m] = E[\theta^m T^{m/\alpha}] = \theta^m \int_0^\infty t^{m/\alpha} e^{-t}\, dt = \theta^m \Gamma\left(1+\frac{m}{\alpha}\right).$$

ゆえに,

$$E[W] = \theta\Gamma\left(1+\frac{1}{\alpha}\right),$$

$$\mathrm{var}(W) = E[W^2] - E[W]^2 = \theta^2\left(\Gamma\left(1+\frac{2}{\alpha}\right) - \Gamma\left(1+\frac{1}{\alpha}\right)^2\right).$$

解答終

2.15 ガンマ函数の函数等式

ガンマ函数の函数等式

$$\Gamma(\alpha+1)=\alpha\Gamma(\alpha)\quad(\alpha>0)$$

を示そう. $(-e^{-x})'=e^{-x}$ に関する部分積分を行うと,

$$\begin{aligned}\Gamma(\alpha+1)&=\int_0^\infty e^{-x}x^\alpha\,dx=\int_0^\infty(-e^{-x})'x^\alpha\,dx\\&=[-e^{-x}x^\alpha]_0^\infty+\int_0^\infty e^{-x}\alpha x^{\alpha-1}\,dx=\alpha\int_0^\infty e^{-x}x^{\alpha-1}\,dx=\alpha\Gamma(\alpha).\end{aligned}$$

3つめの等号で部分積分を行った. 4つめの等号では $\alpha>0$ より $\lim_{x\searrow 0}x^\alpha=0$ となることと, $\lim_{x\to\infty}e^{-x}x^\alpha=0$ となることを使った. 最初と最後の等号でガンマ函数の定義を使った.

2.16 ガンマ函数の函数等式のよく見る使い方

ガンマ函数の函数等式より, $k=0,1,2,\dots$ について,

$$\Gamma(\alpha+k)=(\alpha+k-1)\Gamma(\alpha+k-1)=\cdots=(\alpha+k-1)\cdots(\alpha+2)\alpha\Gamma(\alpha)$$

なので

$$\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}=\alpha(\alpha+k-1)\cdots(\alpha+k-1).$$

ガンマ函数の函数等式はこの形で使われることがよくある.

2.17 階乗のガンマ函数表示

ガンマ函数の函数等式と易しく証明できる $\Gamma(1)=1$ から,

$$\Gamma(n+1)=n!\quad(n=0,1,2,\dots)$$

となることを示せる. 実際, n が 0 以上の整数のとき,

$$\Gamma(n+1)=n\Gamma(n)=n(n-1)\Gamma(n-1)=\cdots=n(n-1)\cdots2\cdot1\,\Gamma(1)=n!.$$

ガンマ函数の函数等式を n 回使い, 最後に $\Gamma(1)=1$ を使った.

2.18 問題: 正の半整数でのガンマ函数の値

次を示せ:

$$\Gamma\left(\frac{2n+1}{2}\right)=\frac{1\cdot3\cdots(2n-1)}{2^n}\sqrt{\pi}=\frac{(2n)!}{2^{2n}n!}\sqrt{\pi}\quad(n=0,1,2,\dots)$$

解答例: n が0以上の整数であるとき,

$$\begin{aligned}\Gamma\left(\frac{2n+1}{2}\right)&=\frac{2n-1}{2}\Gamma\left(\frac{2n-1}{2}\right)=\frac{2n-1}{2}\frac{2n-3}{2}\Gamma\left(\frac{2n-3}{2}\right)=\cdots\\&=\frac{2n-1}{2}\frac{2n-3}{2}\cdots\frac{3}{2}\frac{1}{2}\Gamma\left(\frac{1}{2}\right)=\frac{1\cdot3\cdots(2n-1)}{2^n}\sqrt{\pi}.\end{aligned}$$

ガンマ函数の函数等式を n 回使い, 最後にGauss積分 $\Gamma(1/2)=\sqrt{\pi}$ を使った.

さらに, 分子分母に $2\cdot4\cdots(2n)=2^n n!$ をかけることによって,

$$\Gamma\left(\frac{2n+1}{2}\right)=\frac{1\cdot3\cdots(2n-1)}{2^n}\sqrt{\pi}=\frac{1\cdot3\cdots(2n-1)}{2^n}\frac{2\cdot4\cdots(2n)}{2^n n!}\sqrt{\pi}=\frac{(2n)!}{2^{2n}n!}\sqrt{\pi}.$$

解答終

2.19 問題: 標準正規分布の偶数次のモーメント

次を示せ:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}x^{2n}\,dx = \frac{2^n}{\sqrt{\pi}}\Gamma\left(\frac{2n+1}{2}\right) = \frac{(2n)!}{2^n\,n!} = 1\cdot 3\cdots(2n-1) \quad (n=0,1,2,\ldots)$$

解答例: 被積分函数が偶函数であることを使ってから, $x=\sqrt{y}$ とおくと $dx = y^{-1/2}\,dy/2$ となることより,

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}x^{2n}\,dx &= \frac{2}{\sqrt{2\pi}}\int_0^{\infty} e^{-x^2/2}x^{2n}\,dx = \frac{1}{\sqrt{2\pi}}\int_0^{\infty} e^{-y/2}y^{n-1/2}\,dy\\ &= \frac{1}{\sqrt{2\pi}}\int_0^{\infty} e^{-y/2}y^{(2n+1)/2-1}\,dy = \frac{1}{\sqrt{2\pi}}2^{(2n+1)/2}\Gamma\left(\frac{2n+1}{2}\right)\\ &= \frac{2^n}{\sqrt{\pi}}\Gamma\left(\frac{2n+1}{2}\right) = \frac{(2n)!}{2^n\,n!} = 1\cdot 3\cdots(2n-1).\end{aligned}$$

最後から4つめの等号でガンマ函数のよく使われる形 $\int_0^{\infty} e^{-x/\theta}x^{\alpha-1}\,dx = \theta^{\alpha}\Gamma(\alpha)$ を使い, 終わりの2つの等号で正の半整数でのガンマ函数の値を使った.

解答終

注意: 一般に確率変数 X について $E[X^k]$ $(k=0,1,2,\ldots)$ を X もしくは X が従う分布の **モーメント** (積率, moment)と呼ぶ. 上の結果は標準正規分布に従う確率変数 X のモーメントについて,

$$E[X^{2n}] = \frac{(2n)!}{2^n\,n!} = 1\cdot 3\cdots(2n-1) \quad (n=0,1,2,\ldots)$$

が成立することを意味している. x^{2n+1} は奇函数になるので $E[X^{2n+1}]=0$ となる.

2.20 問題: 標準正規分布のモーメント母函数とキュムラント母函数

確率変数 X について $E[e^{tX}]$ と $\log E[e^{tX}]$ をそれぞれ X の **モーメント母函数** (moment generating function, mgf)と **キュムラント母函数** (cumulant generating function, cgf)と呼ぶ.

標準正規分布に従う確率変数 X のときのモーメント母函数とキュムラント母函数がそれぞれ次のようになることを示せ:

$$E[e^{tX}] = e^{t^2/2}, \quad \log E[e^{tX}] = \frac{t^2}{2}.$$

注意: 標準正規分布のキュムラント母函数がこのようなシンプルな形になることは, 中心極限定理(正規分布による普遍的な近似定理)や正規分布で近似したときの誤差の見積もりなどで使われる.

解答例1 1つ前の問題の結果を使うと,

$$E[e^{tX}] = \sum_{n=0}^{\infty}\frac{t^{2n}E[X^{2n}]}{(2n)!} = \sum_{n=0}^{\infty}\frac{t^{2n}}{(2n)!}\frac{(2n)!}{2^n\,n!} = \sum_{n=0}^{\infty}\frac{(t^2/2)^n}{n!} = e^{t^2/2}, \quad \log E[e^{tX}] = \frac{t^2}{2}.$$

解答終

解答例2: $tx - x^2/2 = t^2/2 - (x-t)^2/2$ であることより,

$$E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx}\frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx = e^{t^2/2}\int_{-\infty}^{\infty}\frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}}\,dx = e^{t^2/2}, \quad \log E[e^{tX}] = \frac{t^2}{2}.$$

解答終

2.21 問題: n-1次元単位球面の面積のガンマ函数表示

$n-1$ 次元単位球面 S^{n-1} を

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

と定める(Sはsphereの頭文字). 例えば, S^1 は平面上の半径が 1 の円周(曲線なので1次元の図形)になり, S^2 は3次元空間上の半径が 1 の球面(曲面なので2次元の図形)になる. S^{n-1} の面積を A_{n-1} がガンマ函数によって次のように表されることを示せ:

$$A_{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)}.$$

解答例: Gauss積分の n 乗

$$I^n = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-(x_1^2+\cdots+x_n^2)}\, dx_1 \cdots dx_n$$

を二通りに計算することによって, $n-1$ 次元単位球面 S^{n-1} の面積 A_{n-1} を求めよう. まず, これがGauss積分の n 乗であることから,

$$I^n = \left(\int_{-\infty}^{\infty} e^{-x^2}\, dx\right)^2 = (\sqrt{\pi})^2 = \pi^{n/2}.$$

$n-1$ 次元球面上の微小領域の面積を $d\omega$ と書くと, 半径 r の $n-1$ 次元球面上の対応する微小領域の面積はその r^{n-1} 倍の $r^{n-1}\, d\omega$ になり, さらにそれに dr の厚みを持たせて作られる n 次元の微小領域の体積はほぼそれらの積 $r^{n-1}\, dr\, d\omega$ に等しくなる. $y = r^2 = x_1^2 + \cdots + x_n^2$ とおくと, $dr = d\sqrt{y} = y^{-1/2}\, dy/2$ なので,

$$r^{n-1}\, dr\, d\omega = \frac{1}{2} y^{(n-1)/2-1/2}\, dy\, d\omega = \frac{1}{2} y^{n/2-1}\, dy\, d\omega.$$

$dx_1 \cdots dx_n$ に関する積分は $\frac{1}{2} y^{n/2-1}\, dy\, d\omega$ に関する積分に書き換えられるので,

$$\pi^{n/2} = I^n = \int_{S^{n-1}} \left(\int_0^{\infty} e^{-y} \frac{1}{2} y^{n/2-1}\, dy\right) d\omega = \frac{A_{n-1}}{2} \int_0^{\infty} e^{-y} y^{n/2-1}\, dy = \frac{A_{n-1}}{2}\Gamma(n/2).$$

すなわち, $A_{n-1} = 2\pi^{n/2}/\Gamma(n/2)$.

解答終

注意: 半径 r の $n-1$ 次元球面の面積は単位球面の面積の r^{n-1} 倍になるので

$$A_{n-1} r^{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1}$$

になる. 例えば,

$$\Gamma(1/2) = \sqrt{\pi}, \quad \Gamma(1) = 1, \quad \Gamma(3/2) = \frac{1}{2}\sqrt{\pi}, \quad \Gamma(2) = 1, \quad \Gamma(5/2) = \frac{3}{4}\sqrt{\pi}$$

なので,

$$A_0 = \frac{2\pi^{1/2}}{\Gamma(1/2)} = 2, \quad A_1 r = \frac{2\pi^{2/2}}{\Gamma(2/2)} r = 2\pi r, \quad A_2 r^2 = \frac{2\pi^{3/2}}{\Gamma(3/2)} r^2 = 4\pi r^2,$$
$$A_3 r^3 = \frac{2\pi^{4/2}}{\Gamma(4/2)} r^3 = 2\pi^2 r^3, \quad A_4 r^4 = \frac{2\pi^{5/2}}{\Gamma(5/2)} r^4 = \frac{8\pi^2 r^4}{3}.$$

2 は集合 $S^0 = \{-1, 1\}$ の点の個数である. $2\pi r$ や $4\pi r^2$ は高校生にとってもおなじみの公式である.

2.22 $n-1$ 次元単位球面上の座標を使った面積の計算

上の問題の解答例中で単位球面上の自然な積分を $\int_{S^{n-1}} \cdots\, d\omega$ と書いた. その解答例のポイントは単位球面上の積分の具体的な座標表示を使わずに, そのように抽象的に書いたまま計算が最後まで遂行できることである.

しかし, 具体的に座標で表示しても同じ結果が得られるかどうがについて疑問に思った人もいるのではないだろうか?

そのような人のために以下では $n-1$ 次元単位球面上の座標を具体的に与えて面積を計算し直してみよう.

$n-1$ 次元球面

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

の上の点は次のようにパラメトライズできる:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \cos\theta_1 \cos\theta_2 \cos\theta_3 \cdots \cos\theta_{n-1} \\ \sin\theta_1 \cos\theta_2 \cos\theta_3 \cdots \cos\theta_{n-1} \\ \sin\theta_2 \cos\theta_3 \cdots \cos\theta_{n-1} \\ \sin\theta_3 \cdots \cos\theta_{n-1} \\ \ddots \quad \vdots \\ \sin\theta_{n-1} \end{bmatrix} \qquad \left(\begin{array}{l} -\pi \le \theta_1 \le \pi, \\ -\frac{\pi}{2} \le \theta_2, \ldots, \theta_{n-1} \le \frac{\pi}{2} \end{array}\right)$$

これは次のような手続きで得られる:

1. (x_1, x_2) 平面上の単位円周を $(x_1, x_2) = (\cos\theta_1, \sin\theta_1)$ とパラメトライズする.

2. (x_1, x_2) 上の単位円周上の点を角度 θ_2 で x_3 軸方向に持ち上げることによって, (x_1, x_2, x_3) 空間内の単位球面をパラメトライズする.
3. (x_1, x_2, x_3) 内の単位球面の点を角度 θ_3 で x_4 軸方向に持ち上げることによって, (x_1, x_2, x_3, x_4) 空間内の単位球面をパラメトライズする.
4. 以下これを繰り返す.

これによって, 単位球面上の面積は以下のようにして計算できることがわかる.

1. (x_1, x_2) 平面上の単位円周上の弧の長さは $d\omega_1 = d\theta_1 = \cos^0 \theta_1 \, d\theta_1$ の積分で計算できる.
2. (x_1, x_2, x_3) 空間内の単位球面上の面積は $d\omega_1$ を $\cos\theta_2$ 倍に縮小した結果と $d\theta_2$ の積である $d\omega_2 = \cos^0 \theta_1 \cos^2 \theta_2 \, d\theta_1 \, d\theta_2$ を積分すれば計算できる.
3. (x_1, x_2, x_3, x_4) 空間内の単位球面上の面積は $d\omega_2$ が2次元の面積要素なのでそれを $\cos\theta_2$ の2乗倍した結果と $d\theta_3$ の積である $d\omega_2 = \cos^0 \theta_1 \cos^1 \theta_2 \cos^2 \theta_3 \, d\theta_1 \, d\theta_2 \, d\theta_3$ を積分すれば計算できる.
4. 以下これを繰り返す.

ゆえに, $n-1$ 次元単位球面 S^{n-1} 上の領域の面積は

$$d\omega_{n-1} = \cos^0 \theta_1 \cos^1 \theta_2 \cos^2 \theta_3 \cdots \cos^{n-2} \theta_{n-1} \, d\theta_1 \, d\theta_2 \, d\theta_3 \cdots d\theta_{n-1}$$

を積分すれば計算できる. 特に S^{n-1} の面積 A_{n-1} は以下のように表される:

$$A_{n-1} = 2 \prod_{k=1}^{n-1} \int_{-\pi/2}^{\pi/2} \cos^{k-1} \theta_k \, d\theta_k .$$

θ_1 に関する $-\pi$ から π までの積分を $-\pi/2$ から $\pi/2$ までの積分に変更し, その代わりに 2 倍してつじつまを合わせたことに注意せよ. (S^0 が2点集合であることを思い出せばこれは自然な処理であることがわかる.)

後で示すベータ函数の三角函数による表示とベータ函数をガンマ函数で表す公式

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

と $\Gamma(1/2) = \sqrt{\pi}$ を使うと,

$$\int_{-\pi/2}^{\pi/2} \cos^k \theta \, d\theta = 2 \int_0^{\pi/2} \cos^{k-1} \theta \, d\theta = B(k/2, 1/2) = \frac{\Gamma(k/2)\sqrt{\pi}}{\Gamma((k+1)/2)}.$$

ゆえに

$$A_{n-1} = 2 \frac{\sqrt{\pi}\,\sqrt{\pi}}{\Gamma(2/2)} \frac{\Gamma(2/2)\sqrt{\pi}}{\Gamma(3/2)} \frac{\Gamma(3/2)\sqrt{\pi}}{\Gamma(4/2)} \cdots \frac{\Gamma((n-1)/2)\sqrt{\pi}}{\Gamma(n/2)} = \frac{2\pi^{n/2}}{\Gamma(n/2)}.$$

上の問題と同じ結果が得られた.

2.23 問題: n次元単位球体の面積のガンマ函数表示

n 次元単位球体 B^n を

$$B^n = \{ (x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1^2 + \cdots + x_n^2 \leq 1 \}$$

と定める(Bはballの頭文字). 例えば, B^1 は直線上の線分 $[-1, 1]$ になり, B^2 は2次元平面上の半径が 1 の円盤になり, B^3 は3次元空間内の半径 1 の単位球体になる. B^n の体積を V_n がガンマ函数によって次のように表されることを示せ:

$$V_n = \frac{\pi^{n/2}}{\Gamma(n/2+1)}.$$

解答例: 半径 r の $n-1$ 次元球面の面積

$$A_{n-1} r^{n-1} = \frac{2\pi^{n/2}}{\Gamma(n/2)} r^{n-1}$$

を r について 0 から 1 まで積分すれば, n 次元単位球体の体積が得られる:

$$V_n = \int_0^1 A_{n-1} r^{n-1} \, dr = \frac{A_{n-1}}{n} = \frac{2\pi^{n/2}}{n\Gamma(n/2)} = \frac{\pi^{n/2}}{(n/2)\Gamma(n/2)} = \frac{\pi^{n/2}}{\Gamma(n/2+1)}.$$

最後の等号でガンマ函数の函数等式を使った.

解答終

注意: この問題の大幅な一般化と解答を19世紀にDirichlet氏が与えている. このノートのずっと下の方で紹介されているDirichlet積分=多変数ベータ函数の解説を参照せよ.

注意: 半径 r の n 次元球体の体積は単位球体の体積の r^n 倍になるので

$$V_n r^n = \frac{\pi^{n/2}}{\Gamma(n/2+1)} r^n$$

になる. 例えば,

$$\Gamma(3/2) = \frac{1}{2}\sqrt{\pi}, \quad \Gamma(2) = 1, \quad \Gamma(5/2) = \frac{3}{4}\sqrt{\pi}\Gamma(3) = 2\sqrt{\pi}\Gamma(7/2) = \frac{15}{8}\sqrt{\pi}$$

なので,

$$V_1 r = \frac{\pi^{1/2}}{\Gamma(1/2+1)} r = 2r, \quad V_2 r^2 = \frac{\pi^{2/2}}{\Gamma(2/2+1)} r^2 = \pi r^2, \quad V_3 r^3 = \frac{\pi^{3/2}}{\Gamma(3/2+1)} r^3 = \frac{4\pi^2 r^3}{3},$$
$$V_4 r^4 = \frac{\pi^{4/2}}{\Gamma(4/2+1)} r^4 = \frac{\pi^2 r^4}{2}, \quad V_5 r^5 = \frac{\pi^{5/2}}{\Gamma(5/2+1)} r^5 = \frac{8\pi^2 r^5}{15}.$$

$2r$ は線分 $[-r, r]$ の長さである. πr^2 や $4\pi r^3/3$ は高校生にとってもおなじみの公式だろう.

注意: m は0以上の整数であるとし, $n = 2m$ または $n = 2m+1$ と仮定する. このとき, n 次元球体の球体の体積の公式に π^m が含まれる理由は, $n = 2m, 2m+1$ 次元の空間では互いに可換な回転軸を m 本取れるからである. 1本の回転軸ごとに1つの π が出て来る. 例えば, $m = 2$, $n = 5$ のとき, 5 次元空間 $\mathbb{R}^5 = \{\, (x_1, x_2, x_3, x_4, x_5) \mid x_1, x_2, x_3, x_4, x_5 \in \mathbb{R} \,\}$ における互いに可換な回転軸として, (x_1, x_2) 平面の回転軸と (x_3, x_4) 平面の回転軸の2本が取れて, 5 次元球体の体積の公式には π^2 が含まれている.

2.24 Stirlingの公式

次の階乗の近似公式を **Stirlingの公式** と呼ぶ:

$$n! = n^n e^{-n} \sqrt{2\pi n}\,(1+\varepsilon_n) \quad (\varepsilon_n \to 0 \text{ as } n \to \infty)$$

Stirlingの公式は後で空気のごとく使われることになる.

Stirlingの公式はガンマ函数の応用として以下のように示される.

$n!/(n^n e^{-n}\sqrt{n})$ が $n \to \infty$ で $\int_{-\infty}^{\infty} e^{-y^2/2}\,dy = \sqrt{2\pi}$ に収束することを示せばよい.

$n! = \Gamma(n+1) = \int_0^\infty e^{-x} x^n\,dx$ で

$$x = n + \sqrt{n}\,y = n\left(1 + \frac{y}{\sqrt{n}}\right)$$

とおくと(これがポイント!),

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

そして,

$$\begin{aligned}
\log\left(e^{-\sqrt{n}\,y}\left(1 + \frac{y}{\sqrt{n}}\right)^n\right) &= n\log\left(1 + \frac{y}{\sqrt{n}}\right)^n - \sqrt{n}\,y \\
&= n\left(\frac{y}{\sqrt{n}} - \frac{y^2}{2n} + \frac{y^3}{3n\sqrt{n}} - \cdots\right) - \sqrt{n}\,y = -\frac{y^2}{2} + \frac{y^3}{3\sqrt{n}} - \cdots \\
&\to -\frac{y^2}{2} \quad (\text{as } n \to \infty)
\end{aligned}$$

ゆえに, $e^{-\sqrt{n}\,y}\left(1 + y/\sqrt{n}\right)^n \to e^{-y^2/2}$ $(n \to \infty)$.

したがって, $n \to \infty$ のとき,

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

これで示すべきことが示された.

注意: 極限と積分の交換可能性が気になる読者はLebesgueの収束定理の適用を考えてみよ. もっと初等的な方法でも極限と積分の交換可能性を示せる. この手の細かいことを気にならない人はそのまま気にする必要はないが, 関連のグラフを大量に描いてみたり, 数値積分を試してみたり, 様々な試行錯誤によって理解を深める努力をした方がよい.

2.25 問題: 小さな n でもStirlingの公式による $n!$ の近似精度がそう悪くないこと

$n = 1, 2, 3, \ldots, 10$ について $n!$ と

$$\mathrm{stirling}(n) = n^n e^{-n} \sqrt{2\pi n}$$

の両方を計算し, 前者の後者による近似の相対誤差を求めよ. ここで $a > 0$ の b による近似の相対誤差の定義は $b/a - 1$ であるとする. $\mathrm{stirling}(n)$ については小数点以下第3桁まで, 相対誤差について小数点以下第5桁まで計算せよ.

拡張版解答例＋解説: $1/(12n)$ によって補正した近似公式

$$\mathrm{stirling}_1(n) = n^n e^{-n} \sqrt{2\pi n} \left(1 + \frac{1}{12n}\right)$$

についてもあわせて計算してみよう.

n	$n!$	$\mathrm{stirling}(n)$	relative error	$n!$	$\mathrm{stirling}_1(n)$	relative error
1	1	0.922	−0.07786	1	0.99898	−0.001018
2	2	1.919	−0.04050	2	1.99896	−0.000519
3	6	5.836	−0.02730	6	5.99833	−0.000279
4	24	23.506	−0.02058	24	23.99589	−0.000171
5	120	118.019	−0.01651	120	119.98615	−0.000115
6	720	710.078	−0.01378	720	719.94038	−0.000083
7	5040	4980.396	−0.01183	5040	5039.68626	−0.000062
8	40320	39902.395	−0.01036	40320	40318.04541	−0.000048
9	362880	359536.873	−0.00921	362880	362865.91796	−0.000039
10	3628800	3598695.619	−0.00830	3628800	3628684.74890	−0.000032

この結果を見れば, $n!$ のStirling公式による近似の相対誤差の絶対値は $n = 2$ ですでに 5% を切っており, $n = 9$ で 1% を切っていることがわかる. $1/(12n)$ で補正した場合には $n = 1$ の段階ですでに相対誤差の絶対値はほぼ 0.1% になっていて非常に小さい. この事実は

$$\frac{13\sqrt{2\pi}}{12e} \approx 0.9989817596371048$$

が非常に 1 に近いことが偶然ではないことも意味している. このように, $1/(12n)$ による補正で誤差が非常に小さくなるという事実は, Stirlingの公式による階乗の近似の相対誤差の大きさがほぼ $1/(12n)$ になることを意味している.

階乗は確率の記述でよく使われるので, 階乗のシンプルな近似式であるStirlingの公式は確率が絡む議論において非常に有用である.

拡張版解答終

In [33]:

```
stirling(n) = n^n * exp(-n) * √(2π*n)
stirling1(n) = n^n * exp(-n) * √(2π*n) * (1 + 1/(12n))
for n in 1:10
    a = factorial(n)
    b = stirling(n)
    c = stirling1(n)
    @printf("%2d & %7d & %11.3f & %8.5f & %7d & %13.5f & %9.6f \\\\\n", n, a, b, b/a - 1, a, c
end
```

```
 1 &       1 &       0.922 & -0.07786 &       1 &       0.99898 & -0.001018 \\
 2 &       2 &       1.919 & -0.04050 &       2 &       1.99896 & -0.000519 \\
 3 &       6 &       5.836 & -0.02730 &       6 &       5.99833 & -0.000279 \\
 4 &      24 &      23.506 & -0.02058 &      24 &      23.99589 & -0.000171 \\
 5 &     120 &     118.019 & -0.01651 &     120 &     119.98615 & -0.000115 \\
 6 &     720 &     710.078 & -0.01378 &     720 &     719.94038 & -0.000083 \\
 7 &    5040 &    4980.396 & -0.01183 &    5040 &    5039.68626 & -0.000062 \\
 8 &   40320 &   39902.395 & -0.01036 &   40320 &   40318.04541 & -0.000048 \\
 9 &  362880 &  359536.873 & -0.00921 &  362880 &  362865.91796 & -0.000039 \\
10 & 3628800 & 3598695.619 & -0.00830 & 3628800 & 3628684.74890 & -0.000032 \\
```

In [34]:
```
13√(2π)/(12e)
```
Out[34]: 0.9989817596371048

In [35]:
```
N = 10
plot(; legend=:topleft, xtick=1:N)
scatter!(1:N, factorial; label="n!")
plot!(stirling, 1, N; label="nⁿ e⁻ⁿ √(2πn)")
title!("Stirling approximation of n!")
```
Out[35]:

In [36]:
```
N = 10
plot(; legend=:topleft, xtick=1:N)
scatter!(1:N, factorial; label="n!")
plot!(stirling, 1, N; label="nⁿ e⁻ⁿ √(2πn)")
title!("Stirling approximation of n!")
plot!(;yscale=:log10)
```
Out[36]:

2.26 Stirlingの公式とガンマ分布の正規分布近似(中心極限定理)の関係

この節では, Stirlingの公式は本質的にガンマ分布の正規分布近似(中心極限定理ともいう)から出て来ることを説明する.

スケールパラメータが $\theta = 1$ で形状パラメータが $\alpha = n + 1$ のガンマ分布 $\mathrm{Gamma}(n+1, 1)$ の確率密度函数は

$$p_n(x) = \frac{e^{-x}x^n}{\Gamma(n+1)} = \frac{e^{-x}x^n}{n!} \qquad (x > 0)$$

という形になる. この $p_n(x)$ の分子の対数を $f_n(x) と書く:

$$f_n(x) = \log(e^{-x}x^n) = n\log x - x.$$

$f_n(x)$ を $x = n$ でTaylor展開してみよう.

$$f_n(n) = n\log n - n, \quad f_n'(x) = \frac{n}{x} - 1, \quad f_n'(n) = 0, \quad f_n''(x) = -\frac{n}{x^2}, \quad f_n''(n) = -\frac{1}{n}.$$

これより, x が n に近いとき,

$$f_n(x) \approx n\log n - n - \frac{(x-n)^2}{2n}.$$

ゆえに, x が n に近いとき, $f_n(x)$ の $x = n$ でのTaylor展開より,

$$e^{-n}x^n = \exp(f_n(x)) \approx \exp\left(n\log n - n - \frac{(x-n)^2}{2n}\right) = n^n e^{-n}\exp\left(-\frac{(x-n)^2}{2n}\right).$$

n が大きなとき, $e^{-n}x^n$ も $\exp(-(x-n)^2/(2n))$ も x が n から離れると急激に 0 に近付くので, それぞれの $x > 0$, $-\infty < x < \infty$ における積分は n に近い x における積分で近似される. したがって,

$$\begin{aligned}
1 &= \int_0^\infty p_n(x)\,dx = \frac{1}{n!}\int_0^\infty e^{-n}x^n\,dx \\
&\approx \frac{1}{n!}\int_{\text{around } n} e^{-n}x^n\,dx \approx \frac{n^n e^{-n}}{n!}\int_{\text{around } n}\exp\left(-\frac{(x-n)^2}{2n}\right)dx \\
&\approx \frac{n^n e^{-n}}{n!}\int_{-\infty}^\infty \exp\left(-\frac{(x-n)^2}{2n}\right)dx = \frac{n^n e^{-n}\sqrt{2\pi n}}{n!}.
\end{aligned}$$

これで $n! \approx n^n e^{-n}\sqrt{2\pi n}$ が再び示された.

以上の計算によってガンマ分布 $\mathrm{Gamma}(n+1, 1)$ の密度函数 $p_n(x)$ が平均 n, 分散 n の正規分布の密度函数で近似されることも示せている:

$$p_n(x) = \frac{e^{-x}x^n}{n!} \approx \frac{1}{\sqrt{2\pi n}}\exp\left(-\frac{(x-n)^2}{2n}\right).$$

以上はかなり雑な議論だが, 議論の全体を正確に作り直すこともできるし, 以下で示すようにグラフを描くことによっても結果が正しいことを確認できる.

In [37]:

```
p_gamma(n, x) = x > 0 ? exp(n*log(x) - x - loggamma(n+1)) : zero(x)
p_gamma_normalapprox(n, x) = 1/√(2π*n) * exp(-(x - n)^2/(2n))
function plot_gamma_normal(n; kwargs...)
    x = range(n - 5√n, n + 6√n, 1000)
    P1 = plot(x, x -> p_gamma(n, x); label="Gamma(n+1,1)", kwargs...)
    plot!(x, x -> p_gamma_normalapprox(n, x); label="Normal(n,√n)", ls=:dash)
    vline!([n]; label="", c=:black, ls=:dot)
    title!("n = $n")
    P2 = plot(x, x -> p_gamma_normalapprox(n, x) - p_gamma(n, x); label="difference", kwargs..
    plot!(x, x -> p_gamma(n, 2n - x) - p_gamma_normalapprox(n, 2n - x); label="180° rotation")
    vline!([n]; label="", c=:black, ls=:dot)
    hline!([0]; label="", c=:black, ls=:dot)
    plot!(ylim=(-1.05p_gamma(n,n), 1.05p_gamma(n,n)))
    plot(P1, P2; size=(800, 250))
end
```

Out[37]:

```
plot_gamma_normal (generic function with 2 methods)
```

In [38]:

```
plot_gamma_normal(3; xtick=-20:20)
```

Out[38]:

In [39]:
```
plot_gamma_normal(10)
```

Out[39]:

In [40]:
```
plot_gamma_normal(30)
```

Out[40]:

In [41]:
```
plot_gamma_normal(100)
```

Out[41]:

3 ベータ函数

3.1 ベータ函数の定義

$\alpha, \beta > 0$ について, ベータ函数 $B(\alpha, \beta)$ を次のように定める:

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt.$$

ベータ函数のこの表示は通常の(第1種)ベータ分布だけではなく, F 分布(Fisherの F 分布)とも関係している. ベータ函数の様々な表示が正規分布に関連した重要な分布達の基礎を与える.

3.2 ベータ分布の定義

$\alpha, \beta > 0$ と仮定する.

次の確率密度函数によって定義される連続分布をパラメータ α, β の **ベータ分布** (Beta distribution)と呼び,

$$\mathrm{Beta}(\alpha, \beta)$$

と表す:

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

二項分布がBernoulli試行(値が 1 と 0 のみの乱数生成の繰り返し)から得られたのと同じように, パラメータが正の整数の場合のベータ分布は一様乱数の生成の繰り返しから得られる. この事実については後で別のノートで解説する予定である. この段階ではベータ分布にも後で説明されるように直観的な意味付けがあることを覚えておけばよい. 実用的に使用される確率分布の世界は豊富かつ複雑であり, それらを単に暗記しようとすることには無理がある. 直観的な意味付けを十分に理解して先に進む方が効率的な場合が多い.

ベータ分布の密度函数は $\alpha < 0$ ならば $t \searrow 0$ のとき無限大になり, $\beta > 0$ ならば $t \nearrow 1$ のとき無限大に発散することには注意しなければいけない.

α と β の両方が大きなとき, ベータ分布は正規分布で近似される.

ベータ分布の密度函数のグラフは以下のようになる.

In [42]:
```
function plot_beta(α, β; kwargs...)
    beta = Beta(α, β)
    normal = Normal(mean(beta), std(beta))
    if α < 1 || β < 1
        x = range(eps(), 1-eps(), 1000)
        plot(x, x -> pdf(beta, x); ylim=(-0.1, 4.1), label="Beta(α, β)")
    else
        plot(x -> pdf(beta, x), 0, 1; label="Beta(α, β)")
    end
    plot!(x -> pdf(normal, x), 0, 1; label="normal approx.", ls=:dash)
    title!("α = $α,  β = $β")
    plot!(; xtick=0:0.1:1, kwargs...)
end
```

Out[42]: plot_beta (generic function with 1 method)

In [43]:
```
plot(plot(plot_beta(0.5, 2)), plot(plot_beta(2, 0.5); legend=:topleft); size=(800, 250))
```

Out[43]:

In [44]:
```
plot(plot(plot_beta(0.5, 0.5); legend=:top),
    plot(plot_beta(1, 1); legend=:bottom, ylim=(-0.05, 1.5)); size=(800, 250))
```

Out[44]:

In [45]:
```
plot(plot(plot_beta(3, 7)), plot(plot_beta(6, 14)); size=(800, 250))
```

Out[45]:

In [46]:
```
plot(plot(plot_beta(30, 70)), plot(plot_beta(120, 280)); size=(800, 250))
```

Out[46]:

3.3 問題: ベータ分布の期待値と分散

分布 $\mathrm{Beta}(\alpha, \beta)$ の期待値と分散がそれぞれ

$$\frac{\alpha}{\alpha+\beta}, \quad \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

となることを示せ. ただし, あとで示すことになる公式 $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$ を自由に用いてよい.

解答例: T は分布 $\mathrm{Beta}(\alpha, \beta)$ に従う確率変数とする.

$$
\begin{aligned}
E[T^k] &= \frac{1}{B(\alpha,\beta)}\int_0^1 t^{\alpha+k-1}(1-t)^{\beta-1}\,dt = \frac{B(\alpha+k,\beta)}{B(\alpha,\beta)} \\
&= \frac{\Gamma(\alpha+k)\Gamma(\beta)}{\Gamma(\alpha+\beta+k)}\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+k-1)}.
\end{aligned}
$$

ゆえに $E[T] = \alpha/(\alpha+\beta)$,

$$
\begin{aligned}
\mathrm{var}(T) &= E[T^2] - E[T]^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} \\
&= \frac{\alpha((\alpha+1)(\alpha+\beta)-\alpha(\alpha+\beta+1))}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.
\end{aligned}
$$

解答終

3.4 ベータ函数は α と β を交換しても不変

$t = 1-u$ とおくと,

$$
B(\beta,\alpha) = \int_0^1 t^{\beta-1}(1-t)^{\alpha-1}\,dt = \int_1^0 (1-u)^{\beta-1}u^{\alpha-1}\,(-du) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1}\,du = B(\alpha,\beta).
$$

3.5 問題: ベータ分布の三角函数を使った表示

次を示せ:

$$
B(\alpha,\beta) = 2\int_0^{\pi/2}(\cos\theta)^{2\alpha-1}(\sin\theta)^{2\beta-1}\,d\theta \quad (\alpha,\beta>0).
$$

解答例: $\alpha,\beta>0$ と仮定する. $t=\cos^2\theta$ とおくと, $dt = -2\cos\theta\,\sin\theta\,d\theta$ なので,

$$
\begin{aligned}
B(\alpha,\beta) &= \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt = \int_{\pi/2}^0 (\cos\theta)^{2(\alpha-1)}(\sin\theta)^{2(\beta-1)}\,(-2\cos\theta\,\sin\theta\,d\theta) = \\
&= 2\int_0^{\pi/2}(\cos\theta)^{2\alpha-1}(\sin\theta)^{2\beta-1}\,d\theta.
\end{aligned}
$$

解答終

注意: 特に

$$
B(1/2,1/2) = 2\int_0^{\pi/2} d\theta = \pi.
$$

となることがわかる. 次の積分は **Wallis積分** と呼ばれている:

$$
\int_0^{\pi/2}\sin^n\theta\,d\theta = \frac{1}{2}B(1/2,(n+1)/2).
$$

3.6 ベータ函数の極限としてガンマ函数を表せること

$\alpha>0$, $b\in\mathbb{R}$, $n+b>0$ と仮定する. このとき, 次が成立している:

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

積分変数を $t = x/n$ で x に変換すると,

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

なので, $n\to\infty$ のとき, $(1-x/n)^{n+b-1}\to e^{-x}$ となることを使うと,

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

例: t 分布の自由度無限大の極限の計算で使った $\sqrt{\nu}B(1/2, \nu/2)$ の $\nu \to \infty$ の極限は上の結果を使うと以下のようにして計算される:

$$\sqrt{\nu}B(1/2, \nu/2) = \sqrt{2}\sqrt{\nu/2}B(1/2, \nu/2) \to \sqrt{2}\Gamma(1/2) = \sqrt{2\pi}.$$

In [47]:

```
beta_approx_of_gamma(n, α) = exp(α*log(n) + logbeta(α, n))
x = range(0, 30, 1000)
plot(; legend=:topleft, yscale=:log10)
plot!(x, gamma; label="Γ(x)", ls=:dash)
for n in (10, 30, 100, 300)
    plot!(x, x -> beta_approx_of_gamma(n, x); label="n = $n")
end
title!("y = nˣ B(x, n)")
```

Out[47]:

3.7 問題: ベータ分布の極限でガンマ分布が得られること

$\alpha, \theta > 0$, $b \in \mathbb{R}$ と仮定する. T_n はパラメータ α, $n/\theta + b > 0$ を持つベータ分布に従う確率変数であるとする:

$$T_n \sim \mathrm{Beta}(\alpha, n/\theta + b).$$

このとき, nT_n が従う分布は $n \to \infty$ でガンマ分布 $\mathrm{Gamma}(\alpha, \theta)$ に近付くことを示せ. すなわち次が成立することを示せ:

$$\lim_{n\to\infty} E[f(nT_n)] = \int_0^\infty f(x)\frac{e^{-x/\theta}x^{\alpha-1}}{\theta^\alpha\Gamma(\alpha)}\,dx.$$

解答例: 定義に戻って $E[f(nT_n)]$ を計算すると $n \to \infty$ で以下が成立することがわかる:

$$\begin{aligned}
E[f(nT_n)] &= \int_0^1 f(nt)\frac{t^{\alpha-1}(1-t)^{n/\theta+b-1}}{B(\alpha, n/\theta+b)}\,dt \\
&= \frac{1}{B(\alpha, n/\theta+b)}\int_0^n f(x)\left(\frac{x}{n}\right)^{\alpha-1}\left(1-\frac{x}{n}\right)^{n/\theta+b-1}\frac{dx}{n} \\
&= \frac{1}{n^\alpha B(\alpha, n/\theta+b)}\int_0^n f(x)x^{\alpha-1}\left(1-\frac{x}{n}\right)^{n/\theta+b-1}dx \\
&= \frac{1}{\theta^\alpha \underbrace{(n/\theta)^\alpha B(\alpha, n/\theta+b)}_{\to\Gamma(\alpha)}}\int_0^n f(x)\underbrace{\left(1-\frac{x}{n}\right)^{n/\theta+b-1}}_{\to\exp(-x/\theta)}x^{\alpha-1}\,dx \\
&\to \frac{1}{\theta^\alpha\Gamma(\alpha)}\int_0^\infty f(x)e^{-x/\theta}x^{\alpha-1}\,dx.
\end{aligned}$$

2つめの等号で $t = x/n$ とおいた. これで示すべきことが示された.

解答例

注意: 正規分布モデルの統計学ではベータ分布がらみの確率分布達(例えば t 分布, F 分布)が大活躍する. 上の結果はそれらが, 正規分布またはガンマ分布がらみの確率分布(例えばχ^2分布)の裾野を太くする拡張になっていることを本質的に含んでいる.

In [48]:
```
pdf_Tn(n, α, θ, b, x) = pdf(Beta(α, n/θ + b), x/n)/n

function plot_nbeta(n, α, θ, b=1)
    xmin, xmax = max(0, α*θ - 4√α*θ), α*θ + 4√α*θ
    plot(x -> pdf_Tn(n, α, θ, b, x), xmin, xmax; label="n Beta(α,n/θ+b)")
    plot!(x -> pdf(Gamma(α, θ), x), xmin, xmax; label="Gamma(α,β)", ls=:dash)
    α < 1 && plot!(; ylim=(-0.3, 3.3))
    title!("n=$n, α=$α, θ=$θ, b=$b")
end
```

Out[48]: plot_nbeta (generic function with 2 methods)

In [49]:
```
plot(plot_nbeta.((1.0, 1.5, 2.0), 0.5, 1)...; size=(800, 200), layout=(1, 3))
```

Out[49]:

In [50]:
```
plot(plot_nbeta.((3, 10, 30), 1, 2)...; size=(800, 200), layout=(1, 3))
```

Out[50]:

In [51]:
```
plot(plot_nbeta.((10, 30, 100), 3, 1)...; size=(800, 200), layout=(1, 3))
```

Out[51]:

3.8 ベータ函数の別の基本的表示

統計学的にはベータ函数の次の表示も重要である:

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

この表示は t 分布(Studentの t 分布)と関係している.

上の表示は以下のようにして証明される.

$t=u/(1+u)=1-1/(1+u)$ とおくと, $dt=du/(1+u)^2$ となるので,

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

となる. ベータ函数は α と β の交換で不変なので被積分函数の分子の α を β で置き換えても等しくなる.

注意: $0 < t < 1$ と $0 < u < \infty$ と $-\infty < x < \infty$ のあいだの変数変換

$$t = \frac{u}{1+u} = \frac{1}{1+e^{-x}} = \mathrm{logistic}(x), \qquad \frac{t}{1-t} = u = e^x, \qquad \mathrm{logit}(t) = \log\frac{t}{1-t} = \log u = x$$

は統計学に関連した計算で非常によく現れる鉄板の基本パターンの一つである. t を確率とみなすとき, $u = t/(1-t)$ をその **オッズ** と呼び, $x = \log(t/(1-t))$ をその **ロジット変換** と呼ぶ.

3.9 問題: t 分布の基礎になる公式

次を示せ:

$$\int_{-\infty}^{\infty}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} dt = \sqrt{\nu}\,B(1/2,\nu/2) \quad (\nu > 0).$$

解答例: $\nu > 0$ と仮定する.

$$\begin{aligned}\int_{-\infty}^{\infty}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} dt &= 2\int_0^{\infty}\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} dt = 2\sqrt{\nu}\int_0^{\infty}(1+s^2)^{-(\nu+1)/2}\,ds \\ &= \sqrt{\nu}\int_0^{\infty}(1+u)^{-(\nu+1)/2}\,u^{-1/2}\,du = \sqrt{\nu}\int_0^{\infty}\frac{u^{1/2-1}\,du}{(1+u)^{1/2+\nu/2}} = \sqrt{\nu}\,B(1/2,\nu/2).\end{aligned}$$

最初の等号で被積分函数が偶函数であることを使った. 2つめの等号で $t = \sqrt{\nu}\,s$ とおいた. 3つめの等号で $s = \sqrt{u}$, $ds = u^{-1/2}\,du/2$ とおいた. 最後の等号でベータ函数の別の基本的表示を使った.

解答終

注意: 上の計算での積分変数の変換では $t = \sqrt{\nu u}$, $u = t^2/\nu$ となっている.

3.10 t 分布の定義

$\nu > 0$ と仮定する.

次の確率密度函数によって定義される連続分布を **自由度** (degree of freedom) が ν の t **分布** (**Student の** t **分布**, Student's t-distribution)と呼び,

$$\mathrm{TDist}(\nu)$$

と表す:

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

t 分布は正規分布モデルによる平均に関する検定や推定で基本的な役目を果たすことになる. この点については後で別のノートで詳しく説明する.

3.11 t 分布の期待値と分散

$T \sim \mathrm{TDist}(\nu)$ と仮定する. $|t| \to \infty$ で自由度 ν の t 分布の確率密度函数は

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

のように振る舞うので, $m \geq 0$ に対する $E[|T|^m]$ が有限の値になって m 次のモーメント $E[T^m]$ が有限の値としてうまく定義されるための必要十分条件は $m < \nu$ が成立することである.

ゆえに, 期待値 $E[T]$ が定義されるための必要十分条件は $\nu > 1$ となることである. t 分布の確率密度函数は偶函数なので, $\nu > 1$ のとき期待値は $E[T] = 0$ となる.

$\nu > 1$ のとき, 分散 $E[T^2]$ が有限の値になるための必要十分条件は $\nu > 2$ となることである. t 分布の分散は次のようになることを示そう:

$$\nu > 2 \implies \mathrm{var}(K) = \frac{\nu}{\nu-2}.$$

証明: $\nu > m \geq 0$ のとき,

$$
\begin{aligned}
E[|T|^m] &= \frac{1}{\sqrt{\nu}\,B(1/2,\nu/2)}\int_{-\infty}^{\infty}|t|^m\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}dt \\
&= \frac{2}{\sqrt{\nu}\,B(1/2,\nu/2)}\int_{0}^{\infty}|t|^m\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2}dt \\
&= \frac{2}{\sqrt{\nu}\,B(1/2,\nu/2)}\int_{0}^{\infty}\nu^{m/2}u^{m/2}(1+u)^{-(\nu+1)/2}\,\frac{1}{2}\sqrt{\nu}\,u^{-1/2}\,du \\
&= \frac{\nu^{m/2}}{B(1/2,\nu/2)}\int_{0}^{\infty}\frac{u^{(m+1)/2-1}}{(1+u)^{((m+1)+(\nu-m))/2}}\,du \\
&= \frac{\nu^{m/2}\,B((m+1)/2,(\nu-m)/2)}{B(1/2,\nu/2)} \\
&= \nu^{m/2}\frac{\Gamma((m+1)/2)}{\Gamma(1/2)}\frac{\Gamma((\nu-m)/2)}{\Gamma(\nu/2)}.
\end{aligned}
$$

3つめの等号で $t=\sqrt{\nu u}$ とおいた. 特に $m=2k<\nu\ (k=0,1,2,\ldots)$ のとき,

$$
\begin{aligned}
E[T^{2k}] &= \nu^k\frac{\Gamma(1/2+k)}{\Gamma(1/2)}\frac{\Gamma(\nu/2-k)}{\Gamma(\nu/2)} \\
&= \nu^k\frac{(1/2)(3/2)\cdots((2k-1)/2)}{((\nu-2)/2)((\nu-4)/2)\cdots((\nu-2k)/2)} \\
&= \frac{1\cdot 3\cdots(2k-1)\,\nu^k}{(\nu-2)(\nu-4)\cdots(\nu-2k)}.
\end{aligned}
$$

特に $k=1$ の場合より,

$$
\mathrm{var}(T) = E[T^2] = \frac{\nu}{\nu-2}.
$$

証明終

3.12 問題: t 分布の自由度無限大の極限で標準正規分布が得られること

t 分布の確率密度函数 $p(t|\nu)$ について次を示せ:

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

解答例: $t\in\mathbb{R}$ であるとする.

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

なので, $\nu\to\infty$ のとき, $\sqrt{\nu}\,B(1/2,\nu/2)\to\sqrt{2\pi}$ かつ $(1+t^2/\nu)^{-(\nu+1)/2}\to e^{-t^2}$ となることを示せばよい.

$N\to\infty$ のとき $(1+x/N)^N\to e^x$ より, $\nu\to\infty$ とすると,

$$
\left(1+\frac{t^2}{\nu}\right)^{-(\nu+1)/2} = \left(1+\frac{t^2/2}{\nu/2}\right)^{-\nu/2}\left(1+\frac{t^2/2}{\nu/2}\right)^{-1/2}\to e^{-t^2/2}.
$$

$u=x/(\nu/2)$ とおくと, $du/u=dx/x$ なので,

$$
\begin{aligned}
\sqrt{\nu}\,B(1/2,\nu/2) &= \sqrt{\nu}\int_0^{\infty}\frac{u^{1/2-1}\,du}{(1+u)^{1/2+\nu/2}} = \frac{\sqrt{\nu}}{(\nu/2)^{1/2}}\int_0^{\infty}\frac{x^{1/2-1}\,dx}{(1+x/(\nu/2))^{1/2+\nu/2}} \\
&= \sqrt{2}\int_0^{\infty}\frac{x^{1/2-1}\,dx}{(1+x/(\nu/2))^{1/2+\nu/2}}
\end{aligned}
$$

なので $\nu\to\infty$ とすると,

$$
\sqrt{\nu}\,B(1/2,\nu/2)\to\sqrt{2}\int_0^{\infty}\frac{x^{1/2-1}\,dx}{e^x} = \sqrt{2}\int_0^{\infty}e^{-x}x^{1/2-1}\,dx = \sqrt{2}\Gamma(1/2) = \sqrt{2\pi}.
$$

解答終

注意: t 分布という新しい複雑に見える確率分布を定義したが, 自由度が大きな場合には, t 分布の代わりに正規分布を使った議論と実践的には(近似的には)同じ結果が得られることになる. 自由度が大体 30 程度で実践的には t 分布はほぼ正規分布だと思ってよい場合が多い. t 分布の自由度はほぼ標本サイズになることが多いので, 標本サイズが大きな場合には複雑に見える t 分布の代わりに

正規分布を使っても大差ないことになる. このように, ざくっとどんぶり勘定することによって, 複雑に見える議論を単純化することは統計学を実用的に使うときに重要である.

3.13 t 分布と標準正規分布を同時プロット

In [52]:
```
function plot_tdist(ν)
    P = plot(x -> pdf(TDist(ν), x), -6, 6; label="", title="TDist($ν)")
    plot!(x -> pdf(Normal(), x), -6, 6; label="Normal()", ls=:dash)
end
```

Out[52]:
```
plot_tdist (generic function with 1 method)
```

In [53]:
```
plot(plot_tdist(1), plot_tdist(2); size=(800, 250))
```

Out[53]:

In [54]:
```
plot(plot_tdist(3), plot_tdist(4); size=(800, 250))
```

Out[54]:

In [55]:
```
plot(plot_tdist(10), plot_tdist(20); size=(800, 250))
```

Out[55]:

4 ベータ函数のガンマ函数表示

ベータ函数をガンマ函数で表せることから, ガンマ分布とベータ分布の関係が得られる.

$\alpha, \beta > 0$ について次が成立している:

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

これを示すためには

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

を示せばよい.

4.1 積分 $J[f]$ を導入とベータ函数のガンマ函数表示の証明

後で, ガンマ分布とベータ分布を関係付けるために, 函数 $f(z,t)$ について次の積分を考える:

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

このとき, 次が成立している:

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

もしもこの公式が証明されたならば, $J[f]$ の定義より, $f(z,t)=1$ のとき,

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

となり, 上の公式より,

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

以上を比較すれば $\Gamma(\alpha)\Gamma(\beta)=\Gamma(\alpha+\beta)B(\alpha,\beta)$ が得られる.

4.2 積分 $J[f]$ に関する公式の証明

前節で説明した $J[f]$ に関する公式は以下のようにして示される. 積分変数 y を $y=z-x$ によって z に変換すると, $0<y<\infty$ と $x<z<\infty$ のあいだの一対一対応が得られるので,

$$\begin{aligned}J[f]&=\int_0^\infty\left(\int_0^\infty f(x+y,x/(x+y))\,e^{-(x+y)}x^{\alpha-1}y^{\beta-1}\,dy\right)dx\\&=\int_0^\infty\left(\int_x^\infty f(z,x/z)\,e^{-z}\,x^{\alpha-1}(z-x)^{\beta-1}\,dz\right)dx.\end{aligned}$$

この二重積分で (x,z) の動く範囲は $0<x<z<\infty$ である. その範囲で (x,z) を動かすためには, まず $0<z<\infty$ を満たす z を与えてから, $0<x<z$ を満たす x を与えればよい. ゆえにその二重積分の積分順序を次のように交換できる:

$$J[f]=\int_0^\infty\left(\int_0^z f(z,x/z)\,e^{-z}\,x^{\alpha-1}(z-x)^{\beta-1}\,dx\right)dz.$$

積分変数 x を $x=zt$ によって t に変換すると, $0<x<z$ と $0<t<1$ のあいだの一対一対応が得られるので,

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

この二重積分の被積分函数の $f(z,t)$ 以外の因子は z,t のそれぞれの単独変数函数の積の形になっており, 次のように整理される:

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

これで示したい公式が示された.

4.3 問題: ガンマ分布とベータ分布の関係 (ベータ分布のガンマ分布表示)

$\alpha, \beta > 0$ と仮定する. 正の実数に値を持つ確率変数 X, Y の組 (X, Y) の確率密度函数は, ガンマ分布達 $\mathrm{Gamma}(\alpha, 1)$, $\mathrm{Gamma}(\beta, 1)$ の確率密度函数の積になっていると仮定する:

$$p(x, y) = \frac{e^{-x} x^{\alpha-1}}{\Gamma(\alpha)} \frac{e^{-y} y^{\beta-1}}{\Gamma(\beta)} = \frac{e^{-(x+y)} x^{\alpha-1} y^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \qquad (x, y > 0)$$

この $p(x, y)$ を確率変数 X, Y の **同時確率密度函数** (simultaneous probability density function)と呼ぶ. 同時確率密度函数がこのように各々の確率変数の確率密度函数の積になっているとき, それらの確率変数達は **独立** (independent)であるという. 以上の設定を次のように書くことにする:

$$X \sim \mathrm{Gamma}(\alpha, 1), \qquad Y \sim \mathrm{Gamma}(\beta, 1) \quad \text{(independent)}$$

このとき,

$$X + Y \sim \mathrm{Gamma}(\alpha + \beta, 1), \qquad \frac{X}{X+Y} \sim \mathrm{Beta}(\alpha, \beta) \quad \text{(independent)}$$

となることを示せ. すなわち, $Z = X + Y$, $T = X/(X+Y) = X/Z$ とおき, 分布 $\mathrm{Gamma}(\alpha + \beta, 1)$ の確率密度函数を $p_Z(z)$ と書き, 分布 $\mathrm{Beta}(\alpha, \beta)$ の確率密度函数を $p_T(t)$ と書くとき, 次が成立することを示せ:

$$E[f(Z, T)] = \int_0^\infty \left(\int_0^1 f(z, t) p_Z(z) p_T(t)\, dt \right) dz.$$

解答例: 分布 $\mathrm{Gamma}(\alpha + \beta, 1)$ の確率密度函数 $p_Z(z)$ と分布 $\mathrm{Beta}(\alpha, \beta)$ の確率密度函数を $p_T(t)$ はそれらの定義より次の形をしている:

$$p_Z(z) = \frac{e^{-z} z^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)}, \qquad p_T(t) = \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha, \beta)} \qquad (z > 0,\ 0 < t < 1)$$

確率変数の期待値の定義と前節の証明で使った $J[f]$ の定義を使うと,

$$\begin{aligned} E[f(Z, T)] &= \int_0^\infty \int_0^\infty f(x+y, x/(x+y)) p(x, y)\, dx\, dy \\ &= \int_0^\infty \int_0^\infty f(x+y, x/(x+y)) \frac{e^{-(x+y)} x^{\alpha-1} y^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}\, dx\, dy = \frac{J[f]}{\Gamma(\alpha)\Gamma(\beta)} =. \end{aligned}$$

$\Gamma(\alpha)\Gamma(\beta) = \Gamma(\alpha+\beta) B(\alpha, \beta)$ と前節で示した $J[f]$ に関する公式と $p_Z(z)$, $p_T(t)$ の定義を使うと,

$$\begin{aligned} E[f(Z, T)] &= \frac{J[f]}{\Gamma(\alpha+\beta) B(\alpha, \beta)} \\ &= \int_0^\infty \left(\int_0^1 f(z, t) \frac{e^{-z} z^{\alpha+\beta-1}}{\Gamma(\alpha+\beta)} \frac{t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha, \beta)}\, dt \right) dz \\ &= \int_0^\infty \left(\int_0^1 f(z, t) p_Z(z) p_T(t)\, dt \right) dz. \end{aligned}$$

これで示すべきことが示された.

解答終

注意: 以上においては計算を簡単にするためにガンマ分布のスケールパラメータを $\theta = 1$ とおいたが, $\theta > 0$ の一般的な場合にも,

$$X \sim \mathrm{Gamma}(\alpha, \theta) \iff \frac{X}{\theta} \sim \mathrm{Gamma}(\alpha, 1)$$

であることから,

$$X \sim \mathrm{Gamma}(\alpha, \theta), \qquad Y \sim \mathrm{Gamma}(\beta, \theta) \quad \text{(independent)}$$

ならば

$$X + Y \sim \mathrm{Gamma}(\alpha + \beta, \theta), \qquad \frac{X}{X+Y} \sim \mathrm{Beta}(\alpha, \beta) \quad \text{(independent)}$$

となることがわかる.

注意: 上の問題の結果はガンマ分布達 $\mathrm{Gamma}(\alpha, 1)$, $\mathrm{Gamma}(\beta, 1)$ に従う独立な確率変数 X, Y からガンマ分布 $\mathrm{Gamma}(\alpha + \beta, 1)$ に従う確率変数を $X + Y$ の形で作れてかつ, ベータ分布 $\mathrm{Beta}(\alpha, \beta)$ に従う確率変数を $X/(X+Y)$ の形で作れることを示した.

これは, ガンマ分布達 $\mathrm{Gamma}(\alpha, 1)$, $\mathrm{Gamma}(\beta, 1)$ に従う乱数達から, ガンマ分布 $\mathrm{Gamma}(\alpha + \beta, 1)$ に従う乱数とベータ分布 $\mathrm{Beta}(\alpha, \beta)$ に従う乱数を作れることを意味している.

上のように積分の煩雑な計算をするだけだと何をやっているのかがピンとこない人が多く出て来るかもしれないが, 確率変数達の以上のような関係は乱数の生成法のあいだの関係のことだと了解しておけば結果的に何をやっているのかわかりやすくなる.

そして, 以下のようにコンピュータによる計算によって数学的結果の正しさを数値的に確認することもできる.

ガンマ分布達の乱数を大量に生成.

In [56]:
```
α, β = 3, 5
n = 10^6
X = rand(Gamma(α, 1), n)
Y = rand(Gamma(β, 1), n);
```

ガンマ函数達の乱数達 X, Y の分布の確認.

In [57]:
```
P1 = histogram(X; norm=true, alpha=0.3, bin=140, xlim=(-1, 23), label="X")
plot!(Gamma(α, 1); label="Gamma($α, 1)", lw=2, ls=:dash)

P2 = histogram(Y; norm=true, alpha=0.3, bin=140, xlim=(-1, 23), label="Y")
plot!(Gamma(β, 1); label="Gamma($β, 1)", lw=2, ls=:dash)

plot(P1, P2; size=(800, 250))
```

Out[57]:

ガンマ函数の乱数達から Z と T を作成.

In [58]:
```
Z = @. X + Y
T = @. X / Z;
```

Z, T の分布達がそれぞれガンマ分布とベータ分布になっていることの確認.

In [59]:
```
P3 = histogram(Z; norm=true, alpha=0.3, bin=140, xlim=(-1, 23), label="Z = X + Y")
plot!(Gamma(α + β, 1); label="Gamma($α + $β, 1)", lw=2, ls=:dash)

P4 = histogram(T; norm=true, alpha=0.3, bin=140, xlim=(-0.02, 1.02), label="T = X / (X + Y)")
plot!(Beta(α, β); label="Beta($α, $β)", lw=2, ls=:dash)

plot(P3, P4; size=(800, 250))
```

Out[59]:

4.4 問題: ガンマ分布の再生成

以下を示せ:

(1) X がスケールパラメータ 1 のガンマ分布 Gamma$(\alpha, 1)$ に従う確率変数ならば θX はスケールパラメータ θ のガンマ分布に従う.

(2) **ガンマ分布の再生成:** 独立な確率変数達 X, Y のそれぞれが同じスケールパラメータのガンマ分布 Gamma(α, θ), Gamma(β, θ) に従うとき, $X + Y$ は形状パラメータがそれらの和であるようなガンマ分布 Gamma$(\alpha + \beta, \theta)$ に従う.

解答例: (1) X がスケールパラメータ 1 のガンマ分布 Gamma$(\alpha, 1)$ に従う確率変数であるとき,

$$\begin{aligned} E[f(\theta X)] &= \int_0^\infty f(\theta x) \frac{e^{-x} x^{\alpha-1}}{\Gamma(\alpha)}\, dx \\ &= \int_0^\infty f(y) \frac{e^{-y/\theta} (y/\theta)^{\alpha-1}}{\Gamma(\alpha)} \frac{dy}{\theta} = \int_0^\infty f(y) \frac{e^{-y/\theta} y^{\alpha-1}}{\theta^\alpha \Gamma(\alpha)}\, dy. \end{aligned}$$

ゆえに $Y = \theta X$ はスケール θ のガンマ分布 $\Gamma(\alpha, \theta)$ に従う.

(2) 独立な確率変数達 X, Y のそれぞれが同じスケールパラメータのガンマ分布 Gamma(α, θ), Gamma(β, θ) に従っていると仮定する.

このとき, (1)の結果を使うと, X, Y はスケールパラメータが 1 のガンマ分布達 Gamma$(\alpha, 1)$, Gamma$(\beta, 1)$ にそれぞれが従う独立な確率変数達 X', Y' によって, $X = \theta X'$, $Y = \theta Y'$ と表される.

ガンマ分布とベータ分布の関係の問題より, $X' + Y'$ はガンマ分布 Gamma$(\alpha + \beta, 1)$ に従うことがわかる.

ゆえに(1)の結果より, $X + Y = \theta(X' + Y')$ はガンマ分布 Gamma$(\alpha + \beta, \theta)$ に従う.

解答終

4.5 積分 $K[f]$ を導入とベータ函数のガンマ函数表示の別証明

函数 $g(z, u)$ に関する以下の積分を使って別証明を作っておく:

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

このとき, 次が成立している:

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

もしもこの公式が証明されたならば, $K[g]$ の定義より, $g(z, u) = 1$ のとき,

$$\begin{aligned} K[1] &= \int_0^\infty \int_0^\infty e^{-(x+y)} x^{\alpha-1} y^{\beta-1}\, dx\, dy \\ &= \int_0^\infty e^{-x} x^{\alpha-1}\, dx \cdot \int_0^\infty e^{-y} y^{\beta-1}\, dy = \Gamma(\alpha)\Gamma(\beta) \end{aligned}$$

となり, 上の公式より,

$$\begin{aligned} K[1] &= \int_0^\infty \left(\int_0^\infty e^{-z} z^{\alpha+\beta-1} \frac{u^{\alpha-1}}{(1+u)^{\alpha+\beta}}\, dt \right) dz \\ &= \int_0^\infty e^{-z} z^{\alpha+\beta-1}\, dz \cdot \int_0^\infty \frac{u^{\alpha-1}\, du}{(1+u)^{\alpha+\beta}} = \Gamma(\alpha+\beta) B(\alpha, \beta). \end{aligned}$$

以上を比較すれば $\Gamma(\alpha)\Gamma(\beta) = \Gamma(\alpha+\beta) B(\alpha, \beta)$ が得られる.

4.6 積分 $K[g]$ に関する公式の証明

前節で説明した $K[g]$ に関する公式は以下のようにして示される. 積分変数 x を $x = yu$ によって u に変換すると, $0 < x < \infty$ と $0 < u < \infty$ のあいだの一対一対応が得られるので,

$$\begin{aligned} K[g] &= \int_0^\infty \left(\int_0^\infty g(x + y, x/y)\, e^{-(x+y)} x^{\alpha-1} y^{\beta-1}\, dx \right) dy \\ &= \int_0^\infty \left(\int_0^\infty g((1+u)y, u)\, e^{-(1+u)y} u^{\alpha-1} y^{\alpha+\beta-1}\, du \right) dy. \end{aligned}$$

この二重積分の積分順序を次のように交換すると,

$$K[g] = \int_0^\infty \left(\int_0^\infty g((1+u)y, u)\, e^{-(1+u)y} u^{\alpha-1}\, y^{\alpha+\beta-1}\, dy \right) du.$$

積分変数 y を $y = z/(1+u)$ によって z に変換すると,

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

この二重積分の被積分函数の $g(z, u)$ 以外の因子は z, u のそれぞれの単独変数函数の積の形になっており, 次のように整理される:

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

これで示したい公式が示された.

$J[f]$ に関する結果を使った別証明: $f(z, t) = g(z, t/(1-t))$ とおくと, $u = t/(1-t) \iff t = u/(1+u)$ より, $g(z, u) = f(z, u/(1+u))$ なので,

$$\begin{aligned} K[g] = J[f] &= \int_0^\infty \int_0^\infty f\left(x+y, \frac{x}{x+y} \right) e^{-(x+y)} x^{\alpha-1} y^{\beta-1}\, dx\, dy \\ &= \int_0^\infty \left(\int_0^1 f(z, t)\, e^{-z} z^{\alpha+\beta-1}\, t^{\alpha-1} (1-t)^{\beta-1}\, dt \right) dz. \end{aligned}$$

$t = u/(1+u) = 1 - 1/(1+u)$, $dt = du/(1+u)^2$ とおくと, $0 < t < 1 \iff 0 < u < \infty$ なので,

$$\begin{aligned} K[g] = J[f] &= \int_0^\infty \left(\int_0^\infty f(z, u/(1+u))\, e^{-z} z^{\alpha+\beta-1} \left(\frac{u}{1+u} \right)^{\alpha-1} \left(\frac{1}{1+u} \right)^{\beta-1} \frac{du}{(1+u)^2} \right) dz \\ &= \int_0^\infty \left(\int_0^\infty g(z, u)\, e^{-z} z^{\alpha+\beta-1} \frac{u^{\alpha-1}}{(1+u)^{\alpha+\beta}}\, du \right) dz \\ &= \int_0^\infty \int_0^\infty g(z, u)\, e^{-z} z^{\alpha+\beta-1} \frac{u^{\alpha-1}}{(1+u)^{\alpha+\beta}}\, dz\, du. \end{aligned}$$

4.7 ベータプライム分布の定義

$\alpha, \beta > 0$ であると仮定する. 確率密度函数

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

によって定義される連続分布をパラメータ (α, β) の **ベータプライム分布** (Beta prime distribution)と呼び,

$$\mathrm{BetaPrime}(\alpha, \beta)$$

と表す.

ベータプライム分布はオッズに座標変換したベータ分布である. $T \sim \mathrm{Beta}(\alpha, \beta)$ のとき, そのオッズ $U = T/(1-T)$ は分布 $\mathrm{BetaPrime}(\alpha, \beta)$ に従い, 逆に $U \sim \mathrm{BetaPrime}(\alpha, \beta)$ のとき, $T = U/(1+U)$ は分布 $\mathrm{Beta}(\alpha, \beta)$ に従う. この結果は本質的にベータ分布の2つの異なる表示の存在

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

と同じことである.

ベータプライム分布と以下で説明する F 分布はスケール変換の違いを除けば同じ分布である. 一般に F 分布の方が使われることが多いように思われるので, 詳しい説明は F 分布の方ですることにし, ベータプライム分布については定義をするだけで詳しい説明を省略することにする.

In [60]:

```
PP = []
α = 0.5
for β in  (1.5, 5, 15)
    bp = BetaPrime(α, β)
    P = plot(bp, 0, α/β + 4√(α*(α+β)/β^2); label="BetaPrime(α, β)")
    α < 1 && plot!(; ylim = (-0.2, 4.2))
    title!("α = $α, β = $β")
    push!(PP, P)
end
plot(PP...; size=(800, 200), layout=(1, 3))
```

Out[60]:

In [61]:

```
PP = []
α = 2
for β in  (1.5, 5, 15)
    bp = BetaPrime(α, β)
    P = plot(bp, 0, α/β + 4√(α*(α+β)/β^2); label="BetaPrime(α, β)")
    α < 1 && plot!(; ylim = (-0.2, 4.2))
    title!("α = $α, β = $β")
    push!(PP, P)
end
plot(PP...; size=(800, 200), layout=(1, 3))
```

Out[61]:

4.8 F 分布の定義

$v_1, v_2 > 0$ であると仮定する. 確率密度函数

$$p(w|v_1, v_2) = \frac{1}{B(v_1/2, v_2/2)} \left(\frac{v_1 w}{v_1 w + v_2} \right)^{v_1/2} \left(1 - \frac{v_1 w}{v_1 w + v_2} \right)^{v_2/2} w^{-1} \quad (w > 0)$$

によって定義される $w > 0$ に関する連続分布を自由度 (v_1, v_2) の F **分布** (F-distribution)と呼び,

$$\mathrm{FDist}(v_1, v_2)$$

と表す. F 分布はベータプライム分布のスケール変換に過ぎないが, こちらの方がよく使われているように思われる.

4.9 F 分布とベータ分布やガンマ分布やχ²分布との関係

$\nu_1, \nu_2, \theta > 0$ であるとし,

$$X \sim \mathrm{Gamma}(\nu_1/2, \theta), \quad Y \sim \mathrm{Gamma}(\nu_2/2, \theta) \quad \text{(independent)}$$

のとき,

$$T = \frac{X}{X+Y}, \quad W = \frac{\nu_2}{\nu_1}\frac{T}{1-T} = \frac{X/\nu_1}{Y/\nu_2}$$

とおくと,

$$T \sim \mathrm{Beta}(\nu_1/2, \nu_2/2), \quad W \sim \mathrm{FDist}(\nu_1, \nu_2).$$

特に $\theta = 1/2$ のとき, $\mathrm{Chisq}(\nu) = \mathrm{Gamma}(\nu/2, 2)$ だったので,

$$X \sim \mathrm{Chisq}(\nu_1), \quad Y \sim \mathrm{Chisq}(\nu_2) \quad \text{(independent)}$$

の場合にも上と同じ結論が成立する.

証明: F 分布の確率密度函数の形は煩雑に見えるので, 見たことがあるよりシンプルな式に変形しよう. オッズ u と確率 t の関係

$$u = \frac{\nu_1}{\nu_2}w, \quad t = \frac{u}{1+u} = \frac{\nu_1 w}{\nu_1 w + \nu_2} = \frac{\nu_1 w}{\nu_1 w + \nu_2}, \quad 1 - t = \frac{1}{1+u}$$

を念頭に置きながら整理し直すと, F 分布の確率密度函数は以下のように変形される:

$$\begin{aligned} p(w|\nu_1, \nu_2) &= \frac{1}{B(\nu_1/2, \nu_2/2)}\left(\frac{\nu_1 w/\nu_2}{1+\nu_1 w/\nu_2}\right)^{\nu_1/2}\left(\frac{1}{1+\nu_1 w/\nu_2}\right)^{\nu_2/2}\left(\frac{\nu_1}{\nu_2}w\right)^{-1}\frac{\nu_2}{\nu_1} \\ &= \frac{1}{B(\nu_1/2, \nu_2/2)}\frac{(\nu_1 w/\nu_2)^{\nu_1-1}}{(1+\nu_1 w/\nu_2)^{(\nu_1+\nu_2)/2}}\frac{\nu_2}{\nu_1} \end{aligned}$$

これより,

$$w = \frac{\nu_1}{\nu_2}u, \quad \alpha = \frac{\nu_1}{2}, \quad \beta = \frac{\nu_1}{2}$$

とおくと,

$$p(w|\nu_1, \nu_2)\,dw = \frac{1}{B(\nu_1/2, \nu_2/2)}\frac{u^{\alpha-1}\,du}{(1+u)^{\alpha+\beta}}.$$

これはすでに何度も見たことがある形の式である! ゆえに $u = t/(1-t)$ とおくと,

$$\frac{u^{\alpha-1}\,du}{(1+u)^{\alpha+\beta}} = t^{\alpha-1}(1-t)^{\beta-1}\,dt$$

となることをすでに知っており,

$$w = \frac{\nu_1}{\nu_2}u = \frac{\nu_1}{\nu_2}\frac{t}{1-t}, \quad \alpha = \frac{\nu_1}{2}, \quad \beta = \frac{\nu_1}{2}$$

とおくと, 次が成立することがわかった:

$$p(w|\nu_1, \nu_2)\,dw = \frac{1}{B(\nu_1/2, \nu_2/2)}t^{\alpha-1}(1-t)^{\beta-1}\,dt.$$

この公式は, ベータ分布に従う確率変数 $T \sim \mathrm{Beta}(\alpha, \beta)$ について

$$W = \frac{\nu_1}{\nu_2}\frac{T}{1-T} = \frac{T/\nu_1}{(1-T)/\nu_2}, \alpha = \frac{\nu_1}{2}, \quad \beta = \frac{\nu_1}{2}$$

とおくと, $W \sim \mathrm{FDist}(\nu_1, \nu_2)$ が成立することを意味している.

さらに,

$$X \sim \mathrm{Gamma}(\alpha, \theta), \quad Y \sim \mathrm{Gamma}(\beta, \theta) \quad \text{(independent)}$$

のとき,

$$X + Y \sim \mathrm{Gamma}(\alpha+\beta, \theta), \quad \frac{X}{X+Y} \sim \mathrm{Beta}(\alpha, \beta) \quad \text{(independent)}$$

となることがすでにわかっているので, $T = X/(X+Y)$ とおくと,

$$W = \frac{T/\nu_1}{(1-T)/\nu_2} = \frac{X/\nu_1}{Y/\nu_2}\alpha = \frac{\nu_1}{2}, \quad \beta = \frac{\nu_1}{2}$$

とおくと, $W \sim \mathrm{FDist}(\nu_1, \nu_2)$ となることもわかる.

特に $\theta = 1/2$ とおくと,

$$X \sim \mathrm{Chisq}(\nu_1), \quad Y \sim \mathrm{Gamma}(\nu_2) \quad \text{(independent)}$$

ならば

$$W = \frac{X/\nu_1}{Y/\nu_2} \sim \mathrm{FDist}(\nu_1, \nu_2)$$

となることがわかる.

証明終

注意: $X_\nu \sim \mathrm{Chisq}(\nu)$ のとき, $\nu \to \infty$ で X_ν/ν の分布は 1 に集中して行く.

In [62]:

```
PP = []
for ν in (10, 100, 1000)
    P = plot(x -> pdf(Chisq(ν), ν*x)*ν, 0, 2; label="", title="Chisq($ν)/$ν")
    push!(PP, P)
end
plot(PP...; size=(800, 200), layout=(1, 3))
```

Out[62]:

ゆえに $X_1 \sim \mathrm{Chisq}(\nu_1)$, $X_2 \sim \mathrm{Chisq}(\nu_2)$ でそれらが独立なとき, $W = (X_1/\nu_1)/(X_2/\nu_2)$ が従う分布は ν_2 が大きなとき, X_1/ν_1 がしたがう分布で近似される. ν_2 が大きくないとき, F 分布は裾が太い分布になる.

In [63]:

```
PP = []
ν₁ = 4
for ν₂ in  (3, 10, 30)
    P = plot(FDist(ν₁, ν₂), 0, 1+5√(2/ν₁); label="FDist(ν₁, ν₂)")
    plot!(x -> pdf(Chisq(ν₁), ν₁*x)*ν₁, 0, 1+5√(2/ν₁); label="Chisq(ν₁)/ν₁", ls=:dash)
    ν₁ < 2 && plot!(; ylim = (-0.1, 1.1))
    title!("ν₁ = $ν₁, ν₂ = $ν₂")
    push!(PP, P)
end
plot(PP...; size=(800, 200), layout=(1, 3))
```

Out[63]:

$T \sim \mathrm{TDist}(\nu) \implies T^2 \sim \mathrm{Fdist}(1, \nu)$ でかつ $Z \sim \mathrm{Normal}(0, 1) \implies Z^2 \sim \mathrm{Chisq}(1)$ なので, 上で述べた結果は t 分布 $\mathrm{TDist}(\nu)$ が ν を大きくすると標準正規分布に近付くことや, ν が大きくないとき裾が太い分布になることの一般化になっている.

In [64]:

```
PP = []
for ν in  (1, 3, 10)
    P = plot(FDist(1, ν), 0, 5; label="FDist(1, ν)")
    plot!(x → pdf(TDist(ν), √x)/√x, 0, 5; label="TDist(ν)²", ls=:dash)
    plot!(x → pdf(Chisq(1), x), 0, 5; label="Chisq(1)", ls=:dashdot)
    plot!(; ylim = (-0.1, 1.1))
    title!("ν = $ν")
    push!(PP, P)
end
plot(PP...; size=(800, 200), layout=(1, 3))
```

Out[64]:

4.10 問題: ガンマ函数のGaussの乗法公式

ベータ函数のガンマ函数表示とガンマ函数の函数等式とベータ函数の極限でガンマ函数を表せることを用いて, 次の公式を証明せよ:

$$\lim_{n\to\infty} \frac{n^\alpha\, n!}{\alpha(\alpha+1)\cdots(\alpha+n)} = \Gamma(\alpha).$$

解答例: ガンマ函数の函数等式より $\Gamma(\alpha+n+1) = (\alpha+n)\cdots(\alpha+1)\alpha\Gamma(\alpha)$ なので, $\Gamma(n+1) = n!$ とベータ函数のガンマ函数表示より,

$$n^\alpha B(\alpha, n+1) = \frac{n^\alpha\Gamma(\alpha)\Gamma(n+1)}{\Gamma(\alpha+n+1)} = \frac{n^\alpha\Gamma(\alpha)n!}{(\alpha+n)\cdots(\alpha+1)\alpha\Gamma(\alpha)} = \frac{n^\alpha\, n!}{\alpha(\alpha+1)\cdots(\alpha+n)}.$$

ベータ函数の極限でガンマ函数が表せることより, $n \to \infty$ のとき,

$$\frac{n^\alpha\, n!}{\alpha(\alpha+1)\cdots(\alpha+n)} = n^\alpha B(\alpha, n+1) \to \Gamma(\alpha)$$

が得られる.

解答終

注意: 上の公式をガンマ函数に関する **Gaussの乗法公式** と呼ぶ. 上の公式の極限は α が0以下の整数以外の任意の複素数であっても収束している.

4.11 問題: sin の無限積表示

$f_\alpha(x)$ は $\cos(\alpha x)$ の $-\pi \le x \le \pi$ への制限の定義域を周期 2π で $\mathbb{R}$ 全体に拡張して得られる連続函数であるとする. このとき, $f_\alpha(x)$ の周期 2π でのFourier級数は $f_n(x)$ に収束することを認めて, 次の sin の無限積表示を示せ:

$$\sin(\pi\alpha) = \pi\alpha \prod_{k=1}^{\infty}\left(1 - \frac{\alpha^2}{k^2}\right).$$

解答例: 偶函数 $f_\alpha(x)$ の周期 2π のFourier級数を計算しよう. $k = 0, 1, 2, \ldots$ について,

$$\begin{aligned} a_k := \int_{-\pi}^{\pi} f_\alpha(x)\cos(kx)\,dx &= \int_{-\pi}^{\pi}\cos(\alpha x)\cos(kx)\,dx = \frac{1}{2}\int_{-\pi}^{\pi}(\cos((\alpha+k)x) + \cos((\alpha-k)x))\,dx \\ &= \frac{1}{2}\left[\frac{\sin((\alpha+k)x)}{\alpha+k} + \frac{\sin((\alpha-k)x)}{\alpha-k}\right]_{-\pi}^{\pi} = (-1)^k\sin(\pi\alpha)\frac{2\alpha}{\alpha^2-k^2}. \end{aligned}$$

$\int_{-\pi}^{\pi}\cos^2(0x)\,dx = 2\pi$ でかつ $k = 1, 2, \ldots$ のとき

$$\int_{-\pi}^{\pi} \cos^2(kx)\,dx = \int_{-\pi}^{\pi} \frac{\cos(2kx)+1}{2}\,dx = \int_{-\pi}^{\pi} \frac{1}{2}\,dx = \pi.$$

ゆえに, $-\pi \le x \le \pi$ のとき,

$$\cos(\alpha x) = f_\alpha(x) = \frac{a_0}{2\pi} + \frac{1}{\pi}\sum_{k=1}^{\infty} a_k \cos(kx) = \sin(\pi\alpha)\left(\frac{1}{2\pi}\frac{2}{\alpha} + \frac{1}{\pi}\sum_{k=1}^{\infty}(-1)^k \frac{2\alpha}{\alpha^2-k^2}\cos(kx)\right).$$

両辺を $\sin(\pi\alpha)/\pi$ で割って, $x=\pi$ とおくと, $\cos(k\pi) = (-1)^k$ より,

$$\pi\cot(\pi\alpha) = \frac{1}{s} + \sum_{k=1}^{\infty} \frac{2\alpha}{\alpha^2-k^2}. \tag{$*$}$$

sin の無限積表示

$$\frac{\sin(\pi\alpha)}{\pi\alpha} = \prod_{k=1}^{\infty}\left(1-\frac{\alpha^2}{k^2}\right)$$

を証明しよう. $\alpha \to 0$ とするとこの両辺はどちらも 1 に等しくなる. ゆえに両辺の対数の導函数が等しければ, 両辺が等しいことがわかる. 左辺の対数の導函数は

$$\frac{\pi\cos(\pi\alpha)}{\sin(\pi\alpha)} - \frac{\pi}{\pi\alpha} = \pi\cot(\pi\alpha) - \frac{1}{\alpha}.$$

右辺の対数の導函数は

$$\sum_{k=1}^{\infty} \frac{-2\alpha/k^2}{1-\alpha^2/k^2} = \sum_{k=1}^{\infty} \frac{2\alpha}{\alpha^2-k^2}.$$

ゆえに($*$)より, 両辺の対数の導函数は互いに等しい. これで sin の無限積表示が証明された.

解答終

4.12 問題: Wallisの公式

Gaussの乗法公式と sin の無限積表示

$$\lim_{n\to\infty} \frac{n^\alpha\, n!}{\alpha(\alpha+1)\cdots(\alpha+n)} = \Gamma(\alpha), \quad \prod_{k=1}^{\infty}\left(1-\frac{\alpha^2}{k^2}\right) = \frac{\sin(\pi\alpha)}{\pi\alpha}$$

を使って, 次の公式を示せ:

$$\lim_{n\to\infty} \frac{2^{2n}(n!)^2}{\sqrt{n}\,(2n)!} = \sqrt{\pi}, \quad \prod_{k=1}^{\infty} \frac{2k\cdot 2k}{(2k-1)(2k+1)} = \frac{\pi}{2}.$$

これらを**Wallisの公式** と呼ぶ.

解答例: $\alpha = 1/2$ とおけばよい. 分子分母に $2\cdot 4\cdots(2n) = 2^n n!$ をかけることによって,

$$\frac{1}{2}\left(\frac{1}{2}+1\right)\cdots\left(\frac{1}{2}+n\right) = \frac{1\cdot 3\cdots(2n-1)}{2^n}\left(\frac{1}{2}+n\right) = \frac{(2n)!}{2^{2n}n!}\left(\frac{1}{2}+n\right).$$

Gaussの乗法公式で極限を取る前の式で $\alpha = 1/2$ とおくと,

$$\frac{\sqrt{n}\,n!}{(1/2)(1/2+1)\cdots(1/2+n)} = \frac{\sqrt{n}\,2^{2n}(n!)^2}{(2n)!(1/2+n)} = \frac{2^{2n}(n!)^2}{\sqrt{n}\,(2n)!}\frac{n}{1/2+n}$$

なので,

$$\lim_{n\to\infty} \frac{2^{2n}(n!)^2}{\sqrt{n}\,(2n)!} = \Gamma(1/2) = \sqrt{\pi}.$$

sin の無限積表示の各因子は $\alpha = 1/2$ とおくと,

$$1 - \frac{(1/2)^2}{k^2} = \frac{(2k)^2-1}{(2k)^2} = \frac{(2k-1)(2k+1)}{2k\cdot 2k}$$

になるので,

$$\prod_{k=1}^{\infty} \frac{(2k-1)(2k+1)}{2k \cdot 2k} = \frac{\sin(\pi/2)}{\pi/2} = \frac{2}{\pi}.$$

これの逆数を取ればもう1つのWallisの公式も得られる.

解答終

以下はWallisの公式の数値的な確認. 階乗を数値計算で扱うときには常に $\log n!$ を扱うようにし, かけ算やわり算を直接的に使わずに対数を取ったものの和や差を計算して, 最後に指数函数を作用させる必要がある.

In [65]:
```
n = 10^6
@show n
@show exp(2n*log(2) + 2logfactorial(n) - (1/2)*log(n) - logfactorial(2n))^2
@show 2prod(2k*2k/((2k-1)*(2k+1)) for k in 1:n);
```

```
n = 1000000
exp((((2n) * log(2) + 2 * logfactorial(n)) - (1 / 2) * log(n)) - logfactorial(2n)) ^ 2 = 3.14159
34154213083
2 * prod((((2k) * (2k)) / ((2k - 1) * (2k + 1)) for k = 1:n)) = 3.141591868192149
```

4.13 問題: ガンマ函数の相反公式(reflection formula)

sin の無限積表示

$$\sin(\pi\alpha) = \pi\alpha \prod_{k=1}^{\infty} \left(1 - \frac{\alpha^2}{k^2}\right)$$

を使って, ガンマ函数のGaussの乗法公式から次の公式を導け:

$$B(\alpha, 1-\alpha) = \Gamma(\alpha)\Gamma(1-\alpha) = \frac{\pi}{\sin(\pi\alpha)}.$$

これをガンマ函数の **相反公式** (reflection formula)と呼ぶ.

解答例: 相反公式の両辺に α をかけて, $\alpha\Gamma(\alpha) = \Gamma(1+\alpha)$ を使った後に, 逆数を取ることによって得られる次の公式を示せば十分である:

$$\frac{1}{\Gamma(1+\alpha)\Gamma(1-\alpha)} = \frac{\sin(\pi\alpha)}{\pi\alpha}.$$

sin の無限積表示を認めて使うと, この公式は次と同値である:

$$\frac{1}{\Gamma(1+\alpha)\Gamma(1-\alpha)} = \prod_{k=1}^{\infty} \left(1 - \frac{\alpha^2}{k^2}\right).$$

これをGaussの乗法公式から導けばよい.

Gaussの情報公式より, 次の式は $n \to \infty$ で $1/(\Gamma(1+\alpha)\Gamma(1-\alpha))$ に収束する:

$$\begin{aligned}
&\frac{(1+\alpha)(2+\alpha)\cdots(n+1+\alpha)}{n^{1+\alpha}\, n!} \frac{(1-\alpha)(2-\alpha)\cdots(n+1-\alpha)}{n^{1-\alpha}\, n!} \\
&= \frac{(n+1)^2}{n^2}\left(1+\frac{\alpha}{1}\right)\left(1+\frac{\alpha}{2}\right)\cdots\left(1+\frac{\alpha}{n+2}\right)\left(1-\frac{\alpha}{1}\right)\left(1-\frac{\alpha}{2}\right)\cdots\left(1-\frac{\alpha}{n+2}\right) \\
&= \frac{(n+1)^2}{n^2} \prod_{k=1}^{n+1} \left(1 - \frac{\alpha^2}{k^2}\right)
\end{aligned}$$

最初の等号で, 分母の n^{α} と $n^{-\alpha}$ をキャンセルさせて, 分子分母を $(n+1)! \times (n+1)!$ で割った. 上の最後の式は $n \to \infty$ で次に

$$\prod_{k=1}^{\infty} \left(1 - \frac{\alpha^2}{k^2}\right)$$

に収束する. これで示すべきことが示された.

解答終

4.14 ディガンマ函数, トリガンマ函数, ポリガンマ函数, Hurwitzのゼータ函数

統計学関係の計算ではガンマ函数やベータ函数の導函数達が必要になる場合がある. ベータ函数はガンマ函数で表示されるのでベータ函数の偏導函数の計算はガンマ函数の導函数の計算に帰着する. ガンマ函数の導函数が必要になる場合として, ガンマ分布の最尤法や二項分布モデルやPoisson分布モデルのベイズ統計における情報量規準の計算などがある.

コンピュータでの基本特殊函数ライブラリの中にはガンマ函数の導函数の計算のために必要な **ディガンマ函数** (digamma function), **トリガンマ函数** (trigamma function), **ポリガンマ函数** (polygamma function)が含まれている. それらはガンマ函数そのものではなく, ガンマ函数の対数の導函数として定義される. ディガンマ函数 $\psi(\alpha)$ は

$$\psi(\alpha) = \frac{d}{d\alpha}\log\Gamma(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

と定義される. これの導函数

$$\psi'(\alpha) = \frac{\Gamma(\alpha)\Gamma''(\alpha) - \Gamma'(\alpha)^2}{\Gamma(\alpha)^2}$$

をトリガンマ函数と呼び, 高階の導函数 $\psi^{(k)}(\alpha)$ をポリガンマ函数と呼ぶ.

ガンマ函数の函数等式 $\gamma(\alpha+1) = \alpha\Gamma(\alpha)$ の両辺の対数を微分することによって,

$$\psi(\alpha+1) = \psi(\alpha) + \frac{1}{\alpha},\ \psi'(\alpha+1) = \psi'(\alpha) - \frac{1}{\alpha^2},\ \psi''(\alpha+1) = \psi''(\alpha) + \frac{2}{\alpha^3},\ \dots \tag{$*$}$$

ガンマ函数のGaussの乗法公式

$$\lim_{n\to\infty}\frac{n^\alpha\, n!}{\alpha(\alpha+1)\cdots(\alpha+n)} = \Gamma(\alpha).$$

と

$$\frac{d}{d\alpha}\log\frac{n^\alpha\, n!}{\alpha(\alpha+1)\cdots(\alpha+n)} = \log n - \sum_{k=0}^{n}\frac{1}{\alpha+k}$$

より,

$$\psi(\alpha) = \lim_{n\to\infty}\left(\log n - \sum_{k=0}^{n}\frac{1}{\alpha+k}\right),\ \psi'(\alpha) = \sum_{k=0}^{\infty}\frac{1}{(\alpha+k)^2},\ \psi''(\alpha) = -2\sum_{k=0}^{\infty}\frac{1}{(\alpha+k)^3},\ \dots \tag{$**$}$$

この($**$)から上の($*$)が成立することを確認することもできる.

調和級数 $\sum_{n=1}^{N} 1/n$ は $N\to\infty$ で無限大に発散するが, それから $\log N$ を引いたものは有限の値 $\gamma \approx 0.5772$ に収束することが知られている. この γ を **Eulerの** γ (オイラーの γ, Euler's gamma)と呼ぶことがある. ($**$)の最初の式より,

$$-\psi(1) = \lim_{N\to\infty}\left(\sum_{n=1}^{N+1}\frac{1}{n} - \log N\right) = \lim_{N\to\infty}\left(\sum_{n=1}^{N}\frac{1}{n} - \log N\right) = \gamma \approx 0.5772.$$

これと($*$)の最初の式から, 次が得られる:

$$\psi(n+1) = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{n} - \gamma \quad (n = 0, 1, 2, \dots)$$

ガンマ函数は階乗 $\Gamma(n+1) = n!$ の一般化になっていたが, ディガンマ函数はEulerの γ の情報を含む調和級数の一般化になっていると考えられる.

Hurwitzのゼータ函数 (フルヴィッツのゼータ函数)が

$$\zeta(s,\alpha) = \sum_{k=0}^{\infty}\frac{1}{(\alpha+k)^s}$$

と定義される. トリガンマ函数以降のポリガンマ函数達は, 定数倍の違いを除いて, Hurwitzのゼータ函数で s を2以上の整数にした場合になっている:

$$\psi^{(k)}(\alpha) = (-1)^{k+1}k!\,\zeta(k+1,\alpha) \quad (k = 1, 2, 3, \dots)$$

$\zeta(s) = \zeta(s,1) = \sum_{n=1}^{\infty} n^{-s}$ は **Riemannのゼータ函数** (リーマンのゼータ函数)と呼ばれている. ゼータ函数達は数論の基本的研究対象である. ガンマ函数のよく使われる形から得られる

$$\frac{1}{(\alpha+k)^s} = \frac{1}{\Gamma(s)}\int_0^\infty e^{-(\alpha+k)t}t^{s-1}\,dt$$

という公式を使えば, Hurwitzのゼータ函数はガンマ函数のような積分表示を持つこともわかる:

$$\zeta(s,\alpha) = \frac{1}{\Gamma(s)}\int_0^\infty \sum_{k=0}^\infty e^{-(\alpha+k)t}t^{s-1}\,dt = \frac{1}{\Gamma(s)}\int_0^\infty \frac{e^{-\alpha t}}{1-e^{-t}}t^{s-1}\,dt.$$

ガンマ函数の対数(対数ガンマ函数 (logarithmic Gamma function))は, 定数差を除いて, Hurwitzのゼータ函数の $s=0$ での微係数に等しいことも知られている(**Lerchの定理**(レルヒの定理)):

$$\log\Gamma(\alpha) = \zeta_s(0,\alpha) + \log\sqrt{2\pi}.$$

この公式を使うと階乗のStirling近似の精密化(**Binetの公式**)も自然に得られる:

$$\varphi(x) = \int_0^\infty \left(\frac{1}{e^t-1} - \frac{1}{t} + \frac{1}{2}\right)e^{-xt}t^{-1}\,dt.$$

とおくと,

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

これはStirlingの公式 $\Gamma(n+1) = n! \approx n^n e^{-n}\sqrt{2\pi n}$ から得られる近似

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

の精密化になっており, この近似の誤差が $\varphi(n)$ と書けることを意味している.

以上のようなことに興味がある人は次のリンク先を参照せよ:

- [10 Gauss積分, ガンマ函数, ベータ函数 (https://nbviewer.org/github/genkuroki/Calculus/blob/master/10%20Gauss%2C%20Gamma%2C%20Beta.ipynb#Lerch%E3%81%](https://nbviewer.org/github/genkuroki/Calculus/blob/master/10%20Gauss%2C%20Gamma%2C%20Beta.ipynb#Lerch%E3%81%)

対数ガンマ函数とディガンマ函数とトリガンマ函数のグラフ:

In [66]:
```
plot(loggamma, eps(), 8; label="y = log Γ(x)", ylim=(-0.5, 10.2))
plot!(x -> x*log(x) - x - (1/2)*log(x) + log(√(2π));
    label="Stirling approx.", ls=:dash)
plot!(; legend=:top)
```

Out[66]:

In [67]:
```
plot(digamma, 0.1, 8; label="y = ψ(x) = digamma(x)")
plot!(x -> log(x) - 1/(2x); label="y = log x - 1/(2x)", ls=:dash)
plot!(; legend=:bottomright)
```

Out[67]:

In [68]:
```
plot(trigamma, 0.1, 8; label="y = ψ'(x) = trigamma(x)", ylim=(-0.2, 10.2))
plot!(x -> 1/x + 1/(2x^2); label="y = 1/x + 1/(2x²)", ls=:dash)
```

Out[68]:

5 Dirichlet積分=多変数ベータ函数

n 次元球体の体積を求める問題は, $a_i, p_i > 0$ に関する

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

の体積を求める問題に一般化される. ここで $\mathbb{R}_{>0}$ は正の実数全体の集合を表す.

さらにこの問題は $\alpha_i > 0$ も与えた場合の次の積分を計算する問題に一般化される:

$$V = \int \cdots \int_X x_1^{\alpha_1 - 1} \cdots x_n^{\alpha_n - 1} \, dx_1 \cdots dx_n$$

$\alpha_1 = \cdots = \alpha_n = 1$ のときこの積分は X の体積になる. この問題の解答は

$$V = \frac{a_1^{\alpha_1} \cdots a_n^{\alpha_n}}{p_1 \cdots p_n} \frac{\Gamma(\alpha_1/p_1) \cdots \Gamma(\alpha_n/p_n)}{\Gamma(\alpha_1/p_1 + \cdots + \alpha_n/p_n + 1)}.$$

例えば, $\alpha_1 = \cdots = \alpha_n = 1$, $a_1 = \cdots = a_n = 1$, $p_1 = \cdots = p_n = 2$ のとき, V の 2^n 倍の

$$2^n V = \frac{\Gamma(1/2)^n}{\Gamma(n/2 + 1)} = \frac{\pi^{n/2}}{\Gamma(n/2 + 1)}$$

は確かに n 次元単位球体の体積になっている.

このような一般化にはどのような価値があるのだろうか. 上の問題の解答はDirichlet(ディリクレ)の1839年の論文

- Lejeune-Dirichlet. Sur une nouvelle méthode pour la détermination des intégrales multiples. Journal de Mathématiques Pures et Appliquées (1839), 164-168. (link (https://eudml.org/doc/235403))

で与えられている. その論文でDirichletはベータ函数の多変数化であるDirichlet積分を発見した. Dirichlet積分はカテゴリカル分布と多項分布の共役事前分布であるDirichlet分布の「分母」となる統計学的にも重要な積分である. 統計学的にも重要なDirichlet積分は

- 高木貞治『解析概論』第8章第96節例2, pp.358-359
- 寺沢寛一『自然科学者のための 数学概論 [増訂版]』第5.21節[4], pp.216-217

でも紹介されている. これらはガンマ函数入門のための文献としても優れている.

5.1 Dirichlet積分=多変数ベータ函数の定義とそのガンマ函数表示

n 次元の領域 Δ_n を次のように定める:

$$\Delta_n = \{ (t_1, \dots, t_n) \in \mathbb{R}^n_{>0} \mid t_1 + \cdots + t_n < 1 \}.$$

ここで $\mathbb{R}_{>0}$ は正の実数全体の集合を表す.

例えば, $\Delta_1 = \{ t_1 \mid 0 < t_1 < 1 \}$ は開区間になり, Δ_2 は頂点 $(1,0), (0,1), (0,0)$ を持つ直角二等辺三角形の内側になり, Δ_3 は頂点 $(1,0,0), (0,1,0), (0,0,1), (0,0,0)$ を持つ四面体の内側になる. (Δ_n を n **次元開単体** と呼ぶことがある.)

Dirichlet積分 とは次の積分のことである:

$$B(\alpha_1, \dots, \alpha_{n+1}) = \int \cdots \int_{\Delta_n} t_1^{\alpha_1 - 1} \cdots t_n^{\alpha_n - 1} (1 - (t_1 + \cdots + t_n))^{\alpha_{n+1} - 1} \, dt_1 \cdots dt_n \quad (\alpha_i > 0)$$

$B(\alpha_1, \dots, \alpha_{n+1})$ を **多変数ベータ函数** とも呼ぶことにする. 多変数ベータ函数は次のガンマ函数表示を持つ.

$$B(\alpha_1, \dots, \alpha_{n+1}) = \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{n+1})}{\Gamma(\alpha_1 + \cdots + \alpha_{n+1})}.$$

5.2 Dirichlet分布の定義

$\mathbb{R}_{>0}$ は正の実数全体の集合を表すとし, $\alpha_1, \dots, \alpha_r \in \mathbb{R}_{>0}$ と仮定する. このとき,

$$\Delta_{r-1} = \{ (t_1, \dots, t_{r-1}) \in \mathbb{R}_{>0}^{r-1} \mid t_1 + \cdots + t_{r-1} < 1 \}$$

の上で定義された確率密度函数を

$$p(t_1, \dots, t_{r-1} | \alpha_1, \dots, \alpha_r) = \frac{t_1^{\alpha_1 - 1} \cdots t_{r-1}^{\alpha_{r-1} - 1} (1 - (t_1 + \cdots + t_{r-1}))^{\alpha_r - 1}}{B(\alpha_1, \dots, \alpha_r)}$$

と定めることができる. この密度函数が定める多変量連続分布をDirichlet分布と呼び,

$$\mathrm{Dirichlet}(\alpha_1, \dots, \alpha_r)$$

と表す. 対称性のためには

$$t_r = 1 - (t_1 + \cdots + t_{r-1})$$

とおいて, r 次元空間 $\mathbb{R}^r$ 内に浮かぶ $r - 1$ 次元の開単体

$$\Delta^{r-1} = \{ (t_1, \dots, t_r) \in \mathbb{R}_{>0}^{r} \mid t_1 + \cdots + t_r = 1 \}$$

上の確率分布だとみなした方がよい. これによって, Dirichlet分布は, カテゴリカル分布

$$\mathrm{Categorical}(t_1, \dots, t_r) \quad (t_i > 0,\ t_1 + \cdots + t_r = 1)$$

および多項分布

$$\mathrm{Multinomial}(n, t_1, \dots, t_r) \quad (t_i > 0,\ t_1 + \cdots + t_r = 1)$$

のパラメータ $(t_1, \dots, t_r)$ に関する確率分布だとみなされる. Dirichlet分布はカテゴリカル分布や多項分布の共役事前分布として重要である.

注意: $\Delta_{r-1} \subset \mathbb{R}^{r-1}$ と $\Delta^{r-1} \subset \mathbb{R}^r$ の微妙な違いに注意せよ. 一点, 線分, 三角形, 四面体, … を単体と呼ぶ. 例えば, Δ^1 は両端が $(1, 0)$ と $(0, 1)$ の線分から両端の点を除いたものになり, Δ^2 は頂点 $(1, 0, 0), (0, 1, 0), (0, 0, 1)$ を持つ正三角形(から境界を除いたもの)になる.

In [69]:

```julia
function _pdf_dirichlet(d::Dirichlet, p, q; valmax=Inf)
    0 ≤ q ≤ min(2p, -2p+2) || return NaN
    x, y = p - 0.5q, q
    clamp(pdf(d, SVector(x, y, 1 - x - y)), 0, valmax)
end

function plot_dirichlet!(d::Dirichlet; valmax=Inf, kwargs...)
    @assert length(d) == 3
    f(p, q) = _pdf_dirichlet(d, p, q; valmax)
    p = range(0, 1, length=401)
    q = range(0, 1, length=401)
    annotate!((1.00, 0.10, text("(1,0,0)", 8, -58.0, :right)))
    annotate!((0.55, 1.00, text("(0,1,0)", 8, -58.0, :left)))
    annotate!((0.00, 0.08, text("(0,0,1)", 8,  58.0, :left)))
    heatmap!(p, q, f.(p',q); kwargs...)
end

function plot_dirichlet(d::Dirichlet ; valmax=Inf, kwargs...)
    plot(; colorbar=false, tick=false, axis=false)
    plot_dirichlet!(d; valmax, kwargs...)
    plot!(; size=(250, 200))
    a, b, c = round.(d.alpha; digits=3)
    title!("Dirichlet($a, $b, $c)")
    plot!(; b=:gray)
end

plot_dirichlet(α₁, α₂, α₃ ; valmax=Inf, kwargs...) =
    plot_dirichlet(Dirichlet([α₁, α₂, α₃]) ; valmax, kwargs...)
```

Out[69]: plot_dirichlet (generic function with 2 methods)

以下はDirichlet分布のヒートマップである. 明るい部分ほど確率密度が大きい. 正三角形は $t_1, t_2, t_3 > 0$, $t_1 + t_2 + t_3 = 1$ を表し, 右下, 上, 左下の頂点はそれぞれ $(t_1, t_2, t_3) = (1, 0, 0), (0, 1, 0), (0, 0, 1)$ に対応しており, 正三角形の中心の座標は $(1/3, 1/3, 1/3)$ になる.

In [70]:

```julia
plot(plot_dirichlet(0.5, 0.5, 0.5; valmax=3),
    plot_dirichlet(0.5, 0.5, 1; valmax=3),
    plot_dirichlet(0.5, 1, 1; valmax=3); size=(800, 240), layout=(1,3))
```

Out[70]:

In [71]:

```julia
plot(plot_dirichlet(2, 3, 5), plot_dirichlet(4, 6, 10), plot_dirichlet(8, 12, 20);
    size=(800, 240), layout=(1,3))
```

Out[71]:

In [72]:
```
plot(plot_dirichlet(10, 15, 25), plot_dirichlet(20, 30, 60), plot_dirichlet(40, 60, 120);
    size=(800, 240), layout=(1,3))
```

Out[72]:

パラメータの大きなDirichlet分布は多変量正規分布で近似され, パラメータを大きくして行くと分布は一点に集中して行く.

5.3 多変数ベータ函数のガンマ函数表示の証明

$z = x_1 + \cdots + x_{n+1}$, $t_i = x_i/z$ とおく. 函数 $f(z, t_1, \ldots, t_n)$ について次の積分を考える:

$$J[f] = \int_0^\infty \cdots \int_0^\infty f(z, t_1, \ldots, t_n)\, e^{-(x_1+\cdots+x_{n+1})} x_1^{\alpha_1-1} \cdots x_{n+1}^{\alpha_{n+1}-1}\, dx_1 \cdots dx_{n+1}.$$

本当は z, t_i に $z = x_1 + \cdots + x_{n+1}$, $t_i = x_i/z$ を代入した式を書きたかったのだが, 式が煩雑に横に長くなり過ぎるのでそうしていない. この式の z, t にはそのように代入してから積分を行うという意味だと解釈する必要がある. そのとき

$$J[f] = \int_0^\infty \cdots \int_0^\infty \left(\int_0^\infty f(z, t_1, \ldots, t_n)\, e^{-(x_1+\cdots+x_{n+1})} x_1^{\alpha_1-1} \cdots x_{n+1}^{\alpha_{n+1}-1}\, dx_{n+1} \right) dx_1 \cdots dx_n$$

における内側の積分変数 x_{n+1} を $x_{n+1} = z - (x_1 + \cdots + x_n)$ を使って z に変換すると, z の動く範囲は $x_1 + \cdots + x_n < z < \infty$ になるので,

$$J[f] = \int_0^\infty \cdots \int_0^\infty \left(\int_{x_1+\cdots+x_n}^\infty f(z, t_1, \ldots, t_n)\, e^{-z} x_1^{\alpha_1-1} \cdots x_n^{\alpha_n-1} (z - (x_1 + \cdots + x_n))^{\alpha_{n+1}-1}\, dz \right) dx_1 \cdots dx_n.$$

この $n+1$ 重積分において $(x_1, \ldots, x_n, z)$ は $x_1, \ldots, x_n > 0$ かつ $x_1 + \cdots + x_n < z$ を満たす範囲を動く. ゆえに,

$$\Delta_n(z) = \{\, (x_1, \ldots, x_n) \in \mathbb{R}^n_{>0} \mid x_1 + \cdots + x_n < z \,\}$$

とおくと, 上の $n+1$ 重積分は次のように積分順序を交換できる:

$$J[f] = \int_0^\infty \left(\int \cdots \int_{\Delta_n(z)} f(z, t_1, \ldots, t_n)\, e^{-z} x_1^{\alpha_1-1} \cdots x_n^{\alpha_n-1} (z - (x_1 + \cdots + x_n))^{\alpha_{n+1}-1}\, dx_1 \cdots dx_n \right) dz.$$

積分変数 $x_1, \ldots, x_n$ を $x_i = zt_i$ によって $t_1, \ldots, t_n$ に変換すると, $(t_1, \ldots, t_n)$ は$t_1, \ldots, t_n > 0$ かつ $t_1 + \cdots + t_n < 1$ を満たす範囲を動くので,

$$\Delta_n = \Delta_n(1) = \{\, (t_1, \ldots, t_n) \in \mathbb{R}^n_{>0} \mid t_1 + \cdots + t_n < 1 \,\}$$

に注意すれば, 次が成立することがわかる:

$$J[f] = \int_0^\infty \left(\int \cdots \int_{\Delta_n} f(z, t_1, \ldots, t_n)\, e^{-z} (zt_1)^{\alpha_1-1} \cdots (zt_n)^{\alpha_n-1} (z(1 - (t_1 + \cdots + t_n)))^{\alpha_{n+1}-1}\, (z\, dt_1) \cdots (z\, dt_n) \right) dz.$$

これは次のように整理される:

$$J[f] = \int_0^\infty \left(\int \cdots \int_{\Delta_n} f(z, t_1, \ldots, t_n)\, e^{-z} z^{\alpha_1+\cdots+\alpha_{n+1}-1}\, t_1^{\alpha_1-1} \cdots t_n^{\alpha_n-1} (1 - (t_1 + \cdots + t_n))^{\alpha_{n+1}-1}\, dt_1 \cdots dt_n \right) dz.$$

これが $J[f]$ について示したかった公式である.

特に $f = 1$ のとき, $J[f]$ の定義より

$$J[1] = \Gamma(\alpha_1) \cdots \Gamma(\alpha_{n+1})$$

であることがわかり, 上で示した公式より,

$$J[1] = \Gamma(\alpha_1 + \cdots + \alpha_{n+1}) B(\alpha_1, \ldots, \alpha_{n+1})$$

となることがわかる. ゆえに

$$\Gamma(\alpha_1) \cdots \Gamma(\alpha_{n+1}) = \Gamma(\alpha_1 + \cdots + \alpha_{n+1}) B(\alpha_1, \ldots, \alpha_{n+1}).$$

すなわち

$$B(\alpha_1, \ldots, \alpha_{n+1}) = \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_{n+1})}{\Gamma(\alpha_1 + \cdots + \alpha_{n+1})}.$$

5.4 ガンマ分布とDirichlet分布の関係

前節の結果を使えば, ガンマ分布とベータ分布の関係の証明と全く同様に以下の結果を示せる.

$X_i \sim \mathrm{Gamma}(\alpha_i, 1)$ $(i = 1, \ldots, n)$ でかつそれらは独立だと仮定する. それらが独立であるとは, $(X_1, \ldots, X_n)$ について

$$E[f(X_1, \ldots, X_n)] = \frac{1}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_n)} \int \cdots \int f(x_1, \ldots, x_n) e^{-(x_1 + \cdots + x_n)} x_1^{\alpha_1 - 1} \cdots x_n^{\alpha_n - 1} \, dx_1 \cdots dx_n$$

が成立することと同値である. このとき,

$$Z = X_1 + \cdots + X_n, \quad T_i = \frac{X_i}{Z}$$

とおくと

$$Z \sim \mathrm{Gamma}(\alpha_1 + \cdots + \alpha_n), \quad (T_1, \ldots, T_{n-1}) \sim \mathrm{Dirichlet}(\alpha_1, \ldots, \alpha_n).$$

Z と $(T_1, \ldots, T_{n-1})$ の独立性(同時確率密度函数がそれぞれの確率密度函数の積になること)も示せる.

5.5 問題: Dirichlet (1839)の公式の証明

$a_i, p_i, \alpha_i > 0$ と仮定する. 領域 X を

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

と定める. ここで $\mathbb{R}_{>0}$ は正の実数全体の集合を表す. 次の積分を求めよ:

$$V = \int \cdots \int_X x_1^{\alpha_1 - 1} \cdots x_n^{\alpha_n - 1} \, dx_1 \cdots dx_n.$$

解答例: 変数 x_i 達と変数 t_i 達のあいだの

$$t_i = \left(\frac{x_i}{a_i}\right)^{p_i} \iff x_i = a_i t_i^{1/p_i}$$

という変換によって, 領域 X と領域

$$\Delta_n = \{ (t_1, \ldots, t_n) \in \mathbb{R}^n_{>0} \mid t_1 + \cdots + t_n < 1 \}$$

のあいだの一対一対応が得られる. ゆえに

$$\begin{aligned}
V &= \int \cdots \int_{\Delta_n} (a_i t_1^{1/p_1})^{\alpha_1 - 1} \cdots (a_n t_n^{1/p_n})^{\alpha_n - 1} \left(\frac{a_1}{p_1} t_1^{1/p_1 - 1} dt_1 \right) \cdots \left(\frac{a_n}{p_n} t_n^{1/p_n - 1} dt_n \right) \\
&= \frac{a_1^{\alpha_1} \cdots a_n^{\alpha_n}}{p_1 \cdots p_n} \int \cdots \int_{\Delta_n} t_1^{\alpha_1/p_1 - 1} \cdots t_n^{\alpha_n/p_n - 1} \, dt_1 \cdots dt_n \\
&= \frac{a_1^{\alpha_1} \cdots a_n^{\alpha_n}}{p_1 \cdots p_n} B(\alpha_1/p_1, \ldots, \alpha_n/p_n, 1) \\
&= \frac{a_1^{\alpha_1} \cdots a_n^{\alpha_n}}{p_1 \cdots p_n} \frac{\Gamma(\alpha_1/p_1) \cdots \Gamma(\alpha_n/p_n) \Gamma(1)}{\Gamma(\alpha_1/p_1 + \cdots + \alpha_n/p_n + 1)} \\
&= \frac{a_1^{\alpha_1} \cdots a_n^{\alpha_n}}{p_1 \cdots p_n} \frac{\Gamma(\alpha_1/p_1) \cdots \Gamma(\alpha_n/p_n)}{\Gamma(\alpha_1/p_1 + \cdots + \alpha_n/p_n + 1)}.
\end{aligned}$$

解答終

5.6 Dirichlet氏の1839年の論文との比較

以下は論文 Dirichlet (1839) (http://portail.mathdoc.fr/JMPA/afficher_notice.php?id=JMPA_1839_1_4_A11_0)の最後のページより.

Comme second exemple, j'indiquerai l'intégrale

$$V = \int x^{a-1} y^{b-1} z^{c-1} \ldots dx dy dz \ldots,$$

qui doit être étendue à toutes les valeurs positives de $x, y, z, \ldots$ telles qu'on ait

$$\left(\frac{x}{\alpha}\right)^p + \left(\frac{y}{\beta}\right)^q + \left(\frac{z}{\gamma}\right)^r + \ldots < 1,$$

les constantes $a, b, c, \ldots p, q, r, \ldots \alpha, \beta, \gamma, \ldots$ étant également positives. Par une analyse toute semblable, on parvient à cette expression très simple, qu'on peut aussi obtenir par d'autres moyens, et qui renferme un grand nombre de résultats relatifs aux volumes, aux centres de gravité, moments d'inertie, etc.:

$$V = \frac{\alpha^a \beta^b \gamma^c \ldots}{pqr} \frac{\Gamma\left(\frac{a}{p}\right) \Gamma\left(\frac{b}{q}\right) \Gamma\left(\frac{c}{r}\right) \ldots}{\Gamma\left(1 + \frac{a}{p} + \frac{b}{q} + \frac{c}{r} + \ldots\right)}.$$

これを見れば上の解答例によって, Dirichlet氏の1839年の論文の結果がぴったり再現されていることが分かる. ただし, 上で $x_1, x_2, \ldots, x_n$ と書いたところを, Dirichlet氏の時代には $x, y, z, \ldots$ のように書くことが普通であった. しかし, 以上を見れば, 19世紀のフランス語で書かれた数学の論文を現代の我々も結構読めそうなことがわかると思う.

Dirichlet氏が $a, b, c, \ldots$ と書いたところを上では $\alpha_1, \alpha_2, \ldots, \alpha_n$ と書き, Dirichlet氏が $\alpha, \beta, \gamma, \ldots$ と書いたところを上では $a_1, a_n, \ldots, a_n$ と書いた. それらの部分では, ギリシャ文字にする方が逆になっていることに注意せよ. この部分についてこのノートでDirichlet氏と逆のスタイルを採用せざるを得なかった理由は, 確率分布のパラメータをギリシャ文字で書く流儀を採用したからである.

Dirichlet氏が $p, q, r, \ldots$ と書いたところ上では $p_1, p_2, \ldots, p_n$ と書いた. 上に引用した論文の中にある2つめの $V =$ の式の右辺の分母の pqr は正しくは $pqr \cdots$ と $\cdots$ が必要である. このような細かい粗は昔の論文にも散見される. 数学の文献はそこに書いてあることを文字通りに正しいと思って読むのではなく, 内容を理解して適切に訂正しながら読む必要がある.

このように19世紀の数学者のDirichlet氏は, n 次元の球体の体積がガンマ函数で書けることの大幅な一般化を発見したのだが, その結果が, 後に統計学や機械学習での定番の道具の一つとして利用されることになることを予想できなかったに違いない.

In [ ]:
```
1
```

確率分布達の解釈

- 黒木玄

- 2022-04-11〜2022-05-21

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.

以下の図の内容を理解できればこのノートの内容をかなりよく理解できたことになる.

$\boxed{\text{ガンマ分布・正規分布ワールド}}$ $\boxed{\text{ベータ分布・t分布ワールド}}$ $\boxed{\text{ガンマ分布・正規分布ワールド}}$

$$\begin{tikzcd}
\mathrm{Gamma}(\alpha,\theta)\times\mathrm{Gamma}(\beta,\theta) \arrow[r, "(X,Y)\mapsto P=\frac{X}{X+Y}"] \arrow[d, equal] & \mathrm{Beta}(\alpha,\beta) \arrow[r, "\text{スケール極限 } \beta\to\infty"] \arrow[d, shift left, "U=\frac{P}{1-P}"] & \mathrm{Gamma}(\alpha,\theta) \arrow[d, equal] \\
\mathrm{Gamma}(\alpha,\theta)\times\mathrm{Gamma}(\beta,\theta) \arrow[r, "(X,Y)\mapsto U=\frac{X}{Y}"] \arrow[d, shift right, "\text{特殊化}"'] & \mathrm{BetaPrime}(\alpha,\beta) \arrow[r, "\beta\to\infty"] \arrow[u, shift left, "P=\frac{U}{1+U}"] \arrow[d, shift right, "\alpha=\nu_1 \;\|\; \beta=\nu_2"'] & \mathrm{Gamma}(\alpha,\theta) \arrow[d, shift right, "\alpha=\frac{\nu_1}{2} \;\|"'] \\
\mathrm{Gamma}(\frac{\nu_1}{2},2)\times\mathrm{Gamma}(\frac{\nu_2}{2},2) \arrow[r, "(X,Y)\mapsto U=\frac{X}{Y}"] \arrow[u, shift right, "\text{一般化}"'] \arrow[d, equal] & \mathrm{BetaPrime}(\frac{\nu_1}{2},\frac{\nu_2}{2}) \arrow[r, "\nu_2\to\infty"] \arrow[u, shift left, "\text{一般化}"] \arrow[d, leftrightarrow, "\text{スケール変換}"] & \mathrm{Gamma}(\frac{\nu_1}{2},\theta) \arrow[d, leftrightarrow, "\text{スケール変換}"] \\
\mathrm{Chisq}(\nu_1)\times\mathrm{Chisq}(\nu_2) \arrow[r, "(X,Y)\mapsto F=\frac{X/\nu_1}{Y/\nu_2}"] \arrow[d, shift right, "\text{特殊化}"'] & \mathrm{FDist}(\nu_1,\nu_2) \arrow[r, "\nu_2\to\infty"] \arrow[d, shift right, "\text{特殊化}"'] & \mathrm{Chisq}(\nu_1) \arrow[d, shift right] \\
\mathrm{Chisq}(1)\times\mathrm{Chisq}(\nu) \arrow[r, "(Z^2,Y)\mapsto T^2=\frac{Z^2}{Y/\nu}"] \arrow[u, shift right, "\text{一般化}"'] & \mathrm{FDist}(1,\nu) \arrow[r, "\nu\to\infty"] \arrow[u, shift left, "\text{一般化}"] & \mathrm{Chisq}(1) \arrow[u, shift left] \\
\mathrm{Normal}(0,1)\times\mathrm{Chisq}(\nu=n-1) \arrow[r, "(Z,Y)\mapsto T=\frac{Z}{\sqrt{Y/\nu}}"] \arrow[u, "Z^2"] \arrow[d, leftrightarrow, "\text{アフィン変換}"'] & \mathrm{TDist}(\nu) \arrow[r, "\nu\to\infty"] \arrow[u, "T^2"] & \mathrm{Normal}(0,1) \arrow[u, "Z^2"] \\
\mathrm{Normal}(\mu,\sigma/\sqrt{n})\times\mathrm{Gamma}(\frac{n-1}{2},\frac{2\sigma^2}{n-1}) & & \\
\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}) \arrow[u] & & \\
\mathrm{Normal}(\mu,\sigma)^n \quad \text{標本分布} & &
\end{tikzcd}$$

$$\begin{tikzcd}
\text{ベータ分布} \arrow[d, leftrightarrow] & \mathrm{Beta}(\alpha,\beta) \arrow[d, leftrightarrow] & \frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1}dp \arrow[d, leftrightarrow] & p=\frac{u}{1+u} \arrow[d, leftrightarrow] \\
\text{ベータプライム分布} & \mathrm{BetaPrime}(\alpha,\beta) & \frac{1}{B(\alpha,\beta)}\frac{u^{\alpha-1}}{(1+u)^{\alpha+\beta}}du & u=\frac{p}{1-p}
\end{tikzcd}$$

これはほぼF分布

$$\begin{tikzcd}
\boxed{\text{Bernoulli試行}} \arrow[r, "t=n/L", "L\to\infty"'] & \boxed{\text{連続時間極限}} \\
n\text{回中平均 } np\text{回} \arrow[r, "n=L", "p=\frac{\lambda}{L}"'] & \text{単位時間に平均 } \lambda\text{回} \\
K\sim\mathrm{Binomial}(n,p) \arrow[d, "p\sim\mathrm{Beta}(\alpha,\beta)"'] & K\sim\mathrm{Poisson}(\lambda) \arrow[d, "\lambda\sim\mathrm{Gamma}(\alpha,\theta)"] \\
K\sim\mathrm{BetaBinomial}(n,\alpha,\beta) \arrow[r, "n=L", "\beta=\frac{L}{\theta}"'] & K\sim\mathrm{NegativeBinomial}(\alpha,\frac{1}{1+\theta})
\end{tikzcd}$$

Pólya's urn: $p=\alpha/(\alpha+\beta)$で
当たりが出ると $(\alpha,\beta)\to(\alpha+1,\beta)$
はずれが出ると $(\alpha,\beta)\to(\alpha,\beta+1)$

$$\begin{tikzcd}
N-k\sim\mathrm{BetaNegativeBinomial}(k,\alpha,\beta) \arrow[r, "\beta=\frac{L}{\theta}"] & T\sim\frac{1}{\theta}\mathrm{BetaPrime}(k,\alpha) \\
p\sim\mathrm{Beta}(\alpha,\beta) \arrow[u] & \lambda\sim\mathrm{Gamma}(\alpha,\theta) \arrow[u] \\
N-k\sim\mathrm{NegativeBinomial}(k,p) \arrow[r, "N=LT", "p=\frac{\lambda}{L}"'] & T\sim\mathrm{Gamma}(k,\frac{1}{\lambda}) \\
\text{ちょうど}k\text{回イベントが起こるまでN回} & \text{ちょうど}k\text{回イベントが起こるまでの時間がT} \\
\boxed{\text{Bernoulli試行}} \arrow[r, "L\to\infty", "\text{単位時間にL回試行}"'] & \boxed{\text{連続時間極限}}
\end{tikzcd}$$

目次

- 1 正規分布
  - 1.1 分布のアフィン変換の定義
  - 1.2 正規分布のアフィン変換も正規分布
  - 1.3 問題: 正規分布の平均と分散
  - 1.4 問題: 正規分布に従う独立な確率変数達の和も正規分布に従う
  - 1.5 問題: 標準正規分布に従う独立な確率変数の和
  - 1.6 必ず解いて欲しい問題: 正規分布における確率が95%または99%になる区間
  - 1.7 問題: 正規分布のモーメント母函数とキュムラント母函数
  - 1.8 問題: キュムラント母函数と期待値と分散
  - 1.9 問題: 対数正規分布の確率密度函数
  - 1.10 問題: 対数正規分布の期待値と分散
- 2 t 分布の「分散が確率的に揺らいでいる正規分布」という解釈
  - 2.1 分散パラメータが確率分布に従う正規分布について
  - 2.2 問題: 分散がχ^2分布の ν 分の1に従う正規分布は自由度 ν の t 分布になる
  - 2.3 必ず解いて欲しい問題: t 分布における確率が95%または99%になる区間
- 3 Poisson分布の導入とその解釈
  - 3.1 Poisson分布の定義
  - 3.2 Poisson分布における確率の総和が1になることの確認
  - 3.3 問題: Poisson分布のキュムラント母函数と期待値と分散
  - 3.4 二項分布の連続時間極限
  - 3.5 Poisson分布は単位時間内に起こるイベントの回数の分布だとみなされる
  - 3.6 Poisson分布の中心極限定理と二項分布の中心極限定理の関係
  - 3.7 問題: Poisson分布の中心極限定理の直接証明
- 4 負の二項分布の「期待値＝分散が確率的に揺らいでいるPoisson分布」という解釈
  - 4.1 Poisson分布のパラメータがガンマ分布に従っていれば負の二項分布が得られる
  - 4.2 問題: Poisson分布とガンマ分布から負の二項分布が得られる
- 5 ガンマ分布の解釈
  - 5.1 問題: 標準正規分布に従う確率変数の2乗は自由度1のχ^2分布に従う
  - 5.2 問題: 標準正規分布に従う n 個の独立な確率変数達の2乗は自由度 n のχ^2分布に従う
  - 5.3 負の二項分布の連続時間極限
  - 5.4 負の二項分布の連続時間極限の様子をプロット
  - 5.5 ガンマ分布はイベントが α 回起こるまでにかかる時間の分布とみなされる
  - 5.6 Poisson分布の累積分布函数とガンマ分布の累積分布函数の関係
  - 5.7 問題: Poisson分布の累積分布函数とガンマ分布の累積分布函数の関係の直接証明
  - 5.8 必ず解いて欲しい問題: χ^2分布における確率が95%または99%になる範囲
  - 5.9 問題: 自由度1のχ^2分布と標準正規分布の関係の数値例
- 6 ベータ分布の一様乱数生成の繰り返しによる解釈
  - 6.1 一様分布のサイズ n の標本分布の順序統計量
  - 6.2 一様分布のサイズ n の標本分布の順序統計量が従う分布
  - 6.3 一様乱数生成の繰り返しからベータ分布が出て来ることの数値的確認
  - 6.4 二項分布の累積分布函数のベータ分布の累積分布函数表示
  - 6.5 二項分布とベータ分布の関係のClopper-PearsonのP値函数への応用
  - 6.6 おまけ: 二項分布とベータ分布の関係のBayes統計への応用
    - 6.6.1 平坦事前分布とimproper事前分布Beta(0,1)の事後分布の比較
  - 6.7 おまけ関連問題: ベータ分布は二項分布の共役事前分布である
    - 6.7.1 様々な事前分布の事後分布のプロット
  - 6.8 負の二項分布の累積分布函数のベータ分布の累積分布函数表示
  - 6.9 負の二項分布とベータ分布の関係の非整数パラメータケース
  - 6.10 二項分布の累積分布函数と負の二項分布の累積分布函数の関係(まとめ)
  - 6.11 Poisson分布の累積分布函数とガンマの累積分布函数の関係(再掲, まとめ)
- 7 ベータ二項分布
  - 7.1 ベータ二項分布の定義
  - 7.2 ベータ二項分布の「成功確率が確率的に揺らいでいる二項分布」という解釈
  - 7.3 ベータ二項分布での確率の総和が 1 になることの確認
  - 7.4 問題: ベータ二項分布の期待値と分散
  - 7.5 問題: ベータ二項分布 = 負の超幾何分布
  - 7.6 超幾何分布, 二項分布, ベータ二項分布の統一的な理解
  - 7.7 問題: 超幾何分布, 二項分布, ベータ二項分布の期待値と分散の統一的な公式
  - 7.8 問題: ベータ二項分布の極限として負の二項分布が得られる
  - 7.9 問題: 二項分布とベータ二項分布のモーメント母函数
- 8 ベータ負の二項分布とベータプライム分布
  - 8.1 ベータ負の二項分布の定義
  - 8.2 問題: ベータ負の二項分布の期待値と分散
  - 8.3 負の二項分布とベータ負の二項分布の期待値と分散の比較
  - 8.4 負の二項分布の連続時間極限の復習
  - 8.5 ベータ分布の極限の復習

8.6 ベータ負の二項分布の連続極限
8.7 ベータ負の二項分布の解釈 (Pólyaの壺)
8.8 Pólyaの壺の確率分布の定義

In [1]:
```
ENV["LINES"], ENV["COLUMNS"] = 100, 100
using BenchmarkTools
using Distributions
using Printf
using QuadGK
using Random
Random.seed!(4649373)
using Roots
using SpecialFunctions
using StaticArrays
using StatsBase
using StatsFuns
using StatsPlots
default(fmt = :png, titlefontsize = 10, size = (400, 250))
using SymPy
```

In [2]:
```
# Override the Base.show definition of SymPy.jl:
# https://github.com/JuliaPy/SymPy.jl/blob/29c5bfd1d10ac53014fa7fef468bc8deccadc2fc/src/types.

@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::SymbolicObject)
    print(io, as_markdown("\\displaystyle " * sympy.latex(x, mode="plain", fold_short_frac=fal
end
@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::AbstractArray{Sym})
    function toeqnarray(x::Vector{Sym})
        a = join(["\\displaystyle " * sympy.latex(x[i]) for i in 1:length(x)], "\\\\")
        """\\left[ \\begin{array}{r}$a\\end{array} \\right]"""
    end
    function toeqnarray(x::AbstractArray{Sym,2})
        sz = size(x)
        a = join([join("\\displaystyle " .* map(sympy.latex, x[i,:]), "&") for i in 1:sz[1]],
        "\\left[ \\begin{array}{" * repeat("r",sz[2]) * "}" * a * "\\end{array}\\right]"
    end
    print(io, as_markdown(toeqnarray(x)))
end
```

1 正規分布

$\mu, \sigma \in \mathbb{R}$, $\sigma > 0$ と仮定する. 確率密度函数

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

で定義される連続分布を平均 μ, 分散 σ^2 (標準偏差 σ)の **正規分布** (normal distribution)と呼び, $\mathrm{Normal}(\mu,\sigma)$ と書くのであった.

さらに, 平均 0, 分散 1^2 の正規分布を **標準正規分布** (standard normal distribution)と呼び, $\mathrm{Normal}()$ と書くのであった. 標準正規分布の密度函数は次の形になる:

$$p(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}.$$

1.1 分布のアフィン変換の定義

確率変数 X が従う分布を D と書くとき, $a, b \in \mathbb{R}$, $a \neq 0$ に対する $aX + b$ が従う分布を $aD + b$ と書いて, 分布 D の **アフィン変換** (affine transformation)と呼ぶ.

1.2 正規分布のアフィン変換も正規分布

$a, b \in \mathbb{R}$, $a \neq 0$ と仮定する. 一般に確率変数 X について,

$$E[aX+b] = aE[X] + b, \qquad \mathrm{var}(aX+b) = a^2\,\mathrm{var}(X)$$

が成立するのであった. X が正規分布に従う確率変数の場合には $aX + b$ も正規分布に従うことを示せる:

$$X \sim \mathrm{Normal}(\mu,\sigma) \implies aX + b \sim \mathrm{Normal}(a\mu + b, |a|\sigma).$$

実際, $X \sim \mathrm{Normal}(\mu, \sigma)$ のとき, $y = ax + b$ すなわち $x = (y - b)/a$ とおくと,

$$\begin{aligned} E[f(aX+b)] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} f(ax+b) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} f(y) \exp\left(-\frac{((y-b)/a-\mu)^2}{2\sigma^2}\right) \frac{dy}{|a|} \\ &= \frac{1}{\sqrt{2\pi(a^2\sigma)^2}} \int_{-\infty}^{\infty} f(y) \exp\left(-\frac{(y-(a\mu+b))^2}{2a^2\sigma^2}\right) dy \end{aligned}$$

なので, $aX + b$ は平均 $a\mu + b$, 分散 $a^2\sigma^2$ の正規分布に従う: $aX + b \sim \mathrm{Normal}(a\mu + b, |a|\sigma)$.

1.3 問題: 正規分布の平均と分散

分布 $\mathrm{Normal}(\mu, \sigma)$ の平均と分散がそれぞれ μ, σ^2 になることを示せ.

解答例1: すでに標準正規分布 $\mathrm{Normal}(0, 1)$ の平均と分散がそれぞれ 0 と 1^2 になることは示してある. $X \sim \mathrm{Normal}(\mu, \sigma)$ のとき, $Z = (X - \mu)/\sigma$ とおくと $Z \sim \mathrm{Normal}((\mu - \mu)/\sigma, \sigma/\sigma) = \mathrm{Normal}(0, 1)$ となるので, $X = \sigma Z + \mu$ の平均と分散はそれぞれ $\sigma \cdot 0 + \mu = \mu$, $\sigma^2 \cdot 1^2 = \sigma^2$ になる.

解答終

解答例2: 直接的に計算してみよう. $X \sim \mathrm{Normal}(\mu, \sigma)$ のとき, 積分変数を $x = \sigma z + \mu$ と変換すると,

$$E[X] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/(2\sigma^2)}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + \mu) e^{-z^2/2}\, dz = \mu.$$

最後の等号で $\int_{-\infty}^{\infty} z e^{-z^2/2}\, dz = 0$ と $\int_{-\infty}^{\infty} e^{-z^2/2}\, dy = \sqrt{2\pi}$ を使った. この結果を使うと,

$$\mathrm{var}(X) = E[(X-\mu)^2] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-(x-\mu)^2/(2\sigma^2)}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 z^2 e^{-z^2/2}\, dx = \sigma^2.$$

最後の等号で $\int_{-\infty}^{\infty} z^2 e^{-z^2/2}\, dz = \sqrt{2\pi}$ を使った. この結果は

$$\int_{-\infty}^{\infty} e^{-\alpha z^2}\, dz = \sqrt{\pi}\, \alpha^{-1/2}$$

の両辺を α で微分して -1 倍して $\alpha = 1/2$ とおいても得られるし, ガンマ函数に帰着する方法でも得られる.

解答終

1.4 問題: 正規分布に従う独立な確率変数達の和も正規分布に従う

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

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

となっているならば,

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

となることを示せ.

解答例: $X \sim \mathrm{Normal}(\mu_X, \sigma_X)$, $Y \sim \mathrm{Normal}(\mu_Y, \sigma_Y)$ でかつ X, Y は独立であると仮定する.

X, Y は独立なので $X + Y$ の平均と分散はそれぞれ $\mu_X + \mu_Y$, $\sigma_X^2 + \sigma_Y^2$ になる.

$X' = X - \mu_X$, $Y' = Y - \mu_Y$ とおくと, $X' \sim \mathrm{Normal}(0, \sigma_X)$, $Y \sim \mathrm{Normal}(0, \sigma_Y)$, $X + Y = (X' + Y') + (\mu_X + \mu_Y)$ となるので, $X' + Y'$ が正規分布に従うことを示せば十分である. ゆえに $\mu_X = 0$, $\mu_Y = 0$ と仮定してよいので, そのように仮定する.

$X + Y$ の確率密度函数を計算して, それが正規分布の密度函数になっていることを示せばよい.

$$E[f(X+Y)] = \frac{1}{\sqrt{(2\pi)^2\sigma_X^2\sigma_Y^2}} \iint_{\mathbb{R}^2} f(x+y) \exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_X^2} + \frac{y^2}{\sigma_Y^2}\right)\right) dx\, dy$$

であり, $x = x + y$ すなわち $y = z - x$ とおいて, x について平方完成すると,

$$\frac{x^2}{\sigma_X^2}+\frac{y^2}{\sigma_Y^2}=\frac{x^2}{\sigma_X^2}+\frac{(z-x)^2}{\sigma_Y^2}=\frac{\sigma_X^2+\sigma_Y^2}{\sigma_X^2\sigma_Y^2}\left(x-\frac{\sigma_X^2 z}{\sigma_X^2+\sigma_Y^2}\right)^2+\frac{z^2}{\sigma_X^2+\sigma_Y^2}.$$

となる(この計算を自分で実行してみること). ゆえに,

$$\begin{aligned}
&E[f(X+Y)]\\
&=\frac{1}{\sqrt{(2\pi)^2\sigma_X^2\sigma_Y^2}}\int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(z)\exp\left(-\frac{\sigma_X^2+\sigma_Y^2}{2\sigma_X^2\sigma_Y^2}\left(x-\frac{\sigma_X^2 z}{\sigma_X^2+\sigma_Y^2}\right)^2-\frac{z^2}{2(\sigma_X^2+\sigma_Y^2)}\right)dx\right)dz\\
&=\frac{1}{\sqrt{(2\pi)^2\sigma_X^2\sigma_Y^2}}\sqrt{\frac{2\pi\sigma_X^2\sigma_Y^2}{\sigma_X^2+\sigma_Y^2}}\int_{\mathbb{R}} f(z)\exp\left(-\frac{z^2}{2(\sigma_X^2+\sigma_Y^2)}\right)dz\\
&=\frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_Y^2)}}\int_{\mathbb{R}} f(z)\exp\left(-\frac{z^2}{2(\sigma_X^2+\sigma_Y^2)}\right)dz
\end{aligned}$$

以上によって $X+Y\sim\mathrm{Normal}\left(\mu_X+\mu_Y,\sqrt{\sigma_X^2+\sigma_Y^2}\right)$ を示せた.

解答終

別解: $z=x+y$, $y=z-x$ とおくと密度函数中の指数函数の中身が z についての二次式になることから, $Z=X+Y$ が正規分布に従うことは計算しなくても明らかだと考えることもできる. 上のように z 以外の変数は積分して消せる. そして, $Z=X+Y$ の期待値は X,Y の期待値の和になり, X,Y が独立という仮定からそれらは無相関になるので $Z=X+Y$ の分散は X,Y の分散の和になることもわかる. これだけで $X+Y\sim\mathrm{Normal}\left(\mu_X+\mu_Y,\sqrt{\sigma_X^2+\sigma_Y^2}\right)$ を証明できたともみなせる.

上の解答中の最も面倒な部分の計算が正しいことはコンピュータで以下のように確認できる.

$a=\sigma_X^2$, $b=\sigma_Y^2$ とおくと,

In [3]:
```
@vars a b x z
(a+b)/(a*b)*(x - a*z/(a+b))^2 + z^2/(a+b) - x^2/a - (z-x)^2/b ▷ simplify
```

Out[3]: 0

以下のように素朴に計算することもできる.

In [4]:
```
@vars a b positive=true
@vars x y z t
expr = x^2/a + y^2/b
```

Out[4]: $\frac{y^2}{b}+\frac{x^2}{a}$

In [5]:
```
expr = expr(y⇒z-x).expand()
```

Out[5]: $\frac{x^2}{b}-\frac{2xz}{b}+\frac{z^2}{b}+\frac{x^2}{a}$

In [6]:
```
A, B = 1/a+1/b, z/b
expr = sympy.poly(expr(x ⇒ t + B/A), t)
```

Out[6]: $\mathrm{Poly}\left(\frac{a+b}{ab}t^2+\frac{z^2}{a+b}, t, domain=\mathbb{Z}(z,a,b)\right)$

In [7]:

```
# 以下のセルでの説明で使う図の準備

μ, σ, c = 2, 3, 1.5
normal = Normal(μ, σ)

P1 = plot(normal, μ - 4σ, μ + 4σ; label="Normal(μ,σ)", xlabel="x = σz + μ")
vline!([μ]; label="x=μ", xtick = ([μ-c*σ, μ, μ+c*σ], ["μ-cσ", "μ", "μ+cσ"]), ytick=false)
plot!(normal, μ - 4σ, μ - c*σ; label="", c=1, frange=0, fc=:red, fa=0.5)
plot!(normal, μ + c*σ, μ + 4σ; label="", c=1, frange=0, fc=:red, fa=0.5)
annotate!(μ - 1.3c*σ, 0.8pdf(normal, μ - c*σ), ("P(X≤μ-cσ)", 10, :red, :right))
annotate!(μ - 1.5c*σ, 0.5pdf(normal, μ - c*σ), ("= α/2", 10, :red, :right))
annotate!(μ + 1.5c*σ, 0.5pdf(normal, μ + c*σ), ("= α/2", 10, :red, :left))
annotate!(μ + 1.3c*σ, 0.8pdf(normal, μ + c*σ), ("1-P(X≤μ+cσ)", 10, :red, :left))

P2 = plot(Normal(), -4, 4; label="Normal(0,1)", xlabel="z = (x - μ)/σ")
vline!([0]; label="z=0", xtick = ([-c, 0, c], ["-c", "0", "c"]), ytick=false)
plot!(Normal(), -4, -c; label="", c=1, frange=0, fc=:red, fa=0.5)
plot!(Normal(),  c,  4; label="", c=1, frange=0, fc=:red, fa=0.5)
annotate!(-1.3c, 0.8pdf(Normal(), -c), ("P(Z≤-c)", 10, :red, :right))
annotate!(-1.5c, 0.5pdf(Normal(), -c), ("= α/2", 10, :red, :right))
annotate!( 1.5c, 0.5pdf(Normal(),  c), ("= α/2", 10, :red, :left))
annotate!( 1.3c, 0.8pdf(Normal(),  c), ("1-P(Z≤c)", 10, :red, :left))

var"P(μ-cσ ≤ X ≤ μ+cσ) = 1-α ⇔ 1 - P(Z ≤ c) = α/2" =
    plot(P1, P2; size=(800, 250), bottommargin=4Plots.mm);
```

1.5 問題: 標準正規分布に従う独立な確率変数の和

確率変数 $X_1, \ldots, X_n$ の各々は標準正規分布に従い, それらは独立な確率変数達である(すなわちそれらの同時確率密度函数が標準正規分布の確率密度函数の積になっている)と仮定する. このとき

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

も標準正規分布に従うことを示せ.

解答例: $X_i/\sqrt{n}$ は平均 0, 分散 $1/n$ の正規分布に従し, $X_1/\sqrt{n}, \ldots, X_n/\sqrt{n}$ は独立になるので, 上の問題の結果より, それらの和は平均 $0 + \cdots + 0 = 0$, 分散 $1/n + \cdots + 1/n = 1$ の正規分布つまり標準正規分布に従う.

解答終

注意: この問題の結果より, 各々が標準正規分布に従う n 個の独立な確率変数達の和の $\sqrt{n}$ 分の 1 が従う分布も標準正規分布になる.

分布 D は平均 0 と分散 1 を持つ分布であるとき, 各々が分布 D に従う n 個の独立な確率変数達の和の $\sqrt{n}$ 分の 1 (以下 Z_n と書く)が従う分布はどうなるだろうか?

その場合にも Z_n の平均と分散はそれぞれ 0 と 1 になるが, Z_n が従う分布は一般に D とは異なる分布になる.

しかし, 応用上無理のないゆるい仮定のもとで, Z_n が従う分布は n を大きくすると標準正規分布に近付くことを示せる.

これが一般的な **中心極限定理** の主張である. もとの分布が標準正規分布なら Z_n が従う分布はぴったり標準正規分布になるが, それ以外の一般の場合にも n を大きくすれば Z_n が従う分布が近似的に標準正規分布になることはいえるのである.

1.6 必ず解いて欲しい問題: 正規分布における確率が95%または99%になる区間

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

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

が $1 - \alpha$ に等しくなるような c を誤差函数

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

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

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

解答例: 標準正規分布の場合の計算に帰着することを考えよう.

$Z = (X - \mu)/\sigma$ とおくと Z は標準正規分布に従うのであった: $Z \sim \mathrm{Normal}(0, 1)$. このとき,

$$\mu - c\sigma \le X \le \mu + c\sigma \iff -c \le Z = \frac{X - \mu}{\sigma} \le c$$

なので, 標準正規分布に従う確率変数 Z について $P(-c \le Z \le c)$ が 95% になる c と 99% になる c を求めればよい. 正規分布は左右対称なので, $P(-c \le Z \le c) = 1 - \alpha$ となることと, $1 - P(Z \le c) = \alpha/2$ すなわち

$$P(Z \le c) = 1 - \alpha/2$$

となることは同値である. 下の図を見よ.

In [8]:
```
var"P(μ-cσ ≤ X ≤ μ+cσ) = 1-α ⇔ 1 - P(Z ≤ c) = α/2"
```

Out[8]:

標準正規分布の累積分布函数

$$F(z) = P(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{t^2/2}\, dt$$

はコンピュータでの基本特殊函数ライブラリに含まれている誤差函数

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

を使えば

$$F(z) = P(Z \le z) = \frac{1 + \mathrm{erf}(z/\sqrt{2})}{2}.$$

と書けるのであった. この公式を証明するためには,

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{t^2/2}\, dt = \frac{1}{2}$$

を使ってから, $t = \sqrt{2}\,u$ とおけばよい. ゆえに誤差函数の逆函数 $\mathrm{erfinv}(y)$ (これもコンピュータでの基本特殊函数ライブラリに含まれている)を使えば, 標準正規分布の累積分布函数 $p = F(z)$ の逆函数(分位点函数)は

$$z = Q(p) = F^{-1}(p) = \sqrt{2}\,\mathrm{erfinv}(2p - 1)$$

と書ける. これを使えば $F(c) = P(Z \le c) = 1 - \alpha/2$ となる c を

$$c = Q(1 - \alpha/2) = \sqrt{2}\,\mathrm{erfinv}(1 - \alpha)$$

と求めることができる.

- $1 - \alpha = 95\%$ のとき, $c = \sqrt{2}\,\mathrm{erfinv}(0.95) \approx 1.96$
- $1 - \alpha = 99\%$ のとき, $c = \sqrt{2}\,\mathrm{erfinv}(0.99) \approx 2.58$

解答終

[Julia言語 (https://julialang.org/)](https://julialang.org/)では以下のように計算できる. (ただし, `using SpecialFunctions, Distributions` が必要.)

In [9]:
```
@show √2 * erfinv(0.95)
@show √2 * erfinv(0.99);
```

```
√2 * erfinv(0.95) = 1.9599639845400576
√2 * erfinv(0.99) = 2.5758293035489053
```

In [10]:
```
@show quantile(Normal(), 0.975)
@show quantile(Normal(), 0.995);
```

```
quantile(Normal(), 0.975) = 1.9599639845400576
quantile(Normal(), 0.995) = 2.5758293035489053
```

WolframAlpha (https://www.wolframalpha.com/)では以下のように計算できる:

- √2 erfinv(0.95) (https://www.wolframalpha.com/input?i=%E2%88%9A2+erfinv%280.95%29)
- √2 erfinv(0.99) (https://www.wolframalpha.com/input?i=%E2%88%9A2+erfinv%280.99%29)
- quantile(NormalDistribution(0,1), 0.975) (https://www.wolframalpha.com/input?i=quantile%28NormalDistribution%280%2C1%29%2C+0.975%29)
- quantile(NormalDistribution(0,1), 0.995) (https://www.wolframalpha.com/input?i=quantile%28NormalDistribution%280%2C1%29%2C+0.995%29)

1.7 問題: 正規分布のモーメント母函数とキュムラント母函数

$X \sim \mathrm{Normal}(\mu, \sigma)$ のとき次が成立することを示せ:

$$E[e^{tX}] = e^{\mu t + \sigma^2 t^2/2}, \quad \log E[e^{tX}] = \mu t + \sigma^2 \frac{t^2}{2}.$$

注意: 一般に確率変数 X に対して, $E[e^{tX}]$ を X の **モーメント母函数** (moment generating function, mgf)と呼び, $\log E[e^{tX}]$ を X の **キュムラント母函数** (cumulant generating function, cgf)と呼ぶ.

注意: 正規分布のキュムラント母函数は上のように非常に単純な形になる. キュムラント母函数が t について2次式になることと分布が正規分布であることは同値であり, キュムラント母函数中の t について3次以上の項は分布が正規分布とどのように違うかを表している. 特にそのうちの最初の2つである $t^3/3!$ と $t^4/4!$ の係数はそれぞれ **歪度** (わいど, skewness) と **尖度** (せんど, kurtosis)と呼ばれている.

注意: モーメント母函数とキュムラント母函数は物理での統計力学での分配函数と自由エネルギーの統計学での類似物になっており, 極めて便利な母函数になっている. 上の t は物理的には逆温度 β の -1 倍の $-\beta$ に対応している. (t が小さい場合の物理的類似は温度が低い場合になる. $t = -\beta$ が小さくなると温度の逆数 β は大きくなる. 温度の逆数 β が大きくなると温度は小さくなる. 注意: 絶対値が大きな負の温度は絶対値が小さな負の逆温度に対応し, 絶対温度無限大は $\beta = 0$ の場合に対応しているので, 絶対値が大きな負の温度は絶対温度無限大よりも少しだけ高温の場合に対応している. 絶対温度ではなく, その逆数の逆温度 $\beta = -t$ で考えた方が混乱が少なくて済む.)

解答例:

$$\begin{aligned}
tx - \frac{(x-\mu)^2}{2\sigma^2} &= -\frac{(x-\mu)^2 - 2\sigma^2 tx}{2\sigma^2} = -\frac{x^2 - 2(\mu + \sigma^2 t)x + \mu^2}{2\sigma^2} \\
&= -\frac{(x - (\mu + \sigma^2 t))^2 + \mu^2 - (\mu + \sigma^2 t)^2}{2\sigma^2} \\
&= -\frac{(x - (\mu + \sigma^2 t))^2 - 2\sigma^2 \mu t + \sigma^4 t^2}{2\sigma^2} \\
&= -\frac{(x - (\mu + \sigma^2 t))^2}{2\sigma^2} + \mu t + \sigma^2 \frac{t^2}{2}
\end{aligned}$$

より

$$\begin{aligned}
E[e^{tX}] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{tx} e^{-(x-\mu)^2)/(2\sigma^2)}\, dx \\
&= \frac{e^{\mu t + \sigma^2 t^2/2}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-(x-(\mu+\sigma^2 t))^2)/(2\sigma^2)}\, dx = e^{\mu t + \sigma^2 t^2/2}.
\end{aligned}$$

こらから $\log E[e^{tX}] = \mu t + \sigma^2 t^2/2$ はただちに得られる.

解答終

1.8 問題: キュムラント母函数と期待値と分散

期待値 μ と分散 σ を持つ確率変数 X について次が成立することを示せ:

$$\log E[e^{tX}] = \mu t + \sigma^2 \frac{t^2}{2} + O(t^3)$$

この結果は期待値と分散の計算に有用な場合がある. この結果は今後空気のごとく使われる.

解答例: $e^{tX} = 1 + Xt + X^2 t^2/2 + O(t^3)$ より,

$$E[e^{tX}] = 1 + E[X]t + E[X^2]\frac{t^2}{2} + O(t^3).$$

$\log(1+a) = a - a^2/2 + O(a^3)$ を使うと,

$$\begin{aligned}\log E[e^{tX}] &= E[X]t + E[X^2]\frac{t^2}{2} - \frac{E[X]t)^2}{2} + O(t^3) \\ &= E[X]t + (E[X^2] - E[X]^2)\frac{t^2}{2} + O(t^3) = \mu t + \sigma^2 \frac{t^2}{2} + O(t^3).\end{aligned}$$

解答終

1.9 問題: 対数正規分布の確率密度函数

$X \sim \mathrm{Normal}(\mu, \sigma)$ のときの $Y = e^X$ が従う分布を対数正規分布と呼び,

$$\mathrm{LogNormal}(\mu, \sigma)$$

と表す. 対数正規分布の確率密度函数を求めよ.

解答例: $x = \log y$ と積分変数を変換すると, $dx = dy/y$ なので,

$$E[f(Y)] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} f(e^x) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} f(y) \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right) \frac{dy}{y}.$$

ゆえに対数正規分布の確率密度函数は次の形になる:

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

解答終

In [11]:
```
plot(LogNormal(0, 1), 0, 8; label="", title="LogNormal(0, 1)")
```

Out[11]:

1.10 問題: 対数正規分布の期待値と分散

$Y \sim \mathrm{LogNormal}(\mu, \sigma)$ のとき次が成立することを示せ:

$$E[Y^m] = e^{m\mu + m^2\sigma^2/2}, \quad E[Y] = e^{\mu + \sigma^2/2}, \quad \mathrm{var}(Y) = e^{2\mu + \sigma^2}(e^{\sigma^2} - 1).$$

解答例: $X = \log Y \sim \mathrm{Normal}(\mu, \sigma)$ となるので, $E[Y^m]$ の計算で正規分布のモーメント母函数に関する結果を使え, 次が得られる:

$$E[Y^m] = E[e^{mX}] = e^{m\mu + m^2\sigma^2/2}.$$

ゆえに

$$
\begin{aligned}
&E[Y] = e^{\mu+\sigma^2/2}, \\
&\mathrm{var}(Y) = E[Y^2] - E[Y]^2 = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1).
\end{aligned}
$$

解答終

2 t 分布の「分散が確率的に揺らいでいる正規分布」という解釈

自由度 $\nu > 0$ のt 分布 $\mathrm{TDist}(\nu)$ の確率密度函数は

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

であり, これは $\nu \to \infty$ で標準正規分布の確率密度函数 $e^{-t^2/2}/\sqrt{2\pi}$ に収束するのであった.

2.1 分散パラメータが確率分布に従う正規分布について

Z は標準正規分布に従う確率変数であるとする.

$\sigma > 0$ について, $Z/\sqrt{\sigma^{-2}} = \sigma Z \sim \mathrm{Normal}(0, \sigma)$ となる.

もしも, 値が正の実数になる確率変数 Y の期待値が ν で標準偏差が ν との比較で小さいならば Y/ν の分布はその期待値 1 の近くに集中する. (以下, Z, Y は独立であると仮定する.) そのとき, $Z/\sqrt{Y/\nu}$ が従う分布は分散が 1 の周囲で確率的に揺らぐ「正規分布」に従うことになる.

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

$$
\begin{aligned}
E[f(Z/\sqrt{Y/\nu})] &= \int_0^\infty \left(\int_{-\infty}^\infty f\left(z/\sqrt{y/\nu}\right) \frac{e^{-z^2/2}}{\sqrt{2\pi}} p(y|\nu)\, dz \right) dy \\
&= \int_0^\infty \left(\int_{-\infty}^\infty f(t) \frac{e^{-yt^2/(2\nu)}}{\sqrt{2\pi}} p(y|\nu) \sqrt{\frac{y}{\nu}}\, dt \right) dy \\
&= \int_{-\infty}^\infty f(t) \left(\frac{1}{\sqrt{2\pi\nu}} \int_0^\infty e^{-t^2 y/(2\nu)} y^{1/2} p(y|\nu)\, dy \right) dt
\end{aligned}
$$

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

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

を持つことになる.

以上のように複数の確率変数を組み合わせて統計モデルを構築することが多い. 実は t 分布もそのような統計モデルの一種とみなせることを次の問題で確認してもらう. 以上を踏まえて次の問題を解け.

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

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

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

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

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

となることを示せ.

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

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

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

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

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

解答終

以上のような積分の計算をすることがどうしても嫌な人は, 以下のようにコンピュータで乱数を発生させてその分布を比較して納得するとよい. 上のような数学の計算が得意な人であっても, 具体的な計算をコンピュータに大量にさせて, その結果を視覚化して確認した方がよい. 「百聞は一見に如かず」は確率分布の世界でも正しい.

In [12]:

```
function plot_nct(ν; L = 10^6)
    Z = rand(Normal(), L) # 標準正規分布の乱数を大量に生成
    Y = rand(Chisq(ν), L) # χ² 分布で乱数達を大量生成
    T_nc = @. Z/√(Y/ν)

    # 比較のための同時プロット
    binmin, binmax = round.(quantile.(TDist(ν), (0.001, 0.999)))
    bin = binmin:0.1:binmax
    histogram(T_nc; norm=true, alpha=0.3, bin, xlim=(-6, 6), label="Z/√(Y/ν)")
    plot!(TDist(ν), binmin, binmax; label="TDist(ν)", lw=1.5)
    plot!(Normal(), binmin, binmax; label="Normal(0,1)", lw=1.5, ls=:dash)
    title!("ν = $ν")
end
```

Out[12]:

```
plot_nct (generic function with 1 method)
```

In [13]:

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

Out[13]:

In [14]:
```
plot(plot_nct(3), plot_nct(4); size=(800, 250))
```

Out[14]:

In [15]:
```
plot(plot_nct(10), plot_nct(20); size=(800, 250))
```

Out[15]:

注意: 一般に $Y \sim \mathrm{Gamma}(\alpha, \theta)$ のとき Y の平均と標準偏差はそれぞれ $\alpha\theta$ と $\sqrt{\alpha}\,\theta$ になるので, で $Y/(\alpha\theta)$ の平均と標準偏差はそれぞれ 1 と $1/\sqrt{\alpha}$ になる. これは α が大きなとき, $Y/(\alpha\theta)$ の分布は 1 の近くに集中することを意味している.

In [16]:
```
function plot_gamoαθ(α, θ = 1; L = 10^6) # Gamma over αθ
    Y = rand(Gamma(α, θ), L) # χ² 分布で乱数達を大量生成
    histogram(Y/(α*θ); norm=true, alpha=0.3, bin=0:0.025:2, label="Y/(αθ)")
    title!("α = $α")
end
```

Out[16]:
```
plot_gamoαθ (generic function with 2 methods)
```

In [17]:
```
plot(plot_gamoαθ(10), plot_gamoαθ(30), plot_gamoαθ(100); size=(800, 200), layout=(1, 3))
```

Out[17]:

2.3 必ず解いて欲しい問題: t 分布における確率が95%または99%になる区間

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

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

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

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

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

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

お願い: この問題の前半部分は少し難しいかもしれないので, どうしても無理ならば以下の解答例をざっと見るだけでもよい. しかし, 後半の「さらに」以降の具体的な数値を求める部分は必ず解けるようになっておいて欲しい. コンピュータを使ってよい. (より正確に言えばコンピュータを使った計算の仕方を1つ以上マスターしておくこと!)

注意: 正規分布モデルの統計学で t 分布に関する計算が不完全ベータ函数に帰着できるという事実から, 正規分布モデルの統計学においてベータ分布が必須であることがわかる.

解答例: 以下 $t \geq 0$ と仮定する. t 分布の累積分布函数を $F(t) = P(T \leq t)$ と書くと,

$$F(t) = \frac{1}{\sqrt{\nu}\, B(1/2, \nu/2)} \int_{-\infty}^{t} \left(1 + \frac{s^2}{\nu}\right)^{-(\nu+2)/2} ds = \frac{1}{2} + \frac{1}{\sqrt{\nu}\, B(1/2, \nu/2)} \int_{0}^{t} \left(1 + \frac{s^2}{\nu}\right)^{-(\nu+2)/2} ds.$$

ベータ函数の様々な表示の計算で学んだように, この積分変数を $s = \sqrt{\nu u}$ ($\iff$ $u = s^2/\nu$)とおいて, さらに $u = v/(1-v) = 1/(1-v) - 1$ ($\iff$ $v = u/(1+u)$)とおくとベータ分布の累積分布函数すなわち正則化された不完全ベータ函数に帰着できることがわかる:

$$\begin{aligned}
F(t) &= \frac{1}{2} + \frac{1}{\sqrt{\nu}\, B(1/2, \nu/2)} \int_0^{t^2/\nu} (1+u)^{-(\nu+2)/2} \frac{\sqrt{\nu}}{2} u^{-1/2}\, du \\
&= \frac{1}{2} + \frac{1}{2B(1/2, \nu/2)} \int_0^{t^2/\nu} \frac{u^{1/2-1}\, du}{(1+u)^{1/2+\nu/2}} \\
&= \frac{1}{2} + \frac{1}{2B(1/2, \nu/2)} \int_0^{(t^2/\nu)/(1+t^2/\nu)} \left(\frac{v}{1-v}\right)^{1/2-1} (1-v)^{1/2+\nu/2} \frac{dv}{(1-v)^2} \\
&= \frac{1}{2} + \frac{1}{2} \frac{\int_0^{(t^2/\nu)/(1+t^2/\nu)} v^{1/2-1}(1-v)^{\nu/2-1}\, dv}{B(1/2, \nu/2)} \\
&= \frac{1}{2}\left(1 + I\left(\frac{t^2/\nu}{1+t^2/\nu}\middle|\alpha, \beta\right)\right).
\end{aligned}$$

ゆえに, $p = u/(1+u)$ と $u = p/(1-p) = 1/(1-p) - 1$ が同値であることを使うと, $t \geq 0$ のとき,

$$\begin{aligned}
p = F(t) &\iff I^{-1}(2p-1|\alpha, \beta) = \frac{t^2/\nu}{1+t^2/\nu} \\
&\iff \frac{t^2}{\nu} = \frac{I^{-1}(2p-1|\alpha, \beta)}{1 - I^{-1}(2p-1|\alpha, \beta)} = \frac{1}{1 - I^{-1}(2p-1|\alpha, \beta)} - 1 \\
&\iff t = \sqrt{\nu\left(\frac{1}{1 - I^{-1}(2p-1|\alpha, \beta)} - 1\right)}.
\end{aligned}$$

ゆえに $t \geq 0$ における $p = F(t)$ の逆函数である分位点函数は

$$t = Q(p) = F^{-1}(p) = \sqrt{\nu\left(\frac{1}{1 - I^{-1}(2p-1|\alpha, \beta)} - 1\right)}$$

と書ける.

$P(-c \leq T \leq c) = 1 - \alpha$ と $F(c) = P(T \leq c) = 1 - \alpha/2$ は同値なので, そのような $c \geq 0$ は次のように表される:

$$c = Q(1 - \alpha/2) = \sqrt{\nu\left(\frac{1}{1 - I^{-1}(2p-1|\alpha, \beta)} - 1\right)}.$$

$\nu = 9$ のとき,

- $P(-c \leq T \leq c) = 95\%$ となる $c \geq 0$ は $c \approx 2.26$
- $P(-c \leq T \leq c) = 99\%$ となる $c \geq 0$ は $c \approx 3.25$

$\nu = 19$ のとき,

- $P(-c \leq T \leq c) = 95\%$ となる $c \geq 0$ は $c \approx 2.09$
- $P(-c \leq T \leq c) = 99\%$ となる $c \geq 0$ は $c \approx 2.86$

$\nu = 29$ のとき,

- $P(-c \le T \le c) = 95\%$ となる $c \ge 0$ は $c \approx 2.05$
- $P(-c \le T \le c) = 99\%$ となる $c \ge 0$ は $c \approx 2.76$

ちなみに $Z \sim \mathrm{Normal}(0, 1)$ のとき,

- $P(-c \le Z \le c) = 95\%$ となる $c \ge 0$ は $c \approx 1.96$
- $P(-c \le Z \le c) = 99\%$ となる $c \ge 0$ は $c \approx 2.58$

解答終

In [18]:

```
quantile_tdist(ν, p) = √(ν * (1/(1 - beta_inc_inv(1/2, ν/2, 2p - 1)[1]) - 1))
quantile_stdnormal(p) = √2 * erfinv(2p - 1)

for ν in (9, 19, 29)
    @show ν
    @show quantile_tdist(ν, 0.975)
    @show quantile_tdist(ν, 0.995)
    println()
end
ν = Inf
@show ν
@show quantile_stdnormal(0.975)
@show quantile_stdnormal(0.995);
```

```
ν = 9
quantile_tdist(ν, 0.975) = 2.262157162798204
quantile_tdist(ν, 0.995) = 3.2498355415921254

ν = 19
quantile_tdist(ν, 0.975) = 2.0930240544083096
quantile_tdist(ν, 0.995) = 2.860934606464978

ν = 29
quantile_tdist(ν, 0.975) = 2.0452296421327034
quantile_tdist(ν, 0.995) = 2.756385903670605

ν = Inf
quantile_stdnormal(0.975) = 1.9599639845400576
quantile_stdnormal(0.995) = 2.5758293035489053
```

In [19]:

```
for ν in (9, 19, 29)
    @show ν
    @show quantile(TDist(ν), 0.975)
    @show quantile(TDist(ν), 0.995)
    println()
end
ν = Inf
@show ν
@show quantile(TDist(ν), 0.975)
@show quantile(TDist(ν), 0.995);
```

```
ν = 9
quantile(TDist(ν), 0.975) = 2.262157162798205
quantile(TDist(ν), 0.995) = 3.249835541592126

ν = 19
quantile(TDist(ν), 0.975) = 2.093024054408309
quantile(TDist(ν), 0.995) = 2.860934606464979

ν = 29
quantile(TDist(ν), 0.975) = 2.0452296421327034
quantile(TDist(ν), 0.995) = 2.756385903670605

ν = Inf
quantile(TDist(ν), 0.975) = 1.9599639845400536
quantile(TDist(ν), 0.995) = 2.5758293035489
```

例えば $\nu = 9$ のとき, WolframAlphaでは以下のように計算できる:

- sqrt(n (1/(1 - InverseBetaRegularized(0.95, 1/2, n/2)) - 1)) where n = 9 (https://www.wolframalpha.com/input?i=sqrt%28n+%281%2F%281+-+InverseBetaRegularized%280.95%2C+1%2F2%2C+n%2F2%29%29+-+1%29%29+where+n+%3D+9),
- sqrt(n (1/(1 - InverseBetaRegularized(0.99, 1/2, n/2)) - 1)) where n = 9 (https://www.wolframalpha.com/input?i=sqrt%28n+%281%2F%281+-+InverseBetaRegularized%280.99%2C+1%2F2%2C+n%2F2%29%29+-+1%29%29+where+n+%3D+9).

- https://www.wolframalpha.com/input?i=quantile%28StudentTDistribution%289%29%2C+0.995%29
- quantile(StudentTDistribution(9), 0.975) (https://www.wolframalpha.com/input?i=quantile%28StudentTDistribution%289%29%2C+0.975%29)

3 Poisson分布の導入とその解釈

3.1 Poisson分布の定義

$\lambda > 0$ と仮定する. 確率質量函数

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

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

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

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

In [20]:

```
PP = []
for (λ, s) in ((1, 1), (3, 3), (10, 5), (30, 10) )
    x = max(0, round(λ-4√λ)):λ+4√λ+3/√λ
    P = bar(x, k → pdf(Poisson(λ), k);
        alpha=0.3, label="", title="Poisson($λ)", xtick=0:s:maximum(x))
    push!(PP, P)
end
plot(PP...; size=(800, 150), layout=(1, 4))
```

Out[20]:

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

3.2 Poisson分布における確率の総和が1になることの確認

指数函数のTaylor展開によって,

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

Poisson分布は指数函数のTaylor展開から得られる離散分布である.

Poisson分布よりも複雑な負の二項分布は函数 $p^{-\alpha} = (1-(1-p))^{-\alpha}$ の $1-p$ に関するTaylor展開

$$p^{-\alpha} = (1-(1-p))^{-\alpha} = \sum_{m=0}^{\infty}(-1)^m\binom{-\alpha}{m}(1-p)^m = \sum_{m=0}^{\infty}\binom{\alpha+m-1}{m}(1-p)^m$$

から得られるのであった. このパターンはよく現れる.

3.3 問題: Poisson分布のキュムラント母函数と期待値と分散

$K_\lambda \sim \mathrm{Poisson}(\lambda)$ のとき次が成立することを示せ:

$$\log E[e^{tK_\lambda}] = \lambda(e^t - 1), \qquad E[K_\lambda] = \mathrm{var}(K_\lambda) = \lambda.$$

さらに K_λ の標準化 Z_λ を $Z_\lambda = (K_\lambda - \lambda)/\sqrt{\lambda}$ と定めると次が成立することも示せ:

$$\log E[e^{tZ_\lambda}] = \log E\left[e^{t(K_\lambda-\lambda)/\sqrt{\lambda}}\right] = \frac{t^2}{2} + O(\lambda^{-1/2}).$$

ここで $O(\lambda^{-1/2})$ の部分は $O(\lambda)^{-1/2} = \lambda^{-1/2}\times$ ($\lambda \to \infty$ で有界な量)であることを表している. たとえば, $f(\lambda)$ が $\lambda \to \infty$ で収束するとき, $\lambda^{-1/2} f(\lambda)$ は $O(\lambda^{-1/2})$ と表される. 特に $O(\lambda^{-1/2})$ の部分は $\lambda \to \infty$ で 0 に収束する.

解答例:

$$E[e^{tK_\lambda}] = e^{-\lambda}\sum_{k=0}^{\infty} e^{tk}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^t)^k}{k!} = e^{\lambda(e^t-1)},$$

$$\log E[e^{tK_\lambda}] = \lambda(e^t-1) = \lambda\left(t + \frac{t^2}{2} + O(t^3)\right) = \lambda t + \lambda\frac{t^2}{2} + O(t^3).$$

$\log E[e^{tK_\lambda}]$ の展開における t, $t^2/2$ の係数がそれぞれ K_λ の期待値, 分散になるので

$$E[K_\lambda] = \mathrm{var}(K_\lambda) = \lambda.$$

さらに, $Z_\lambda = (K_\lambda - \lambda)/\sqrt{\lambda})$ について

$$\begin{aligned}\log E[e^{tZ_\lambda}] &= \log\left(e^{-\sqrt{\lambda}\,t}E\left[e^{(t/\sqrt{\lambda})K_\lambda}\right]\right) = \log E\left[e^{(t/\sqrt{\lambda})K_\lambda}\right] - \sqrt{\lambda}\,t) \\ &= \lambda\left(e^{t/\sqrt{\lambda}}-1\right) - \sqrt{\lambda}\,t) = \lambda\left(\frac{t}{\sqrt{\lambda}} + \frac{t^2}{2\lambda} + O(\lambda^{-3/2})\right) - \sqrt{\lambda}\,t \\ &= \left(\sqrt{\lambda}\,t + \frac{t^2}{2} + O(\lambda^{-1/2})\right) - \sqrt{\lambda}\,t = \frac{t^2}{2} + O(\lambda^{-1/2}).\end{aligned}$$

解答終

注意: 上の問題の結果より, $K_\lambda \sim \mathrm{Poisson}(\lambda)$, $Z_\lambda = (K_\lambda - \lambda)/\sqrt{\lambda}$, $Z_\infty \sim \mathrm{Normal}(0,1)$ のとき, λ を大きくすると $\log E[e^{tZ_\lambda}]$ が $\log E[e^{tZ_\infty}]$ に収束する. (一般に確率変数 X の分布はそのキュムラント母函数 $\log E[e^{tX}]$ から一意的に決まる. この点については別のノートで説明する.) これは実は λ が大きなとき, $Z_\lambda = (K_\lambda - \lambda)/\sqrt{\lambda}$ の従う分布が標準正規分布で近似されることを意味している. そのとき, もとの K_λ の分布は平均 λ, 分散 λ の正規分布で近似される. この結果を, **Poisson分布の中心極限定理** と呼ぶ.

In [21]:

```
PP = []
for λ in (10, 30)
    P = bar(max(0, round(λ-4√λ)):λ+4√λ+3/√λ, k -> pdf(Poisson(λ), k);
        alpha=0.3, label="", title="Poisson(λ = $λ)")
    plot!(Normal(λ, √λ); label="Normal(λ, √λ)", lw=2)
    push!(PP, P)
end
plot(PP...; size=(800, 250), layout=(1, 2))
```

Out[21]:

3.4 二項分布の連続時間極限

期待値 λ のPoisson分布は同じ期待値を持つ p が小さな二項分布によって近似されることを示そう. 二項分布 $\mathrm{Bonomial}(n,p)$ の期待値は np なので, $p = \lambda/n$ とおくと期待値は λ になる. そのように p を定めて, $n \to \infty$ とすると, 以下のように, 二項分布の確率質量函数はPoisson分布の確率質量函数に収束する:

$$
\begin{aligned}
\binom{n}{k} p^k (1-p)^{n-k} &= \frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \underbrace{\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}_{\to 1} \underbrace{\left(1-\frac{\lambda}{n}\right)^{n-k}}_{\to \exp(-\lambda)} \frac{\lambda^k}{k!} \to e^{-\lambda}\frac{\lambda^k}{k!}.
\end{aligned}
$$

1つめの等号で二項係数の定義と $p = \lambda/n$ を使い, 2つめの等号で分子の $n(n-1)\cdots(n-k+1)$ を $(\lambda/n)^k = \lambda^k/n^k$ の分母の n^k で割った. 最後に $n \to \infty$ の極限を取った.

3.5 Poisson分布は単位時間内に起こるイベントの回数の分布だとみなされる

前節の結果は以下のように解釈される.

(1) 二項分布は n 回のBernoulli試行で出る 1 の個数の分布であった. 1 が出ることを「イベントが起こった」と解釈することにする.

(2) n 回の試行を単位時間内に等間隔で行う状況を考える. これは時間の刻み幅が $1/n$ の離散時間を考えることに相当する. そのとき, 二項分布は単位時間のあいだに平均して np 回起こるイベントが単位時間内で何回起こるかに関する分布になっている.

(3) 単位時間内に起こる二項分布に従うイベントの回数の期待値 np を $np = \lambda$ と固定したままで, $n \to \infty$ とすることは, 時間の刻み幅を細かくする連続時間極限を取ることだと考えられる. この極限によって, 二項分布はPoisson分布に収束する.

(4) **Poisson分布は, 連続時間の場合に単位時間のあいだに平均して λ 回起こるイベントが単位時間内で起こる回数の分布になっている.**

In [22]:
```
function plot_binpoi(λ, n)
    @assert n > λ
    p = λ/n
    x = range(max(-0.5, λ - 5√λ), λ + 6√λ, 1000)
    plot(x, x -> pdf(Binomial(n, p), round(Int, x)); label="Binomial(n, p)")
    plot!(x, x -> pdf(Poisson(λ), round(Int, x)); label="Poisson(λ)", ls=:dash)
    title!("λ = $λ, n = $n, p = λ/n")
end
```

Out[22]: plot_binpoi (generic function with 1 method)

In [23]:
```
plot(plot_binpoi(1, 5), plot_binpoi(1, 15), plot_binpoi(1, 50); size=(800, 200), layout=(1, 3)
```

Out[23]:

In [24]:
```
plot(plot_binpoi(3, 15), plot_binpoi(3, 45), plot_binpoi(3, 150); size=(800, 200), layout=(1,
```

Out[24]:

In [25]:
```
plot(plot_binpoi(10, 50), plot_binpoi(10, 150), plot_binpoi(10, 500); size=(800, 200), layout=
```

Out[25]:

以上のグラフを見れば, $p = \lambda/n$ とおくとき, n を大きくすると, 二項分布が確かにPoisson分布で近似されていることがわかる.

3.6 Poisson分布の中心極限定理と二項分布の中心極限定理の関係

一般に期待値 μ, 標準偏差 σ を持つ確率変数 X の **標準化** Z は $Z = (X - \mu)/\sigma$ と定義される.

そのとき $\log E[e^{tZ}]$ を X の **標準化キュムラント母函数** と呼ぶ.

標準化キュムラント母函数が正規分布の標準化キュムラント母函数 $t^2/2$ にどれだけ近いかは, その分布が正規分布にどれだけ近いかを表している(この点については後で中心極限定理について一般的に説明するときに再度触れる).

Poisson分布のキュムラント母函数に関する問題の結果によれば,

$$K_\lambda \sim \mathrm{Poisson}(\lambda), \qquad Z_\lambda = \frac{K_\lambda - \lambda}{\sqrt{\lambda}} \qquad \Longrightarrow \qquad \log E[e^{tZ_\lambda}] = \frac{t^2}{2} + O(\lambda^{-1/2}).$$

これより, λ を大きくすると, Poisson分布は正規分布で近似されるようになることがわかる. この結果を**Poisson分布の中心極限定理**と呼ぶことにする.

さらに, 前節の結果より, $\lambda = np$ を一定としたままで, n を大きくすると, 二項分布 $\mathrm{Binomial}(n, p)$ はPoisson分布 $\mathrm{Poisson}(\lambda)$ で近似される.

そのとき注意するべきことは $\mathrm{Binomial}(n, p)$ の分散 $np(1-p)$ は $np = \lambda$ のとき $np(1-p) = \lambda(1-p)$ になり, p が小さくないと, $\mathrm{Poisson}(\lambda)$ の分散 λ と全然違う値になってしまうことである(たとえば $p = 1/2$ だと二項分布の分散が対応するPoisson分布の分散の半分になってしまう). 二項分布によるPoisson分布の近似は p が小さくないと精度が低くなる.

しかし, 以上の2つの結果を合わせると, Poisson分布の正規分布近似(=中心極限定理)を通して, p は小さいが np が大きい場合の二項分布 $\mathrm{Binomial}(n, p)$ が正規分布で近似されることがわかる. 実際には, p が小さくなくても, np と $n(1-p)$ が大きければ, 二項分布は正規分布でよく近似される. 二項分布が正規分布で近似されるという結果を **二項分布の中心極限定理** と呼ぶ.

p が小さい場合の二項分布の中心極限定理は応用上あまりにも不完全なので, 後で二項分布を直接正規分布で近似するという p が小さくない場合にも通用する方法で二項分布の中心極限定理を証明する.

注意: 「中心極限定理」は「極限として中心に収束する定理」という意味では **ない**. 「中心極限定理」は「確率論における中心的な極限定理」という意味である. 「中心極限定理」はある種の状況で分布が正規分布で近似されるようになるという結果に付けられた名前である.

In [26]:

```
PP = []
for (λ, n) in ((15, 20), (15, 300))
    p = λ/n
    ks = max(0, round(Int, λ-5.5√λ)):round(Int, λ+5.5√λ+3/√λ)
    P = groupedbar(ks,
        [pdf.(Poisson(λ), ks) pdf.(Binomial(n, λ/n), ks)];
        alpha=0.3,
        label=["Poisson(λ)" "Binomial(n,λ/n)"],
        title="λ = $λ,  n = $n,  p = λ/n = $p")
    plot!(Normal(λ, √λ); label="Normal(λ,√λ)", lw=1.5, c=:blue)
    plot!(Normal(λ, √(λ*(1-p))); label="Normal(λ,√(λ(1-p)))", lw=1.5, ls=:dash, c=:red)
    push!(PP, P)
end
plot(PP...; size=(800, 250), layout=(1, 2))
```

Out[26]:

左側のグラフでは $p = \lambda/n = 3/4$ なので, 二項分布の側の分散はPoisson分布の側の分散の4分の1になっており, 標準偏差については半分になっている.

右側のグラフでは $p = \lambda/n = 1/20$ が小さくなっているので, 二項分布のグラフとPoisson分布のグラフはほとんど一致し, 正規分布による近似もほとんど一致する.

3.7 問題: Poisson分布の中心極限定理の直接証明

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

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

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

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

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

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

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

注意: 以下の解答例は二項分布の中心極限定理の証明のひな型になる.

解答例: まず, 問題文よりも弱い次の条件を仮定する:

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

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

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

そして,

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

を $f(k)$ と書くと, $f(\lambda)=0$, $f'(k)=\log(k/\lambda)$, $f'(\lambda)=0$, $f''(k)=1/k$, $f''(\lambda)=1/\lambda$, $f'''(k)=-1/k^2$, $f'''(\lambda)=-1/\lambda^2$, etc. なので, $f(k)$ は $k=\lambda$ で次のようにTaylor展開される:

$$k\log\frac{k}{\lambda}+k-\lambda=f(k)=\frac{1}{2}\frac{(k-\lambda)^2}{\lambda}-\frac{1}{3!}\frac{(k-\lambda)^3}{\lambda^2}+\cdots$$

ゆえに, 問題文のように

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

と仮定すると,

$$\frac{k-\lambda}{\lambda}=\frac{x}{\sqrt{\lambda}}+o(\lambda^{-1/2})=O(\lambda^{-1/2}),\qquad \frac{(k-\lambda)^2}{\lambda^2}=\frac{x^2}{\lambda}+o(\lambda^{-1})=O(\lambda^{-1}),\qquad \text{etc.}$$

なので

$$\frac{(k-\lambda)^2}{\lambda}=x^2+o(1),\qquad \frac{(k-\lambda)^3}{\lambda^2}=O(\lambda^{-1/2})=o(1),\qquad \text{etc.}$$

となり, 次が得られる:

$$k\log\frac{k}{\lambda}+k-\lambda=\frac{x^2}{2}+o(1).$$

したがって,

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

これで示したい公式が得られた.

解答終

注意: 上の計算の途中で出て来た式

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

における $k\log(k/\lambda)$ の部分は **Kullback-Leibler情報量**

$$D(q||p)=\sum_i q_i\log\frac{q_i}{p_i}$$

の各項に化ける式になっている.

注意: 上の計算の途中で出て来た式

$$\frac{(k-\lambda)^2}{\lambda}=x^2+o(1)$$

の左辺は, **Pearsonのχ^2検定量**

$$\sum_i\frac{(\text{observation}_i-\text{expectation}_i)^2}{\text{expectation}_i}$$

の各項に化ける式になっている.

注意: このように, Poisson分布の中心極限定理を直接証明することには, 統計学的に重要な数学的対象の構成要素を得る行為にもなっている!

4 負の二項分布の「期待値＝分散が確率的に揺らいでいるPoisson分布」という解釈

4.1 Poisson分布のパラメータがガンマ分布に従っていれば負の二項分布が得られる

Poisson分布は単位時間内に起こるイベントの回数の分布とみなされるのであった.

Poisson分布でモデル化できると期待される現実の現象のデータを複数の対象について取得したとする. そのとき, そのすべてについて, 一定期間のあいだにイベントが起こる回数の期待値 λ が等しいと期待できるならば, Poisson分布モデルを単純に適用すればよい.

しかし, データを取得した対象ごとに, 一定期間のあいだにイベントが起こる回数の期待値 λ が異なると考えられる場合にはどうすればよいだろうか?

例えば, 野球選手達が年間何本ホームランを打つかについてデータを取得したとする. 選手ごとに打つホームランの本数の期待値は大きく異なる.

そのような場合には, ホームランの本数の期待値の選手間でのばらつき方も確率分布でモデル化すればよい.

たとえば一定期間内で打つホームランの本数の期待値の分布をガンマ分布でモデル化するとどうなるだろうか?

実はその場合には, ホームランの本数の分布を負の二項分布でモデル化したのと同じことになる.

次の問題ではそのことを実際に計算して確認してもらおう.

4.2 問題: Poisson分布とガンマ分布から負の二項分布が得られる

$\alpha, \theta > 0$ であるとし, 確率変数 Λ は次のようにガンマ分布に従っていると仮定する:

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

さらに確率変数 M はパラメータが確率変数 Λ であるようなPoisson分布に従っていると仮定する:

$$M \sim \mathrm{Poisson}(\Lambda).$$

これの意味は乱数の生成を考えるとわかりやすい. この確率変数 M に対応する乱数は以下の手続きで生成されると考える:

1. 分布 $\mathrm{Gamma}(\alpha, \theta)$ に従って乱数 Λ を生成する.
2. 分布 $\mathrm{Poisson}(\Lambda)$ に従って乱数 M を生成する.

このとき, M の分布が従う確率質量函数は次になることを示せ:

$$P(m|\alpha, \theta) = \binom{\alpha + m - 1}{m} \left(\frac{1}{1+\theta}\right)^{\alpha} \left(1 - \frac{1}{1+\theta}\right)^{m} \quad (m = 0, 1, 2, \ldots)$$

これは負の二項分布 $\mathrm{NegativeBinomial}(\alpha, \theta/(1+\theta))$ の確率質量函数に一致する. これによって, 負の二項分布の通常の解釈「α 回成功するまでに失敗する回数の分布」とは別の解釈が得られた. 負の二項分布は「期待値パラメータが固定されてなくてガンマ分布に従っているようなPoisson分布」であるとも考えられる. (この場合には「成功回数」にあたる α は整数でなくても任意の正の実数になれる.)

ただし, $\mathrm{Gamma}(\alpha, \theta)$ の確率密度函数を $p(\lambda|\alpha, \theta)$ と書き, $\mathrm{Poisson}(\lambda)$ の確率質量函数を $P(m|\lambda)$ と書くとき, 上のような状況における M が従う分布の確率質量函数 $P(m|\alpha, \theta)$ は

$$P(m|\alpha, \theta) = \int_0^{\infty} P(m|\lambda) p(\lambda|\alpha, \theta)\, d\lambda$$

と定義される.

解答例: $p(\lambda|\alpha, \theta)$, $P(k|\lambda)$ の定義は次の通り:

$$p(\lambda|\alpha, \theta) = \frac{e^{-\lambda/\theta} x^{\alpha-1}}{\theta^{\alpha}\Gamma(\alpha)}, \qquad P(k|\lambda) = \frac{e^{-\lambda}\lambda^{k}}{k!}.$$

ゆえに, ガンマ函数の基本的な使い方である

$$\int_0^{\infty} e^{-\lambda/\eta} \lambda^{\beta-1}\, d\lambda = \eta^{\beta}\Gamma(\beta)$$

を $\eta = \theta/(1+\theta)$, $\beta = \alpha + m$ の場合に使うと,

$$
\begin{aligned}
P(m|\alpha,\theta) &= \int_0^\infty P(m|\lambda)p(\lambda|\alpha,\theta)\,d\lambda = \int_0^\infty \frac{e^{-\lambda}\lambda^m}{m!}\frac{e^{-\lambda/\theta}x^{\alpha-1}}{\theta^\alpha\Gamma(\alpha)}\,d\lambda \\
&= \frac{1}{m!\theta^\alpha\Gamma(\alpha)}\int_0^\infty e^{-((1+\theta)/\theta)\lambda}\lambda^{\alpha+m-1}\,d\lambda\,d\lambda \\
&= \frac{1}{m!\theta^\alpha\Gamma(\alpha)}\left(\frac{\theta}{1+\theta}\right)^{\alpha+m}\Gamma(\alpha+m) \\
&= \frac{(\alpha+m-1)\cdots(\alpha+1)\alpha}{m!}\left(\frac{1}{1+\theta}\right)^{\alpha}\left(\frac{\theta}{1+\theta}\right)^{m} \\
&= \binom{\alpha+m-1}{m}\left(\frac{1}{1+\theta}\right)^{\alpha}\left(1-\frac{1}{1+\theta}\right)^{m}.
\end{aligned}
$$

解答終

この手の積分計算をサボりたい人は, 以下のようにコンピュータでガンマ分布の乱数を発生させて, それをパラメータとするPoisson分布の乱数を発生させることを大量に実行し, それと負の二項分布を直接使って生成した乱数の分布を比較してみるとよい.

In [27]:
```
function plot_gampoi(α, θ; L = 10^6)
    p = 1/(1 + θ)
    Λ = rand(Gamma(α, θ), L) # ガンマ分布で乱数達を大量生成
    M_gampoi = @. rand(Poisson(Λ)) # その各々を期待値とするPoisson分布の乱数を生成
    M_negbin = rand(NegativeBinomial(α, p), L) # 直接的に負の二項分布の乱数を大量生成

    # 比較のための同時プロット
    binmin, binmax = round.(quantile.(Ref(M_negbin), (0.001, 0.999)))
    stephist(M_gampoi; norm=true, bin=binmin-0.5:binmax+0.5, label="Gamma-Poisson")
    stephist!(M_negbin; norm=true, bin=binmin-0.5:binmax+0.5, ls=:dash, label="NegativeBinomia
end
```

Out[27]:
```
plot_gampoi (generic function with 1 method)
```

In [28]:
```
plot(plot_gampoi(1, 3), plot_gampoi(3, 1); size=(800, 250), layout=(1, 2))
```

Out[28]:

In [29]:
```
plot(plot_gampoi(6, 2), plot_gampoi(10, 1); size=(800, 250), layout=(1, 2))
```

Out[29]:

このようにMonte Carloシミュレーション(コンピュータで乱数を発生させることによるシミュレーション)でも, 期待値パラメータがガンマ分布に従っているようなPoisson分布と負の二項分布はぴったり一致している.

注意: $p = 1/(1+\theta)$ とおくと, $1-p = \theta/(1+\theta)$ になるので, 負の二項分布 $\mathrm{NegativeBinomial}(\alpha, \theta/(1+\theta))$ の期待値と分散は

それぞれ次のように書ける:

$$E[M] = \frac{\alpha(1-p)}{p} = \alpha\theta, \qquad \mathrm{var}(M) = \frac{\alpha(1-p)}{p^2} = \alpha\theta(1+\theta).$$

Poisson分布では期待値と分散が等しかったが, この場合には $1+\theta$ 倍の分だけ分散が期待値よりも大きくなっている. これより, $\mu = \alpha\theta$ を一定に保ったままで $\theta \searrow 0$ とすると負の二項分布 $\mathrm{NegativeBinomial}(\alpha, \theta/(1+\theta))$ がPoisson分布 $\mathrm{Poissom}(\mu)$ に収束すると予想されるが実際にそうなる: $\alpha = \mu/\theta$ とおいて, $\theta \searrow 0$ とすると,

$$\begin{aligned}
P(m|\alpha,\theta) &= \binom{\alpha+m-1}{m}\left(\frac{1}{1+\theta}\right)^{\alpha}\left(1-\frac{1}{1+\theta}\right)^{m} \\
&= \frac{(\mu/\theta+m-1)\cdots(\mu/\theta+1)(\mu/\theta)}{m!}\left(\frac{1}{1+\theta}\right)^{\mu/\theta}\left(\frac{\theta}{1+\theta}\right)^{m} \\
&= \underbrace{\frac{(\mu+(m-1)\theta)\cdots(\mu+\theta)\mu}{m!}}_{\to\mu^m/m!}\underbrace{\left(\frac{1}{1+\theta}\right)^{\mu/\theta}}_{\to\exp(-\mu)}\underbrace{\left(\frac{1}{1+\theta}\right)^{m}}_{\to 1} \to \frac{e^{-\mu}\mu^m}{m!}.
\end{aligned}$$

$\alpha = \mu/\theta$ とおいて, θ を小さくすると, $\mathrm{Gamma}(\mu/\theta, \theta)$ の分布は μ に集中して行くのでこうなるとも考えられる.

In [30]:

```
PP = []
μ = 3
for α in (10, 30, 100)
    θ = μ/α
    P = plot(Gamma(α,θ), 0, 6; label="", title="Gamma($α, $θ)")
    push!(PP, P)
end
plot(PP...; size=(800, 200), layout=(1, 3))
```

Out[30]:

In [31]:

```
PP = []
μ = 3
for α in (10, 30, 100)
    θ = μ/α
    p = 1/(1 + θ)
    x = max(0, round(Int, μ - 4√μ)):round(Int, μ + 4√μ)
    y1 = pdf.(NegativeBinomial(α, p), x)
    y2 = pdf.(Poisson(μ), x)
    P = groupedbar(x, [y1 y2]; alpha=0.5, label=["NegBin" "Poisson"])
    title!("μ = $μ, α = $α, θ = $θ")
    push!(PP, P)
end
plot(PP...; size=(800, 200), layout=(1, 3))
```

Out[31]:

5 ガンマ分布の解釈

形状パラメータ $\alpha > 0$, スケールパラメータ $\theta > 0$ を持つガンマ分布 $\mathrm{Gamma}(\alpha, \theta)$ の確率密度函数は

$$p(y|\alpha,\beta) = \frac{1}{\theta^\alpha \Gamma(\alpha)} e^{-y/\theta} y^{\alpha-1} \qquad (y > 0)$$

であった. 我々は別のノートでガンマ分布を「正規分布に関係した分布」として導入したのであった(χ²分布はガンマ分布の特別な場合であった). 以下の2つの問題を見よ. その後にガンマ分布の別の解釈を説明する.

5.1 問題: 標準正規分布に従う確率変数の2乗は自由度1のχ²分布に従う

標準正規分布に従う確率変数の2乗は自由度1のχ²分布に従うことを示せ.

解答例: $Z \sim \mathrm{Normal}(0,1)$ のとき, $Y = Z^2$ とおくと,

$$\begin{aligned} E[f(Y)] &= \int_{-\infty}^{\infty} f(z^2) \frac{e^{-z^2/2}}{\sqrt{2\pi}}\, dz = 2\int_0^{\infty} f(z^2) \frac{e^{-z^2}/2}{\sqrt{2\pi}}\, dz \\ &= \int_0^{\infty} f(y) \frac{e^{-y/2}}{\sqrt{2\pi}} y^{-1/2}\, dy = \int_0^{\infty} f(y) \frac{e^{-y/2} y^{1/2-1}}{2^{1/2}\Gamma(1/2)}\, dy \end{aligned}$$

2つめの等号で被積分函数が偶函数であることを使い, 3つめの等号で $z = \sqrt{y}$ とおき, 最後の等号で $\Gamma(1/2) = \sqrt{\pi}$ を使った. これは Y がχ²分布に従うことを意味している.

解答終

In [32]:

```
L = 10^6
Z = rand(Normal(), L)
Y = @. Z^2
histogram(Y; norm=true, alpha=0.3, label="Y = Z²")
plot!(Chisq(1); label="Chisq(1)", lw=1.5)
plot!(; xlim=(-0.2, 4.2), ylim=(-0.2, 4.2))
```

Out[32]:

5.2 問題: 標準正規分布に従う n 個の独立な確率変数達の2乗は自由度 n のχ²分布に従う

標準正規分布に従う n 個の独立な確率変数達の2乗は自由度 n のχ²分布に従うことを示せ.

解答例: $Z_1, \dots, Z_n$ は独立な確率変数達であるとし, $Z_i \sim \mathrm{Normal}(0,1)$ となっていると仮定する. このとき, 1つ上の問題より, 各 i ごとに $Z_i^2 \sim \mathrm{Chisq}(1) = \mathrm{Gamma}(1/2, 1/2)$ となり, $Z_1^2, \dots, Z_n^2$ は独立な確率変数達になる.

一般に, X, Y が独立な確率変数達で $X \sim \mathrm{Gamma}(\alpha,\theta)$, $Y \sim \mathrm{Gamma}(\beta,\theta)$ のとき, $X+Y \sim \mathrm{Gamma}(\alpha+\beta,\theta)$, $X/(X+Y) \sim \mathrm{Beta}(\alpha,\beta)$ で $X+Y$, $X/(X+Y)$ は独立になるのであった. この結果を上の状況で使うと,

$$Z_1^2 + \cdots + Z_n^2 \sim \mathrm{Gamma}(\underbrace{1/2 + \cdots + 1/2}_{n}, 1/2) = \mathrm{Gamma}(n/2, 1/2) = \mathrm{Chisq}(n)$$

となることがわかる.

解答終

In [33]:

```
L = 10^6
n = 5
Z = rand(Normal(), n, L)
Y = [sum(z -> z^2, Z) for Z in eachcol(Z)]
histogram(Y; norm=true, alpha=0.3, bin=100, label="Z₁² + ⋯ + Zₙ²")
plot!(Chisq(n); label="Chisq(1)", lw=1.5)
plot!(; xlim=(-1, 20))
title!("n = $n")
```

Out[33]:

5.3 負の二項分布の連続時間極限

ガンマ分布 $\mathrm{Gamma}(\alpha, \theta)$ が負の二項分布の連続極限になっていることを説明する. 以下ではガンマ分布が時間幅に関する確率分布だとみなしたいので, ガンマ分布に従う確率変数を T と書き, 対応する通常の変数を t と書くことにする.

負の二項分布は成功確率 $0 < p \leq 1$ のBernoulli試行を $\alpha > 0$ 回成功するまで続けたときに失敗する回数 M が従う分布であり, その確率質量函数は

$$P(m|\alpha, p) = \binom{\alpha + m - 1}{m} p^\alpha (1-p)^m \quad (m = 0, 1, 2, \ldots)$$

と書けるのであった(分布がうまく定義されるために α は整数でなくてもよい). このときの総試行回数 N は $N = \alpha + M$ と書ける. N が取り得る値の全体は $\alpha, \alpha+1, \alpha+2, \ldots$ になる.

一般に二項係数は

$$\binom{\alpha + m - 1}{m} = \frac{\Gamma(\alpha+m)}{\Gamma(m+1)\Gamma(\alpha)} = \frac{\Gamma(\alpha+m+1)}{(\alpha+m)\Gamma(m+1)\Gamma(\alpha)} = \frac{1}{(\alpha+m)B(\alpha, m+1)}$$

とベータ函数で書けるので, N が従う分布の確率質量函数は次のように書ける:

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

$L > 0$ であるとし, 単位時間に L 回にBenoulli試行を行い, 単位時間内での成功回数(イベントが起こる回数)の期待値 Lp は一定であり,

$$Lp = \frac{1}{\theta}$$

を満たしていると仮定する(p はこれによって L の函数とみなす). これは固定された θ について, 単位時間に平均して $1/\theta$ 回のイベントが起こるという仮定であり, この仮定はイベントが起こる間隔の期待値が θ であるという仮定と同等である.

さらにこのとき, 試行回数を表す通常の変数 n から離散時間変数 t が

$$t = \frac{n}{L}$$

と定義される. このとき単位時間分の試行が行われたとき $n = L$ となり, $t = 1$ となるのでつじつまが合っている.

以上の設定のもとで, 時間の刻み幅を 0 に近付ける $L \to \infty$ で負の二項分布の確率質量函数がどのように振る舞うかを調べよう. $n = Lt$ と $p = 1/(L\theta)$ を上の式に代入すると, $L \to \infty$ において,

$$
\begin{aligned}
P(Lt-\alpha|\alpha, 1/(L\theta)) &= \frac{1}{Lt\,B(\alpha, Lt-\alpha+1)}\frac{1}{(L\theta)^\alpha}\left(1-\frac{1}{L\theta}\right)^{Lt-\alpha} \\
&= \frac{1}{(Lt)^\alpha\,B(\alpha, Lt-\alpha+1)}\frac{(Lt)^{\alpha-1}}{(L\theta)^\alpha}\left(1-\frac{1}{L\theta}\right)^{Lt-\alpha} \\
&= \frac{1}{\theta^\alpha}\underbrace{\frac{1}{(Lt)^\alpha\,B(\alpha, Lt-\alpha+1)}}_{\to\Gamma(\alpha)}\underbrace{\left(1-\frac{1}{L\theta}\right)^{Lt-\alpha}}_{\to\exp(-t/\theta)}t^{\alpha-1}\frac{1}{L} \\
&= \frac{1}{\theta^\alpha\Gamma(\alpha)}e^{-t/\theta}t^{\alpha-1}\frac{1}{L}\,(1+O(1/L)).
\end{aligned}
$$

ここで $K^\alpha B(\alpha, K+b) \to \Gamma(\alpha)$ $(K \to \infty)$ となることを使った.

$1/L$ は離散時間 t の刻み幅なので, この式は時間の刻み幅を 0 に近付ける $L \to \infty$ の極限で, 負の二項分布がガンマ分布に収束することを意味している:

$$
\frac{\alpha + \mathrm{NegativeBinomial}(\alpha, p = 1/(L\theta))}{L} \to \mathrm{Gamma}(\alpha, \theta) \quad (L \to \infty)
$$

5.4 負の二項分布の連続時間極限の様子をプロット

前節の計算より, L が大きなとき, ガンマ分布 $\Gamma(\alpha, \theta)$ の確率密度函数 $p(t|\alpha, \theta)$ を近似する函数を, 負の二項分布の確率質量函数 $P(m|\alpha, p)$ を使って次のように作れることがわかる:

$$
p_{\mathrm{approx}}(t|\alpha, \theta, L) = P\Big(\mathrm{round}(Lt-\alpha)\,\Big|\,\alpha, 1/(L\theta)\Big)L.
$$

$T = N/L$ を $M = N - \alpha$ について解くと, $M = (LT - \alpha)$ であることを使った. 前節の計算から右辺には L をかけておかばければいけないことがわかる. ここで $\mathrm{round}(x)$ は x を四捨五入して整数に変換する函数であるとする.

この事実を使って, 負の二項分布とガンマ分布を比較するグラフを作成してみよう.

In [34]:
```
function plot_nbgam(α, θ, L)
    p = 1/(L*θ)
    @show mean((α + NegativeBinomial(α, p))/L), mean(Gamma(α, θ))
    @show std((α + NegativeBinomial(α, p))/L), std(Gamma(α, θ))
    t = range(max(-1/(2L), α*θ - 4√α*θ), α*θ + 6√α*θ, 1000)
    plot(t, t -> pdf(NegativeBinomial(α, p), round(Int, L*t - α))*L; label="(α+NB(α,p))/L")
    plot!(t, t -> pdf(Gamma(α, θ), t); label="Gamma(α, θ)", ls=:dash)
    title!("α=$α, θ=$θ, L=$L, p=1/(Lθ)")
end
```

Out[34]:
```
plot_nbgam (generic function with 1 method)
```

In [35]:
```
plot(plot_nbgam.(1, 1, (1.2, 3, 10))...; size=(800, 200), layout=(1, 3))
```

```
(mean((α + NegativeBinomial(α, p)) / L), mean(Gamma(α, θ))) = (1.0, 1.0)
(std((α + NegativeBinomial(α, p)) / L), std(Gamma(α, θ))) = (0.40824829046386296, 1.0)
(mean((α + NegativeBinomial(α, p)) / L), mean(Gamma(α, θ))) = (1.0, 1.0)
(std((α + NegativeBinomial(α, p)) / L), std(Gamma(α, θ))) = (0.816496580927726, 1.0)
(mean((α + NegativeBinomial(α, p)) / L), mean(Gamma(α, θ))) = (1.0, 1.0)
(std((α + NegativeBinomial(α, p)) / L), std(Gamma(α, θ))) = (0.9486832980505137, 1.0)
```

Out[35]:

In [36]:
```
plot(plot_nbgam.(2, 2, (1.2, 3, 10)./2)...; size=(800, 200), layout=(1, 3))
```

```
(mean((α + NegativeBinomial(α, p)) / L), mean(Gamma(α, θ))) = (4.0, 4.0)
(std((α + NegativeBinomial(α, p)) / L), std(Gamma(α, θ))) = (1.1547005383792515, 2.8284271247461
903)
(mean((α + NegativeBinomial(α, p)) / L), mean(Gamma(α, θ))) = (4.0, 4.0)
(std((α + NegativeBinomial(α, p)) / L), std(Gamma(α, θ))) = (2.3094010767585034, 2.8284271247461
903)
(mean((α + NegativeBinomial(α, p)) / L), mean(Gamma(α, θ))) = (4.0, 4.0)
(std((α + NegativeBinomial(α, p)) / L), std(Gamma(α, θ))) = (2.6832815729997477, 2.8284271247461
903)
```

Out[36]:

In [37]:
```
plot(plot_nbgam.(5, 4, (1.2, 3, 10)./4)...; size=(800, 200), layout=(1, 3))
```

```
(mean((α + NegativeBinomial(α, p)) / L), mean(Gamma(α, θ))) = (20.0, 20.0)
(std((α + NegativeBinomial(α, p)) / L), std(Gamma(α, θ))) = (3.651483716701107, 8.9442719099991
6)
(mean((α + NegativeBinomial(α, p)) / L), mean(Gamma(α, θ))) = (20.0, 20.0)
(std((α + NegativeBinomial(α, p)) / L), std(Gamma(α, θ))) = (7.302967433402216, 8.9442719099991
6)
(mean((α + NegativeBinomial(α, p)) / L), mean(Gamma(α, θ))) = (20.0, 20.0)
(std((α + NegativeBinomial(α, p)) / L), std(Gamma(α, θ))) = (8.48528137423857, 8.94427190999916)
```

Out[37]:

5.5 ガンマ分布はイベントが α 回起こるまでにかかる時間の分布とみなされる

前節の結果は以下のように解釈される.

(1) 負の二項分布は成功確率 p のBernoulli試行を 1 が α 回出るまでに出た 0 の個数 M の分布であった. 以下では 1 が出たことを「イベントが起こった」と解釈することにし, イベントが α 回起こるまでの試行回数 $N = \alpha + M$ の分布を考える.

(2) さらに, 試行を単位時間あたり L 回の等間隔で行う状況を考える. これは刻み幅が $1/L$ の離散時間を考えることに相当する. そのとき, $T = N/L$ は α 回イベントが起こるまでにかかる時間を意味する.

(3) 単位時間内に起こるイベントの回数の期待値 $Lp = 1/\theta$ を固定したままで, $L \to \infty$ とすることは, イベントが起こる間隔の期待値を θ に固定して連続時間極限を取ることだと考えられる. この極限によって, 負の二項分布における $T = N/L$ の分布はガンマ分布に収束する.

(4) **ガンマ分布は, 連続時間の場合に, イベント起こる間隔の期待値が θ であるときに, α 回のイベントが起こるまでにかかる時間の分布になっている.**

5.6 Poisson分布の累積分布函数とガンマ分布の累積分布函数の関係

Poisson分布は単位時間内に生じるイベントの回数の分布だと解釈でき, ガンマ分布はイベントがα回おこるまでにかかる時間の分布だと解釈できるのであった. これらの解釈から, Poisson分布における確率をガンマ分布における確率と解釈したり, その逆向きの解釈をしたりできることがわかる.

Poisson分布 $\mathrm{Poisson}(\lambda)$ は単位時間のあいだに平均して λ 回起こるイベントが単位時間内で起こる回数の分布だと解釈できるのであった. その状況のもとでは, より一般に時間 t のあいだにイベントが起こる回数 K の分布は $\mathrm{Poisson}(\lambda t)$ になる.

ガンマ分布 $\mathrm{Gamma}(\alpha, \theta)$ は, イベントが起こる時間間隔の期待値が θ であるときに, α 回イベントが起こるまでにかかる時間 T の分布だと解釈できるのであった.

イベントが起こる時間間隔の期待値が θ であることは, 単位時間のあいだに平均して $1/\theta$ 回のイベントが起こることと同値である. そこで, 以下では次のように仮定する:

$$\lambda = 1/\theta \quad \theta = 1/\lambda.$$

「時間 τ 以内にイベントが起こる回数が κ 回未満であること」と「イベントが κ 回起こるまでにかかる時間が τ より大きいこと」は論理的に同値なので, それらの確率も等しくなるはずである: $\tau > 0$, $\kappa = 1, 2, \ldots$ のとき,

$$K \sim \mathrm{Poisson}(\lambda\tau), \quad T \sim \mathrm{Gamma}(\alpha = \kappa, \theta = 1/\lambda), \quad \Longrightarrow \quad P(K < \kappa) = P(T > \tau).$$

実際にこれが成立している. それを式で書くと,

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

注意: この結果より, Poisson分布における累積分布函数の大量の和が必要な計算をガンマ分布における累積分布函数の計算に帰着できることがわかる. ガンマ分布の累積分布函数は基本特殊函数ライブラリに含まれている正則化された(不完全)ガンマ函数を使って効率的に計算できる.

5.7 問題: Poisson分布の累積分布函数とガンマ分布の累積分布函数の関係の直接証明

次の公式を証明せよ:

$$\sum_{k=0}^{\kappa-1} \frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!} = \frac{\int_\tau^\infty e^{-\lambda t} t^{\kappa-1}\, dt}{\Gamma(\kappa)/\lambda^\kappa} \quad (\lambda, \tau > 0,\ \kappa = 1, 2, \ldots)$$

解答例: 左辺を $f(\tau)$ とおく. そのとき,

$$f(t) = \sum_{k=0}^{\kappa-1} \frac{e^{-\lambda t}\lambda^k t^k}{k!}.$$

$t \searrow 0$ のとき $f(t) \to 0$ となり,

$$\begin{aligned} f'(t) &= \sum_{k\le\kappa-1} \frac{e^{-\lambda t}\lambda^k k t^{k-1}}{k!} - \sum_{k\le\kappa-1} \frac{e^{-\lambda t}\lambda\lambda^k t^k}{k!} = \sum_{k\le\kappa-1} \frac{e^{-\lambda t}\lambda^k t^{k-1}}{(k-1)!} - \sum_{k\le\kappa-1} \frac{e^{-\lambda t}\lambda^{k+1} t^k}{k!} \\ &= \sum_{k\le\kappa-1} \frac{e^{-\lambda t}\lambda^k t^{k-1}}{(k-1)!} - \sum_{k\le\kappa} \frac{e^{-\lambda t}\lambda^k t^{k-1}}{(k-1)!} = \frac{e^{-\lambda t}\lambda^\kappa t^{\kappa-1}}{(\kappa-1)!} = \frac{e^{-\lambda t} t^{\kappa-1}}{\Gamma(\kappa)/\lambda^\kappa}. \end{aligned}$$

ゆえに

$$f(\tau) = \frac{\int_\tau^\infty e^{-\lambda t} t^{\kappa-1}\, dt}{\Gamma(\kappa)/\lambda^\kappa}.$$

解答終

上のような煩雑な計算をやりたくない人も楽々遂行できる人も以下のように結果が本当に正しいかをコンピュータで計算して確認した方がよい.

In [38]:
```
function plot_cdfpoigam(λ, τ)
    poi = Poisson(λ*τ)
    gam(κ) = Gamma(κ, 1/λ)
    κ = range(1, round(mean(poi) + 5*std(poi))+0.5, 1000)
    plot(κ, κ -> cdf(poi, κ-1); label="cdf(Poisson(λτ),κ-1)")
    plot!(κ, κ -> ccdf(gam(floor(κ)), τ); ls=:dash, label="ccdf(Gamma(⌊κ⌋),1/λ),τ)")
    plot!(; legend=:bottomright, xlabel="κ")
    title!("λ=$λ, τ=$τ")
end
```

Out[38]: plot_cdfpoigam (generic function with 1 method)

In [39]:
```
plot(plot_cdfpoigam(3, 1), plot_cdfpoigam(3, 5); size=(800, 250), bottommargin=4Plots.mm)
```

Out[39]:

5.8 必ず解いて欲しい問題: χ²分布における確率が95%または99%になる範囲

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

$$P(Y \le c)$$

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

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

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

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

お願い: 後半の「さらに」以降の具体的な数値を求める部分は必ず解けるようになっておいて欲しい. コンピュータを使ってよい. (より正確に言えばコンピュータを使った計算の仕方を1つ以上マスターしておくこと!)

注意: 正規分布モデルの統計学でχ²分布に関する計算が不完全ガンマ函数に帰着できるという事実から, 正規分布モデルの統計学においてガンマ分布が必須であることがわかる.

解答例: $F(y) = P(Y \le y)$ とおく. $p = F(y)$ の逆函数 $y = Q(p) = F^{-1}(p)$ を正則化された不完全ガンマ函数を使って表そう:

$$\begin{aligned} F(y) = P(Y \le y) &= \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^y e^{-x/2} x^{\nu/2-1}\, dx \\ &= \frac{1}{\Gamma(\nu/2)} \int_0^{y/2} e^{-u} u^{\nu/2-1}\, du = P(y/2|\nu/2). \end{aligned}$$

2つめの等号で $x = 2u$ とおいた. これより, $p = F(y)$ の逆函数(分位点函数) $y = Q(p) = F^{-1}(p)$ が次のように表されることがわかる:

$$y = Q(p) = 2P^{-1}(p|\nu/2).$$

したがって, $P(Y \le c) = 1-\alpha$ となる c は

$$c = 2P^{-1}(1-\alpha|\nu/2)$$

と表される.

$\nu = 1$ のとき,

- $P(Y < c) = 95\%$ となる c は $c \approx 3.84$

- $P(Y \leq c) = 99\%$ となる c は $c \approx 6.63$

$\nu = 3$ のとき,

- $P(Y \leq c) = 95\%$ となる c は $c \approx 5.99$
- $P(Y \leq c) = 99\%$ となる c は $c \approx 9.21$

$\nu = 3$ のとき,

- $P(Y \leq c) = 95\%$ となる c は $c \approx 7.81$
- $P(Y \leq c) = 99\%$ となる c は $c \approx 11.34$

解答終

In [40]:

```
quantile_chisq(ν, p) = 2gamma_inc_inv(ν/2, p, 1-p)

for ν in (1, 2, 3)
    @show ν
    @show quantile_chisq(ν, 0.95)
    @show quantile_chisq(ν, 0.99)
    println()
end
```

```
ν = 1
quantile_chisq(ν, 0.95) = 3.8414588206941245
quantile_chisq(ν, 0.99) = 6.634896601021214

ν = 2
quantile_chisq(ν, 0.95) = 5.99146454710798
quantile_chisq(ν, 0.99) = 9.21034037197618

ν = 3
quantile_chisq(ν, 0.95) = 7.814727903251177
quantile_chisq(ν, 0.99) = 11.34486673014437
```

In [41]:

```
for ν in (1, 2, 3)
    @show ν
    @show quantile(Chisq(ν), 0.95)
    @show quantile(Chisq(ν), 0.99)
    println()
end
```

```
ν = 1
quantile(Chisq(ν), 0.95) = 3.8414588206941245
quantile(Chisq(ν), 0.99) = 6.634896601021214

ν = 2
quantile(Chisq(ν), 0.95) = 5.99146454710798
quantile(Chisq(ν), 0.99) = 9.21034037197618

ν = 3
quantile(Chisq(ν), 0.95) = 7.814727903251177
quantile(Chisq(ν), 0.99) = 11.34486673014437
```

例えば $\nu = 2$ の場合について, WolframAlphaでは以下のように計算できる:

- 2 InverseGammaRegularized(n/2, 1 - 0.95) where n = 2 (https://www.wolframalpha.com/input?i=2+InverseGammaRegularized%28n%2F2%2C+1+-+0.95%29+where+n+%3D+2)
- 2 InverseGammaRegularized(n/2, 1 - 0.99) where n = 2 (https://www.wolframalpha.com/input?i=2+InverseGammaRegularized%28n%2F2%2C+1+-+0.99%29+where+n+%3D+2)
- quantile(ChisqDistribution(2), 0.95) (https://www.wolframalpha.com/input?i=quantile%28ChisqDistribution%282%29%2C+0.95%29)
- quantile(ChisqDistribution(2), 0.99) (https://www.wolframalpha.com/input?i=quantile%28ChisqDistribution%282%29%2C+0.99%29)

注意: このノートでは分位点函数を $Q(p)$ と書いていて紛らわしいのだが, 不完全ガンマ函数として

$$Q(x|\alpha) = Q(\alpha, x) = 1 - P(\alpha, x) = \frac{\int_x^\infty e^{-u} u^{\alpha-1}}{\Gamma(\alpha)}$$

の側が使われることもよくある. $p = Q(x|\alpha)$ の逆函数を $Q^{-1}(p|\alpha)$ と書くとき, $P(Y \leq y) = 2Q^{-1}(1 - p|\alpha)$ となることに注意せよ. 上のWolframAlphaでは実際にそうなっている.

この. 上の Normal.alpha とは矢印でつながっている.

5.9 問題: 自由度1のχ²分布と標準正規分布の関係の数値例

$Z \sim \mathrm{Normal}(0,1)$, $Y \sim \mathrm{Chisq}(1)$ のとき,

$$P(-c \leq Z \leq c) = 0.95, \quad P(Y \leq d) = 0.95$$

を満たす c, d の値はそれぞれ

$$c \approx 1.96, \quad d \approx 3.84$$

であった. このとき

$$c^2 \approx 1.96^2 = 3.8416 \approx 3.84 \approx d$$

となる. これが必然である理由を説明せよ.

In [42]:
```
(196//100)^2 ▷ float
```

Out[42]: 3.8416

解答例1: 実際には正確に $c^2 = d$ となっている. なぜならば, $z > 0$ を $z = \sqrt{y}$ とおくと,

$$\begin{aligned} P(-c \leq Z \leq c) &= \int_{-c}^{c} \frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz = 2\int_0^c \frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz \\ &= 2\int_0^{c^2} \frac{e^{-y/2}}{\sqrt{2\pi}}\frac{1}{2}y^{-1/2}\,dy = \int_0^{c^2} \frac{e^{-y/2}y^{1/2-1}}{2^{1/2}\Gamma(1/2)}\,dy = \\ &= P(Y \leq c^2). \end{aligned}$$

つまり $P(-c \leq Z \leq c) = P(y \leq c^2)$ なので, $P(-c \leq Z \leq c) = 1-\alpha$ となる c の二乗と $P(Y \leq d) = 1-\alpha$ となる d は一致する.

解答終

解答例2: $Z^2 \sim \mathrm{Chisq}(1)$ となるのであった. ゆえに

$$P(-c \leq Z \leq c) = P(Z^2 \leq c^2) = P(Y \leq c^2).$$

これより, $P(-c \leq Z \leq c) = 1-\alpha$ となる c の二乗と $P(Y \leq d) = 1-\alpha$ となる d は一致することがわかる.

解答終

6 ベータ分布の一様乱数生成の繰り返しによる解釈

$\alpha, \beta > 0$ と仮定する. ベータ分布は 0 から 1 のあいだに分布する連続確率分布でその確率密度函数は

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

と書けるのであった. 別のノートで我々はこの確率密度函数を天下り的に与えて, この分布のある種の極限でガンマ分布が得られることや, ガンマ分布からベータ分布を作れることなどを示した. ガンマ分布は正規分布がらみの確率分布なので, この事実は, ベータ分布が正規分布モデルの統計学で大活躍する確率分布であることを示唆し, 実際にそうなっている.

以下ではこれとは別のベータ分布の解釈を紹介する. それによって, ベータ分布を二項分布や負の二項分布と繋げることに成功する. ベータ分布は二項分布モデルの統計学や負の二項分布モデルの統計学でも役に立つ.

6.1 一様分布のサイズ n の標本分布の順序統計量

独立な確率変数達 $T_1, T_2, \ldots, T_n$ のそれぞれは一様分布 $\mathrm{Uniform}(0,1)$ に従っていると仮定する. $T_1, \ldots, T_n$ の中で下から k 番目に小さなものを $T_{(k)}$ と書く.

$T_{(k)}$ も確率変数になることに注意せよ. $T_{(k)}$ に対応する乱数は次の手続きで作れる:

(1) rand() を n 回実行してその結果を $T_1, \ldots, T_n$ とする. (2) $T_1, \ldots, T_n$ の中で下から k 番目に小さな値を $T_{(k)}$ とする.

これによって 0 から 1 のあいだの数値 $T_{(k)}$ で以上の手続きによってランダムに与えられるものが得られる.

$T_{(k)}$ を **一様分布のサイズ n の標本分布の順序統計量** と呼ぶ.

6.2 一様分布のサイズ n の標本分布の順序統計量が従う分布

一様分布のサイズ n の標本分布の順序統計量 $T_{(k)}$ が従う分布はベータ分布 $\mathrm{Beta}(k, n-k+1)$ になる:

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

このようにパラメータが正の整数のすべてのベータ分布は一様分布の標本分布の順序統計量の分布として得られる.

証明1: $0 < t < t + dt < 1$ と $1, 2, \ldots, n$ を並び替えたもの $i_1, i_2, \ldots, i_n$ について,

$$P(T_{i_1}, \ldots, T_{i_{k-1}} < t < T_{i_k} < t + dt < T_{i_{k+1}}, \ldots, T_{i_n}) = t^{k-1}\, dt\, (1 - (t + dt))^{n-k}.$$

さらに, $1, 2, \ldots, n$ を $k-1$ 個の $i_1, \ldots, i_{k-1}$ と1個の i_k と $n-k$ 個の $i_{k+1}, \ldots, i_n$ の3グループへの分割の仕方の個数は

$$\frac{n!}{(k-1)!1!(n-k)!} = \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} = \frac{1}{B(k, n-k+1)}$$

と書ける. これと上の確率をかけて, $dt \to 0$ のときの dt より高次の微小量を無視したものが $T_{(k)}$ が従う分布の密度函数 $\times dt$ になる. (順序統計量 $T_{(k)}$ は $T_1, \ldots, T_n$ の中で k 番目に小さなものであった.) ゆえに, $T_{(k)}$ が従う分布の確率密度函数は

$$p(t|n, k) = \frac{t^{k-1}(1-t)^{n-k}}{B(k, n-k+1)} \quad (0 < t < 1)$$

になる. これはベータ分布 $\mathrm{Beta}(k, n-k+1)$ の確率密度函数である. これで $T_{(k)}$ が従う分布はベータ分布 $\mathrm{Beta}(k, n-k+1)$ になることがわかった.

証明終

証明2: 一様分布 $\mathrm{Uniform}(0, 1)$ の累積分布函数と確率密度函数をそれぞれ $F(t) = t$ $(0 < t < 1)$, $p(t) = F'(t) = 1$ $(0 < t < 1)$ と書き, 一様分布のサイズ n の標本分布の順序統計量 $T_{(k)}$ の累積分布函数を $G(t)$ と書く:

$$G(t) = P(T_{(k)} \leq t).$$

$T_{(k)} \leq t$ が成立することは, $T_1, \ldots, T_n$ の中の k 個以上が t 以下になることと同値なので,

$$G(t) = \sum_{j=k}^{n} \binom{n}{j} F(t)^j (1 - F(t))^{n-j}.$$

これの導函数が $T_{(k)}$ の確率密度函数になる:

$$\begin{aligned}
G'(t) &= \sum_{j \geq k} j \binom{n}{j} F(t)^{j-1} F'(t)(1 - F(t))^{n-j} - \sum_{j \geq k} (n-j) \binom{n}{j} F(t)^j F'(t)(1 - F(t))^{n-j-1} \\
&= \sum_{j \geq k} n \binom{n-1}{j-1} F(t)^{j-1} F'(t)(1 - F(t))^{n-j} - \sum_{j \geq k} n \binom{n-1}{j} F(t)^j F'(t)(1 - F(t))^{n-j-1} \\
&= \sum_{j \geq k} n \binom{n-1}{j-1} F(t)^{j-1} F'(t)(1 - F(t))^{n-j} - \sum_{j \geq k+1} n \binom{n-1}{j-1} F(t)^j F'(t)(1 - F(t))^{n-j} \\
&= n \binom{n-1}{j-1} F(t)^{k-1} F'(t)(1 - F(t))^{n-k} \\
&= \frac{n!}{(k-1)!(n-k)!} F(t)^{k-1} F'(t)(1 - F(t))^{n-k} \\
&= \frac{1}{B(k, n-k+1)} F(t)^{k-1} F'(t)(1 - F(t))^{n-k} \\
&= \frac{1}{B(k, n-k+1)} t^{k-1}(1-t)^{n-k}.
\end{aligned}$$

これはベータ分布 $\mathrm{Beta}(k, n-k+1)$ の確率密度函数である. これで $T_{(k)}$ が従う分布はベータ分布 $\mathrm{Beta}(k, n-k+1)$ になることがわかった.

証明終

注意: 上の $G'(t)$ の計算は最後の等号を除けば任意の函数 $F(t)$ について成立している. そのことを使えば一様分布に限らない場合にも順序統計量の確率密度函数を計算できる.

6.3 一様乱数生成の繰り返しからベータ分布が出て来ることの数値的確認

実際にコンピュータで n 個の一様乱数の組を大量に発生させて, 各々の中で k 番目に小さな値の分布がベータ分布になることを確認してみよう.

In [43]:
```
function sim_uniformbeta(n; L = 10^6)
    T = Matrix{Float64}(undef, n, L) # 順序統計量の乱数を保存する2次元行列
    tmp = [Vector{Float64}(undef, n) for _ in 1:Threads.nthreads()]
    Threads.@threads for j in 1:L
        t = rand!(tmp[Threads.threadid()]) # 一様乱数をn個生成
        T[:, j] .= sort!(t) # 小さい順序にソートして順序統計量の乱数を作る
    end
    T
end

function plot_sim_uniformbeta(T::Matrix; ran = axes(T, 1), L = 10^6, kwargs...)
    n = size(T, 1)
    PP = []
    for k in ran
        P = stephist(T[k,:]; norm=true, alpha=true, bin=[0; 0.01:0.02:0.99; 1], label="")
        plot!(; title="n=$n, k=$k", xtick=0:0.1:1)
        plot!(Beta(k, n-k+1), -0.01, 1.01; ls=:dash, label="")
        push!(PP, P)
    end
    plot(PP...; kwargs...)
end

function plot_sim_uniformbeta(n::Integer; ran = 1:n, L = 10^6, kwargs...)
    T = sim_uniformbeta(n; L)
    plot_sim_uniformbeta(T; ran, L, kwargs...)
end
```

Out[43]: plot_sim_uniformbeta (generic function with 2 methods)

In [44]:
```
plot_sim_uniformbeta(9; size=(800, 500), layout=(3, 3), tickfontsize=6)
```

Out[44]:

乱数の大量発生によるシミュレーションの結果のステップヒストグラムと理論的な分布 $\mathrm{Beta}(k, n-k+1)$ の確率密度函数をプロットした破線がぴったり一致している.

6.4 二項分布の累積分布函数のベータ分布の累積分布函数表示

$0 < p < 1$ のとき, 前節の証明2より,

$$P(T_{(k)} \leq p) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}.$$

この右辺は二項分布 $\mathrm{Binomial}(n, p)$ において k 以上になる確率に一致している.

さらに, $T_{(k)}$ が従う分布が $\mathrm{Beta}(k, n-k+1)$ だったので, その確率はベータ分布 $\mathrm{Beta}(k, n-k+1)$ において p 以下になる確率に一致することになる.

これで, 二項分布とベータ分布の累積分布函数の関係が得られた:

$$\begin{aligned} &K \sim \mathrm{Binomial}(n, p), \quad T \sim \mathrm{Beta}(k, n-k+1) \\ &\implies P(K \geq k) = P(T \leq p), \quad P(K \leq k-1) = 1 - P(T \leq p). \end{aligned}$$

k を $k+1$ に置き換えると,

$$\begin{aligned} &K \sim \mathrm{Binomial}(n, p), \quad T \sim \mathrm{Beta}(k+1, n-k) \\ &\implies P(K \geq k+1) = P(T \leq p), \quad P(K \leq k) = 1 - P(T \leq p). \end{aligned}$$

これらの公式の利点は右辺のベータ分布側の確率の計算を基本特殊函数ライブラリによって効率よく可能なことである. 左辺の二項分布における確率は n が大きいときには平均して $n/2$ 個の数値を足し上げる計算が必要になってしまう.

注意: 前節の証明1の結果($T_{(k)}$ の確率密度函数の形)を認めれば, 証明2における $G'(t)$ に関する少し面倒な計算を省略して,

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

が得られるので, 平易に得られる

$$P(T_{(k)} \leq p) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}$$

と合わせて,

$$\sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j} = \frac{\int_0^p t^{k-1}(1-t)^{n-k}\,dt}{B(k, n-k+1)}$$

が得られる. このことからも上の方で述べた二項分布の累積分布函数がベータ分布の累積分布函数で書けるという結果が得られる.

注意: 以上の結果はBernoulli試行を一様乱数の生成の繰り返しに持ち上げることによって得られたと考えられる. 0 と 1 の間の一様乱数の値が p 以下のときに値 1 を生成し, それ以外のときに 0 を生成するようにすれば, 一様乱数の生成の繰り返しからBernoulli試行が得られる.

そのときに, n 回分の一様乱数の生成結果の中で k 番目に小さな乱数のなす分布がベータ分布になっていることから, 二項分布とベータ分布の関係が自然に導かれた.

In [45]:

```
function plot_cdfbinbeta(n, p)
    bin = Binomial(n, p)
    μ, σ = mean(bin), std(bin)
    beta(k) = Beta(k, n-k+1)
    k = range(max(1, μ-4σ), min(n, μ+4σ), 1000)
    plot(k, k -> ccdf(bin, k-1); label="ccdf(Binomial(n,p),k-1)")
    plot!(k, k -> cdf(beta(floor(k)), p); ls=:dash, label="cdf(Beta(⌊k⌋,n-⌊k⌋+1)),p)")
    plot!(; legend=:topright, xlabel="k")
    title!("n=$n, p=$p")
end
```

Out[45]:

```
plot_cdfbinbeta (generic function with 1 method)
```

In [46]:
```
plot(plot_cdfbinbeta(20, 0.2), plot_cdfbinbeta(20, 0.4); size=(800, 250), bottommargin=4Plots.
```

Out[46]:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

この辺りの話は後で詳しく説明することになる.

6.6 おまけ: 二項分布とベータ分布の関係のBayes統計への応用

この一連のノートではBayes統計についての詳しい説明はしないつもりでいるのだが, 以下では **用語の詳しい定義を一切せずに** 二項分布とベータ分布の関係のBayes統計について説明してみたい. (以下の部分は理解できなくてもよい.)

二項分布モデルで事前分布として共役事前分布 $\mathrm{Beta}(a, b)$ を採用したとき, 「n 回中 k 回成功した」(例: 「n 任中 k 人の病気が治った」, 「n 人中 k 人が商品を購入してくれた」「n 個中 k 個が不良品だった」)というデータが得られたときの成功確率 p に関する事後分布は $\mathrm{Beta}(a + k, b + (n - k))$ になる.

特に形式的にimproper共役事前分布 $\mathrm{Beta}(0, 1)$ や $\mathrm{Beta}(1, 0)$ を採用したときの事後分布は前節や前々節で出て来た $\mathrm{Beta}(k, n-k+1)$ や $\mathrm{Beta}(k+1, n-k)$ になる.

そして, 二項分布の累積分布函数とベータ分布の累積分布函数の関係は以下のようになっていた:

$$K \sim \mathrm{Binomial}(n, p_0)$$

のとき,

$$
\begin{aligned}
T \sim \mathrm{Beta}(k, n-k+1) &\implies P(K \geq k) = P(T \leq p_0), \\
T \sim \mathrm{Beta}(k+1, n-k) &\implies P(K \leq k) = P(T \geq p_0).
\end{aligned}
$$

これは, 以下を意味している:

- 事前分布 $\mathrm{Beta}(0,1)$ に関する事後分布において「$p \leq p_0$」という仮説が成立する確率は二項分布モデルにおける「$p \leq p_0$」という仮説の片側検定のP値にぴったり等しい.
- 事前分布 $\mathrm{Beta}(1,0)$ に関する事後分布において「$p \geq p_0$」という仮説が成立する確率は二項分布モデルにおける「$p \geq p_0$」という仮説の片側検定のP値にぴったり等しい.

このように, 特定のケースではP値とBayes統計の事後分布において仮説が成立する確率がぴったり等しくなる. 事前分布を例えば平坦事前分布 $\mathrm{Beta}(1,1)$ に変えても, $(0,1)$ や $(1,0)$ と $(1,1)$ の違いは小さいので, n が数十程度で差は小さくなる. 「$p \leq p_0$」や「$p \geq p_0$」のような仮説の片側検定については, 通常のP値を使う統計学と事前事後分布を使うBayes統計学は数学的にはほぼ同じものだと思ってよい. (注意: ベータ分布 $\mathrm{Beta}(\alpha, \beta)$ は α, β がともに正の場合にのみ定義されているので, $\mathrm{Beta}(0,1)$ や $\mathrm{Beta}(1,0)$ は確率分布としては定義されていない. しかし, Bayes法はその場合にも有効になっている. $\mathrm{Beta}(0,1)$ や $\mathrm{Beta}(1,0)$ は improper事前分布と呼ばれている.)

この事実から, **「Bayes統計は通常の統計学とは異なる主義思想哲学に基く. そしてP値の使用には害がある. P値を使う統計学は捨ててBayes統計を使うべきである」のような主張をすることによって, Bayes統計を宣伝している人達が完全に間違っている** ことがわかる.

数学的には同じデータから常に同じ結果が出ることがわかっているのに, 主義思想哲学によって結果が変わると主張することは論理的に誤りである. 事前事後分布ではなくP値を使った場合にも, 数学的に同等だという理由で事前事後分布の場合と同じような解釈をP値についても許すならば論理的に整合性は取れているが, 「Bayes統計は通常の統計学とは異なる主義思想哲学に基く」のような主張をする人達はそういう論理的に整合性のある議論を決してしない.

以上の結果から, 「怪しげな事前分布を使うのでBayes統計は信用でいない」という主張も完全に間違っていることもわかる. 通常のP値を使った統計学に価値があるならば, 事前分布を適当に選べばそれと完全に同じことをできるBayes統計にも価値があるということは明らかである. むしろ, 事前分布という数学的選択肢が増えたおかげで, それなしにはできなかったこともできるようになると予想され, その予想は実際に正しい.

各種の技術には長所と短所の両方があることが普通である. 方法Aを方法Bに切り替える場合には, それで失われる方法A側の相対的長所とそれによって得られる方法Bの側の相対的長所のどちらを選ぶかになることが多い. 所謂「トレードオフ」の問題になる. P値を使う方法を使うか, Bayes統計の方法を使うかの問題も典型的なトレードオフの問題に過ぎない. そこに「主義思想哲学」を持ち込む必要はない.

6.6.1 平坦事前分布とimproper事前分布Beta(0,1)の事後分布の比較

上で述べたように次が成立しているのであった.

- 事前分布 $\mathrm{Beta}(0,1)$ に関する事後分布において「$p \leq p_0$」という仮説が成立する確率は二項分布モデルにおける「$p \leq p_0$」という仮説の片側検定のP値にぴったり等しい.

次の節の「おまけ関連問題」で扱うように, 事前分布 $\mathrm{Beta}(\alpha, \beta)$ のもとでデータ「n 回中 k 回成功」が得られたときの事後分布は $\mathrm{Beta}(\alpha+k, \beta+n-k)$ になる. 以下では事前分布が $\mathrm{Beta}(0,1)$ の場合の事後分布 $\mathrm{Beta}(k, n-k+1)$ と平坦事前分布 $\mathrm{Beta}(1,1)$ の事後分布 $\mathrm{Beta}(k+1, n-k+1)$ の確率密度函数を同時プロットして比較してみる. k と $n-k$ が大きな場合には差は小さい.

In [47]:
```
function plot_2betas(n, k; α=1, β=1, kwargs...)
    plot(Beta(k, n-k+1); label="Beta(k, n-k+1)")
    plot!(Beta(k+α, n-k+β); label="Beta(k+$α,n-k+$β)", ls=:dash)
    plot!(; xtick=0:0.1:1)
    title!("n=$n, k=$k"; kwargs...)
end
```

Out[47]:
```
plot_2betas (generic function with 1 method)
```

In [48]:
```
plot(plot_2betas(20, 5), plot_2betas(40, 10), plot_2betas(80, 20);
    size=(800, 200), layout=(1, 3), tickfontsize=7)
```

Out[48]:

このように k, $n-k$ が大きくなると, 事後分布への事前分布のパラメータの違いの影響は小さくなって行く.

6.7 おまけ関連問題: ベータ分布は二項分布の共役事前分布である

現実から「n 回中 k 回成功」の型のデータが得られる状況を考える(「n 人中 k 人が病気になった」「n 人中 k 人が商品を購入した」など).

そのデータと成功確率 p の二項分布 $\mathrm{Binomial}(n, p)$ の整合性がどうなっているかを見たいとする. その確率質量函数を次のように書く:

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

Bayes統計では, 成功確率パラメータ p に関する確率密度函数 $\varphi(p)$ で与えられる確率分布(**事前分布** (prior)と呼ぶ)をさらに与えて, 統計モデル内部では次のようになっていると考える:

1. 最初に成功確率パラメータの値 p が事前分布に従ってランダムに決まる.
2. その p の値を採用した二項分布 $\mathrm{Binomial}(n, p)$ に従って k がランダムに決まる.

この統計モデル内における k と p の同時分布は $P(k|n,p)\varphi(p)$ によって与えられ, モデル内での k, p のすべての組み合わせに関する確率の総和が 1 になることは次のようにして確認される:

$$\int_0^1 \underbrace{\sum_{k=0}^{n} P(k|n,p)}_{=1} \varphi(p)\,dp = \int_0^1 \varphi(p)\,dp = 1.$$

現実のデータを取得における成功確率 p に対応する値はこのモデル内設定のように確率分布しておらず, 定数かもしれない. 例えば, S市の住人をランダムに選んだときのBMI(=体重/身長の2乗)の値が25以上である確率は, S市におけるBMIが25以上である人達の割合に一致し, 確率分布しているわけではない. そういう状況であっても, 統計モデル内では成功確率パラメータ p がランダムに決まっていると考えてもよい. (その理由を一般的な主義思想哲学に求めることはよく見る典型的な誤りである. Bayes統計に限らず一般的に, 統計モデルとして現実にぴったり一致する可能性があるものを選ぶ必要はない.)

「n 人中 k 人成功」の型のデータが得られたとき, データとモデルを比較するために, その統計モデル内でデータと同じ数値が生成されたという条件で制限された **条件付き確率分布** (条件付き確率については後で説明したい)を考える. 統計モデルと現実から得たデータを比較するために, 数学的フィクションであるモデル内で現実から得たデータと同じ数値がランダムに生成されたという状況を考えるのである. Bayes法とは異なる最尤法の予測分布もそのような条件付き確率分布になっているので, そのような条件付き確率分布を考えるか否かはBayes統計であるか否かとは無関係である.

そのモデル内条件付き確率分布におけるパラメータ p の分布は **事後分布** (posterior)と呼ばれる. 条件付き確率の定義より, 「n 回中 k 回成功」というデータに関する事後分布 $\varphi(p|n,k)$ の確率密度函数は次の形になる:

$$\varphi(p|n,k) = \frac{P(k|n,p)\varphi(p)}{\int_0^1 P(k|n,p)\varphi(p)\,dp}.$$

事後分布は統計モデルとデータの整合性を見るために使われる. 例えば, 事後分布に従う確率変数 P について $P \le 0.3$ となる確率が小さいならば, その統計モデル+仮説「成功確率は 0.3 以下である」の組み合わせとデータとの整合性は低いと考える. 注意するべきことは「成功確率は 0.3 以下である」という仮説単体とデータの整合性を見ているのではなく, 統計モデル+仮説とデータの整合性しかわからないということである. 統計モデルを適切に与えないと仮説とデータの整合性について信頼できる結果は得られない.

問題: 以上の設定のもとで, 事前分布がベータ分布 $\mathrm{Beta}(a, b)$ で与えられたとき, 「n 回中 k 回成功」というデータに関する事後分布はベータ分布 $\mathrm{Beta}(a+k, b+n-k)$ になることを示せ.

解答例: 事前分布がベータ分布 $\mathrm{Beta}(a, b)$ で与えられているという設定は事前分布 $\varphi(p)$ が次のように書けているということを意味する:

$$\varphi(p) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}.$$

このとき,

$$C = \binom{n}{k} \Big/ B(a,b)$$

とおくと,

$$P(k|n,p)\varphi(p) = C\,p^k(1-p)^{n-k} \cdot p^{a-1}(1-p)^{b-1} = C\,p^{a+k-1}(1-p)^{b+n-k-1}$$
$$\int_0^1 P(k|n,p)\varphi(p)\,dp = C\int_0^1 p^{a+k-1}(1-p)^{b+n-k-1}\,dp = C\,B(a+k, b+n-k).$$

ゆえに, 事後分布 $\varphi(p|n,k)$ は次のようになる:

$$\varphi(p|n,k) = \frac{P(k|n,p)\varphi(p)}{\int_0^1 P(k|n,p)\varphi(p)\,dp} = \frac{C\,p^{a+k-1}(1-p)^{b+n-k-1}}{C\,B(a+k,b+n-k)} = \frac{p^{a+k-1}(1-p)^{b+n-k-1}}{B(a+k,b+n-k)}.$$

これは事後分布がベータ分布 $\mathrm{Beta}(a+k, b+n-k)$ になることを意味している.

解答終

In [49]:
```
function plot_betaposterior(a, b, n, k; kwargs...)
    beta = Beta(a+k, b+n-k)
    plot(beta, -0.01, 1.01; label="", xtick=0:0.1:1)
    title!("posterior Beta(a+k, b+n-k)\n(a,b)=($a,$b), n=$n, k=$k")
    plot!(; kwargs...)
end
```

Out[49]:
```
plot_betaposterior (generic function with 1 method)
```

6.7.1 様々な事前分布の事後分布のプロット

事後分布 $\mathrm{Beta}(a+k, b+n-k)$ を色々プロットしてみよう.

様々な事前分布について, データが「n 回中3割成功」の場合の事後分布を $n = 0, 10, 100, 1000$ の場合にプロットしてみる. ベイズ法ではデータサイズ n が 0 の場合(データが取得される前の状況)も扱える.

以下は事前分布が一様分布 $\mathrm{Beta}(1,1) = \mathrm{Uniform}(0,1)$ の場合.

In [50]:
```
a, b = 1, 1
plot(plot_betaposterior(a, b, 0, 0),
    plot_betaposterior(a, b, 10, 3);
    size=(800, 200), layout=(1, 2), topmargin=4Plots.mm)
```

Out[50]:

In [51]:
```
a, b = 1, 1
plot(plot_betaposterior(a, b, 100, 30),
    plot_betaposterior(a, b, 1000, 300);
    size=(800, 200), layout=(1, 2), topmargin=4Plots.mm)
```

Out[51]:

以下は事前分布がJeffreys事前分布 $\mathrm{Beta}(0.5, 0.5)$ の場合だが, 上の一様事前分布の場合とほとんど同じになっている.

In [52]:
```
a, b = 0.5, 0.5
plot(plot_betaposterior(a, b, 0, 0; ylim=(-0.3, 3.3)),
    plot_betaposterior(a, b, 10, 3);
    size=(800, 200), layout=(1, 2), topmargin=4Plots.mm)
```

Out[52]:

In [53]:
```
a, b = 0.5, 0.5
plot(plot_betaposterior(a, b, 100, 30),
    plot_betaposterior(a, b, 1000, 300);
    size=(800, 200), layout=(1, 2), topmargin=4Plots.mm)
```

Out[53]:

以下では違いを出すために事前分布を $\mathrm{Beta}(18, 2)$ にしてみた.

この事前分布を採用するとデータがない状態では「p は 0.9 に近い」という仮説と(事前分布を含めての)統計モデルの整合性が高いということになる. しかし, データが「n 回中3割成功」の形をしている場合には n を大きくすると, 「p は 0.3 に近い」という仮説と統計モデル+データの整合性が高くなる.

In [54]:
```
a, b = 18, 2
plot(plot_betaposterior(a, b, 0, 0),
    plot_betaposterior(a, b, 10, 3);
    size=(800, 200), layout=(1, 2), topmargin=4Plots.mm)
```

Out[54]:

In [55]:
```
a, b = 18, 2
plot(plot_betaposterior(a, b, 100, 30),
    plot_betaposterior(a, b, 1000, 300);
    size=(800, 200), layout=(1, 2), topmargin=4Plots.mm)
```

Out[55]:

6.8 負の二項分布の累積分布函数のベータ分布の累積分布函数表示

独立な確率変数達 $T_1, T_2, T_3, \dots$ のそれぞれは一様分布 $\mathrm{Uniform}(0,1)$ に従っていると仮定する. その最初の n 個 $T_1, \dots, T_n$ の中で下から k 番目に小さなものをこの節では $T_{(n,k)}$ と書くことにする. $T_{(n,k)}$ はベータ分布に従うのであった:

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

$0 < p < 1$ と仮定する. このとき, $T_i \le p$ ならば $X_i = 1$ とし, そうでないならば $X_i = 0$ とすることによって, 成功確率 p の Bernoulli分布に従う独立な確率変数達 $X_1, X_2, X_3, \dots$ が得られる. これはBernoulli試行そのものである.

Bernoulli試行を 1 がちょうど k 回出るまで続けたときの試行回数 N (これは確率変数になる)が n 以下になる確率は負の二項分布の確率質量函数を使って次のように書ける:

$$P(N \le n) = \sum_{j \le n} \binom{j-1}{j-k} p^k (1-p)^{j-k}.$$

この確率は $T_1, \dots, T_n$ の中に p 以下のものが k 個以上存在する確率に一致し, さらにその確率は $T_1, \dots, T_n$ の中で k 番目に小さな値 $T_{(n,k)}$ が p 以下になる確率に一致する.

以上より以下が成立することがわかる:

$$\begin{aligned} &N - k \sim \mathrm{NegativeBinomial}(k, p), \quad T \sim \mathrm{Beta}(k, n-k+1) \\ &\implies P(N \le n) = P(T \le p), \quad P(N > n) = 1 - P(T \le p). \end{aligned}$$

6.9 負の二項分布とベータ分布の関係の非整数パラメータケース

前節の議論では k は正の整数でなければいけない. 以下では正の整数 k を正の実数 α に置き換えても同様の結果が成立することを前々節の証明2の方法で示そう.

$M \sim \mathrm{NegativeBinomial}(\alpha, p)$ であるとし,

$$H(p) = P(M \le m) = \sum_{i \le m} \binom{\alpha + i - 1}{i} p^\alpha (1-p)^i$$

とおく. このとき,

$$\begin{aligned}
\frac{\partial}{\partial p}(p^\alpha (1-p)^i) &= \alpha t^{\alpha-1}(1-t)^i - i t^\alpha (1-p)^{i-1} \\
&= \alpha t^{\alpha-1}(1-t)^i - i(t^{\alpha-1}(-(1-t)+1)(1-t)^{i-1} \\
&= \alpha t^{\alpha-1}(1-t)^i + i t^{\alpha-1}(1-t)^i - i t^{\alpha-1}(1-t)^{i-1} \\
&= (\alpha+i) t^{\alpha-1}(1-t)^i - i t^{\alpha-1}(1-t)^{i-1}
\end{aligned}$$

より

$$\begin{aligned}
H'(t) &= \sum_{i \le m} (\alpha+i)\binom{\alpha+i-1}{i} t^{\alpha-1}(1-t)^i - \sum_{i \le m} i \binom{\alpha+i-1}{i} t^{\alpha-1}(1-t)^{i-1} \\
&= \sum_{i \le m} (i+1)\binom{\alpha+i}{i+1} t^{\alpha-1}(1-t)^i - \sum_{i \le m} i \binom{\alpha+i-1}{i} p^{\alpha-1}(1-p)^{i-1} \\
&= \sum_{i \le m+1} i \binom{\alpha+i-1}{i} t^{\alpha-1}(1-t)^{i-1} - \sum_{i \le m} i \binom{\alpha+i-1}{i} t^{\alpha-1}(1-t)^{i-1} \\
&= (m+1)\binom{\alpha+m}{m+1} t^{\alpha-1}(1-t)^m \\
&= -\frac{\Gamma(\alpha+m+1)}{\Gamma(\alpha)\Gamma(m+1)} t^{\alpha-1}(1-t)^m \\
&= -\frac{1}{B(\alpha, m+1)} t^{\alpha-1}(1-t)^m.
\end{aligned}$$

これはベータ分布 $\mathrm{Beta}(\alpha, m+1)$ の確率密度函数の形をしている. $H(0) = 0$ より,

$$P(M \le m) = H(p) = \int_0^p H'(t)\,dt = \frac{\int_0^p t^{\alpha-1}(1-t)^m\,dt}{B(\alpha, m+1)}.$$

以上より, 整数とは限らない $\alpha > 0$ についても以下が成立することがわかった:

$$\begin{aligned}
&M \sim \mathrm{NegativeBinomial}(\alpha, p), \quad T \sim \mathrm{Beta}(\alpha, m+1) \\
&\implies P(M \le m) = P(T \le p), \quad P(M > m) = 1 - P(T \le p).
\end{aligned}$$

注意: 負の二項分布の累積分布函数 $P(M \le m)$ の計算を定義通りに行うと, $m+1$ 個の数値の和を取る計算をすることになる. しかし, 以上の結果を使って, $P(M \le m)$ の計算をベータ分布の累積分布函数 $P(T \le p)$ の値として計算すれば, 基本特殊函数ライブラリの中に含まれている正則化された(不完全)ガンマ函数を使って効率的に計算できる.

In [56]:

```julia
function plot_cdfnegbinbeta(α, p)
    negbin = NegativeBinomial(α, p)
    μ, σ = mean(negbin), std(negbin)
    beta(m) = Beta(α, m+1)
    m = range(max(0, μ-4σ), μ+4σ, 1000)
    plot(m, m -> cdf(negbin, m); label="cdf(NegBin(α,p),m)")
    plot!(m, m -> cdf(beta(floor(m)), p); ls=:dash, label="cdf(Beta(α, ⌊m⌋+1)")
    plot!(; legend=:bottomright, xlabel="m")
    title!("α=$α, p=$p")
end
```

Out[56]:

```
plot_cdfnegbinbeta (generic function with 1 method)
```

In [57]:
```
plot(plot_cdfnegbinbeta(5, 0.7), plot_cdfnegbinbeta(5, 0.3); size=(800, 250), bottommargin=4Pl
```

Out[57]:

6.10 二項分布の累積分布函数と負の二項分布の累積分布函数の関係(まとめ)

以上によって, 以下が示された:

$$K \sim \mathrm{Binomial}(n, p), \quad T \sim \mathrm{Beta}(k, n-k+1), \quad N-k \sim \mathrm{NegativeBinomial}(k, p)$$

のとき,

$$P(K \geq k) = P(T \leq p) = P(N \leq n), \quad P(K < k) = P(T > p) = P(N > n).$$

これを具体的な式で書くと,

$$\sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i} = \frac{\int_0^p t^{k-1}(1-t)^{n-k}}{B(k, n-k+1)} = \sum_{j=k}^{n} \binom{j-1}{j-k} p^k (1-p)^{j-k},$$

$$\sum_{i=0}^{k-1} \binom{n}{i} p^i (1-p)^{n-i} = \frac{\int_p^1 t^{k-1}(1-t)^{n-k}}{B(k, n-k+1)} = \sum_{j=n+1}^{\infty} \binom{j-1}{j-k} p^k (1-p)^{j-k}.$$

In [58]:
```
function plot_cdf_bin_beta_nb(n, p; xtick=0:n)
    bin = Binomial(n, p)
    beta(k) = Beta(k, n-k+1)
    nb(k) = NegativeBinomial(k, p)
    μ, σ = mean(bin), std(bin)
    k = range(max(1, μ-4σ), μ+5σ, 1000)
    plot(k, k -> ccdf(bin, k-1); label="cdf(Bin(n,p),k-1)")
    plot!(k, k -> cdf(beta(floor(k)), p); ls=:dash, label="cdf(Beta(k,n-⌊k⌋+1),p)")
    plot!(k, k -> cdf(nb(floor(k)), n - floor(k)); ls=:dashdot, label="cdf(NegBin(⌊k⌋,p),n-⌊k⌋
    plot!(; xlabel="k", xtick)
    title!("n=$n, p=$p")
end
```

Out[58]: `plot_cdf_bin_beta_nb (generic function with 1 method)`

In [59]:
```
plot(plot_cdf_bin_beta_nb(10, 0.3), plot_cdf_bin_beta_nb(30, 0.3);
    size=(800, 250), bottommargin=4Plots.mm)
```

Out[59]:

In [60]:
```
plot(plot_cdf_bin_beta_nb(100, 0.3; xtick=0:5:100), plot_cdf_bin_beta_nb(300, 0.3; xtick=0:10:
    size=(800, 250), bottommargin=4Plots.mm)
```

Out[60]:

6.11 Poisson分布の累積分布函数とガンマの累積分布函数の関係(再掲, まとめ)

二項分布と負の二項分布の連続時間極限でそれぞれPoisson分布とガンマ分布が得られるのであった. ゆえに, 前節でまとめた結果は以下の連続時間版の結果の離散時間版だとみなされる:

$$K \sim \mathrm{Poisson}(\lambda\tau), \quad T \sim \mathrm{Gamma}(\alpha = \kappa, \theta = 1/\lambda)$$

のとき,

$$P(K \geq \kappa) = P(T \leq \tau), \quad P(K < \kappa) = P(T > \tau).$$

これを具体的な式で書くと,

$$\sum_{j=\kappa}^{\infty} \frac{e^{-\lambda\tau}(\lambda\tau)^j}{j!} = \frac{\int_0^\tau e^{-\lambda t} t^{\kappa-1}\, dt}{\lambda^{-\kappa}\Gamma(\kappa)},$$
$$\sum_{j=0}^{\kappa-1} \frac{e^{-\lambda\tau}(\lambda\tau)^j}{j!} = \frac{\int_\tau^\infty e^{-\lambda t} t^{\kappa-1}\, dt}{\lambda^{-\kappa}\Gamma(\kappa)}.$$

In [61]:
```
function plot_cdf_poi_gam(λ, τ; xtick=0:round(Int, λ*τ+10√(λ*τ)))
    poi = Poisson(λ*τ)
    gam(κ) = Gamma(κ, 1/λ)
    μ, σ = mean(poi), std(poi)
    κ = range(max(1, μ-4σ), μ+5σ, 1000)
    plot(κ, κ -> ccdf(poi, κ-1); label="ccdf(Poisson(λτ),κ-1)")
    plot!(κ, κ -> cdf(gam(floor(κ)), τ); label="cdf(Gamma(⌊κ⌋,1/λ),τ)", ls=:dash)
    plot!(; xlabel="κ", xtick)
    title!("λ=$λ, τ=$τ")
end
```

Out[61]: plot_cdf_poi_gam (generic function with 1 method)

In [62]:
```
plot(plot_cdf_poi_gam(1, 1), plot_cdf_poi_gam(1, 3; xtick=0:5:100);
    size=(800, 250), bottommargin=4Plots.mm)
```

Out[62]:

In [63]:
```
plot(plot_cdf_poi_gam(3, 3; xtick=0:5:1000), plot_cdf_poi_gam(3, 10; xtick=0:20:1000);
    size=(800, 250), bottommargin=4Plots.mm)
```

Out[63]:

7 ベータ二項分布

我々は以下が成立することを知っている:

(a) 期待値パラメータ λ がガンマ分布 $\mathrm{Gamma}(\alpha, \theta)$ に従う確率変数になっているようなPoisson分布 $\mathrm{Poisson}(\lambda)$ に従う確率変数が従う分布は負の二項分布 $\mathrm{NegativeBinomial}(\alpha, 1/(1+\theta))$ になる.

そしてさらに

(1) 二項分布 $\mathrm{Binomial}(n, \lambda/n)$ は $n \to \infty$ でPoisson分布 $\mathrm{Poisson}(\lambda)$ に収束する.

(2) $T_n \sim \mathrm{Beta}(\alpha, n+b)$ のとき $\theta n T_n$ が従う分布は $n \to \infty$ でガンマ分布 $\Gamma(\alpha, \theta)$ に収束する.

以上の結果から, まだ埋まっていない穴を埋めるとベータ二項分布が自然に得られる. 結果的に以下が成立することになる:

(b) 二項分布における成功確率パラメータ p がベータ分布 $\mathrm{Beta}(\alpha, \beta)$ に従う確率変数になっているような二項分布 $\mathrm{Binomial}(n, p)$ に従う確率変数が従う分布は **ベータ二項分布** になる.

(3) ベータ二項分布の連続時間極限として, 負の二項分布が得られる. しかも, その極限は(1), (2)と整合的である.

7.1 ベータ二項分布の定義

次の確率質量函数で定義される有限離散分布を **ベータ二項分布** (Beta-binomial disttribution)と呼び,

$$\mathrm{BetaBinomial}(n, \alpha, \beta)$$

と表す:

$$P(k|n, \alpha, \beta) = \binom{n}{k} \frac{B(\alpha+k, \beta+n-k)}{B(\alpha, \beta)} \qquad (k = 0, 1, \ldots, n)$$

次が成立していることに注意せよ:

$$P(k|n, \alpha, \beta) = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)} dp.$$

この式は二項分布 $\mathrm{Binomial}(n, p)$ の成功確率パラメータ p がベータ分布 $\mathrm{Beta}(\alpha, \beta)$ に従って確率的に揺らいでいる場合の分布を表す確率質量函数になっている.

ベータ二項分布の乱数は以下の手続きで生成可能である:

1. ベータ分布 $\mathrm{Beta}(\alpha, \beta)$ に従って乱数 p を生成する.
2. その p を採用した二項分布 $\mathrm{Binomial}(n, p)$ に従って乱数 k を生成する.

このように k を大量に生成したときの k の分布がベータ二項分布になる.

以下はその確認である.

In [64]:

```
function plot_bb(n, α, β; L = 10^6, kwargs...)
    # Monte Carlo シミュレーション
    P = rand(Beta(α, β), L) # ベータ分布の乱数を大量に生成
    K = @. rand(Binomial(n, P)) # 各p∈Pごとに二項分布Binomial(n,p)で乱数を生成

    # 比較のための同時プロット
    kmin, kmax = round.(Int, quantile.(BetaBinomial(n, α, β), (0.001, 0.999)))
    x = range(kmin-0.6, kmax+0.6, 1000)
    histogram(K; norm=true, alpha=0.3, bin=kmin-0.5:kmax+0.5, label="mc sim")
    plot!(x, x -> pdf(BetaBinomial(n, α, β), round(Int, x)); label="BetaBin(n,α,β)", lw=1.5)
    plot!(x, x -> pdf(Binomial(n, α/(α+β)), round(Int, x)); label="Bin(n,α/(α+β))", ls=:dash)
    title!("n=$n, α=$α, β=$β")
    plot!(; kwargs...)
end

function plot_beta(α, β)
    plot(Beta(α, β); label="", title="Beta(α=$α, β=$β)", xtick=0:0.1:1)
end
```

Out[64]: plot_beta (generic function with 1 method)

In [65]:

```
plot(plot_bb(20, 3, 7), plot_bb(20, 9, 21); size=(800, 250))
```

Out[65]:

In [66]:

```
plot(plot_bb(20, 30, 70), plot_bb(20, 90, 210); size=(800, 250))
```

Out[66]:

ベータ分布で p をランダムに生成して, その p を採用した二項分布で k をランダムに生成することを繰り返した Monte Carloシミュレーション(mc sim)の結果と, ベータ二項分布の確率質量函数のプロット(BetaBin(α, β))がぴったり一致している. 破線は二項分布のグラフである.

$\alpha/(\alpha+\beta)$ を保ったまま α, β を大きくすると分布 BetaBinomial(n, α, β) は 分布 Binomial($n, \alpha/(\alpha+\beta)$ に近付く. そうなる理由は $\alpha/(\alpha+\beta)$ を保ったまま α, β を大きくすると, 分布 Beta(α, β) が $\alpha/(\alpha+\beta)$ に集中するようになるからである.

In [67]:
```
plot(plot_beta(3, 7), plot_beta(9, 21); size=(800, 250))
```

Out[67]:

In [68]:
```
plot(plot_beta(30, 70), plot_beta(90, 210); size=(800, 250))
```

Out[68]:

7.2 ベータ二項分布の「成功確率が確率的に揺らいでいる二項分布」という解釈

ある母集団に含まれる医師を m 人選び, $i = 1, 2, \ldots, m$ のそれぞれについて i 番目の医師に n_i 人の患者を方法Aで治療してもらい, 治療に成功した人数 k_i を記録したとする. (本当は対照群を設ける必要があるが, 説明の単純化のためにここでは扱わない.)

このとき n_i 人中の治療成功数 k_i 達のデータの1つの単純なモデル化は, 医師によらずに方法Aによる共通の治療成功確率 p が決まっていて, 成功回数 k_i は共通の成功確率 p を持つ二項分布 $\mathrm{Binomial}(n_i, p)$ に従ってランダムに決まっていると想定することである.

統計モデルとして現実には確実に一致しないことが明らかな確率分布を選ぶことが多い. (しかし, 目的のために十分な精度で現実の状況をうまく把握できれば統計モデルの設定は成功したことになる.)

例えば, 上の単純な統計モデルの前提である「医師によらずに治療成功確率は同じである」という仮定は非常に疑わしいかもしれない. その場合には, 上の統計モデルの採用は失敗に終わる可能性が高い. 医師達のあいだで方法Aによる治療技術が違っている可能性を考慮する必要があるだろう.

そのような場合には, 「医師達に共通の成功確率 p が決まっている」という設定を, 「医師ごとに成功確率は違っていて, 各医師固有の成功確率はある確率分布にしたがってばらついている」という設定に緩めれば良さそうである.

そのような統計モデルの有力候補の1つがベータ二項分布である. ベータ二項分布における成功回数 k は以下のように生成されるのであった:

1. ベータ分布 $\mathrm{Beta}(\alpha, \beta)$ に従って乱数 p を生成する.
2. その p を成功確率として採用した二項分布 $\mathrm{Binomial}(n, p)$ に従って乱数 k を生成する.

ベータ分布で乱数 p を生成するところが, 選んだ医師の治療成功確率 p がベータ分布に従ってランダムに決まることを記述していると解釈可能である. 現実の医師達についてはそう単純ではないことは確実だが, このようなモデル設定を試せることも知っていると, 統計分析の成功可能性が高まるだろう.

上で見たように, α, β が大きくなると, ベータ分布がランダムに生成する成功確率 p のばらつきの幅が小さくなり, ベータ二項分布はただの二項分布に近付く. このような場合にはよりシンプルな二項分布モデルを採用した方が合理的な可能性も出て来る.

7.3 ベータ二項分布での確率の総和が 1 になることの確認

ベータ二項分布の確率質量函数の次の表示を使う:

$$P(k|n,\alpha,\beta) = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} \, \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} \, dp.$$

これを使うと,

$$\begin{aligned}\sum_{k=0}^{n} P(k|n,\alpha,\beta) &= \int_0^1 \underbrace{\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k}}_{=1} \, \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} \, dp \\ &= \int_0^1 \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)} \, dp = 1.\end{aligned}$$

一般にパラメータ θ を持つ確率質量函数 $P(x|\theta)$ と θ に関するパラメータ η を持つ確率密度函数 $p(\theta|\eta)$ に対して,

$$\tilde{P}(x|\eta) = \int P(x|\theta) p(\theta|\eta) \, d\theta$$

によって, パラメータ η を持つ確率質量函数 $\tilde{P}(x|\eta)$ を構成できる. 確率の総和が 1 になることは上の計算と同様で以下のようにして示される:

$$\sum_x \tilde{P}(x|\eta) = \int \underbrace{\sum_x P(x|\theta)}_{=1} \, p(\theta|\eta) \, d\theta = \int p(\theta|\eta) \, d\theta = 1.$$

確率質量函数 $P(x|\theta)$ を確率密度函数 $p(x|\theta)$ で置き換えた場合も同様である.

7.4 問題: ベータ二項分布の期待値と分散

ベータ二項分布 $\mathrm{Beta}(n,\alpha,\beta)$ の期待値と分散が以下のようになることを示せ: $K \sim \mathrm{Beta}(n,\alpha,\beta)$ とすると,

$$E[K] = \frac{n\alpha}{\alpha+\beta}, \quad \mathrm{var}(K) = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

注意: これらを二項分布と比較するために仮に $p = \alpha/(\alpha+\beta)$ とおくと,

$$E[K] = np, \quad \mathrm{var}(K) = np(1-p)\left(1 + \frac{n-1}{\alpha+\beta+1}\right).$$

期待値の方は二項分布と同じ形になったが, **分散の方は** 二項分布の場合の分散 $np(1-p)$ よりも $1 + (n-1)/(\alpha+\beta+1)$ 倍だけ**大きくなっている**. 大雑把に言うと, $\alpha+\beta$ が n よりずっと大きい場合にはベータ二項分布は二項分布でよく近似されるようになる.

解答例: $K \sim \mathrm{Binomial}(n,p)$ について

$$\begin{aligned}E[K(K-1)\cdots(K-m+1)] &= \sum_k k(k-1)\cdots(k-m+1)\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k} \\ &= n(n-1)\cdots(n-m+1)p^m \underbrace{\sum_k \frac{(n-m)!}{(k-m)!(n-k)!}p^{k-m}(1-p)^{n-k}}_{=(p+(1-p))^{n-m}=1} \\ &= n(n-1)\cdots(n-m+1)p^m.\end{aligned}$$

$P \sim \mathrm{Beta}(\alpha,\beta)$ について

$$
\begin{aligned}
E[P^m] &= \int_0^1 p^m \frac{p^{\alpha-1}(1-p)^{\beta-1}\,dp}{B(\alpha,\beta)} = \frac{\int_0^1 p^{\alpha+m-1}(1-p)^{\beta-1}\,dp}{B(\alpha,\beta)} \\
&= \frac{B(\alpha+m,\beta)}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+m)\Gamma(\beta)}{\Gamma(\alpha+\beta+m)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= \frac{\alpha(\alpha+1)\cdots(\alpha+m-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+m-1)}.
\end{aligned}
$$

ゆえに, $K \sim \mathrm{BetaBinomial}(n,\alpha,\beta)$ について,

$$
\begin{aligned}
E[K(K-1)\cdots(K-m+1)] &= \sum_k k(k-1)\cdots(k-m+1)\int_0^1 \binom{n}{k} p^k(1-p)^{n-1}\frac{p^{\alpha-1}(1-p)^{\beta-1}\,dp}{B(\alpha,\beta)} \\
&= \int_0^1 \left(\sum_k k(k-1)\cdots(k-m+1)\binom{n}{k}p^k(1-p)^{n-1}\right)\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\,dp \\
&= \int_0^1 n(n-1)\cdots(n-m+1)p^m\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\,dp \\
&= n(n-1)\cdots(n-m+1)\int_0^1 p^m\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\,dp \\
&= n(n-1)\cdots(n-m+1)\frac{\alpha(\alpha+1)\cdots(\alpha+m-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+m-1)}.
\end{aligned}
$$

特に,

$$
E[K] = \frac{n\alpha}{\alpha+\beta}, \qquad E[K(K-1)] = \frac{n(n-1)\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}.
$$

ゆえに

$$
E[K]-E[K]^2 = E[K](1-E[K]) = \frac{n\alpha}{\alpha+\beta}\left(1-\frac{n\alpha}{\alpha+\beta}\right) = \frac{n\alpha(\alpha+\beta-n\alpha)}{(\alpha+\beta)^2}
$$

なので

$$
\begin{aligned}
\mathrm{var}(K) &= E[K^2]-E[K]^2 = E[K(K-1)]+E[K]-E[K]^2 \\
&= \frac{n(n-1)\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} + \frac{n\alpha(\alpha+\beta-n\alpha)}{(\alpha+\beta)^2} \\
&= \frac{n\alpha((n-1)(\alpha+1)(\alpha+\beta)+(\alpha+\beta-n\alpha)(\alpha+\beta+1))}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
&= \frac{n\alpha(\alpha\beta+\beta^2+n\beta)}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{n\alpha\beta(\alpha+\beta+n)}{(\alpha+\beta)^2(\alpha+\beta+1)}.
\end{aligned}
$$

解答終

In [69]:

```
@vars a b n
var"E[K]" = n*a/(a+b)
var"E[K(K-1)]" = n*(n-1)*a*(a+1)/((a+b)*(a+b+1))
var"var(K)" = var"E[K(K-1)]" + var"E[K]" - var"E[K]"^2 ▷ factor
```

Out[69]:

$$
\frac{abn\,(a+b+n)}{(a+b)^2\,(a+b+1)}
$$

7.5 問題: ベータ二項分布 = 負の超幾何分布

ベータ二項分布の確率質量函数

$$
P(k|n,\alpha,\beta) = \binom{n}{k}\frac{B(\alpha+k,\beta+n-k)}{B(\alpha,\beta)} \qquad (k=0,1,\ldots,n)
$$

は以下のように書き換えられることを示せ:

$$
\begin{aligned}
P(k|n,\alpha,\beta) &= \binom{\alpha+k-1}{k}\binom{\beta+(n-k)-1}{n-k}\Big/\binom{\alpha+\beta+n-1}{n} \\
&= \binom{-\alpha}{k}\binom{-\beta}{n-k}\Big/\binom{-\alpha-\beta}{n}.
\end{aligned}
$$

注意: この形の確率質量函数は別に定義される有限離散分布のパラメータ $s, f, n \in \mathbb{Z}_{\geq 0}, n \leq s+f$ に関する **超幾何分布** $\mathrm{Hypergeometric}(s, f, n)$ の 確率質量函数

$$k \mapsto \binom{s}{k}\binom{f}{n-k}\Big/\binom{s+f}{n}.$$

における s, f に負の値 $-\alpha, -\beta$ を形式的に代入したものに一致しているので, ベータ二項分布は **負の超幾何分布** と呼ばれることがある.

注意: 上に書いた超幾何分布の確率は, s 個の赤い球と f 個の白い球が入っている袋から n 個をまとめて無作為に取り出したとき, その中にちょうど k 個の赤い球が含まれる確率になっている(s はsuccessの, f はfailureの頭文字). 実際, 分母の $\binom{n+m}{n}$ は $s+f$ 個から n 個を選ぶときの組み合わせの数であり, 分子は s 個の赤い球から k 個を選び, f 個の白い球から $n-k$ 個選ぶ組み合わせの個数であり, 分子を分母で割れば, n 個中ちょうど k 個が赤い球である確率になる. 超幾何分布自体は「超幾何」という難しそうな形容詞が付いているが, 初等的に理解できる確率分布である.

注意: 負の二項分布の確率質量函数は

$$P(m|\alpha, p) = \binom{\alpha+m-1}{m}p^{\alpha}(1-p)^m = (-1)^m\binom{-\alpha}{m}p^{\alpha}(1-p)^m$$

と書けるのであった. 二項分布と負の二項分布の関係は超幾何分布と負の超幾何分布=ベータ二項分布の関係に似ている.

解答例:

$$\begin{aligned}
P(k|n, \alpha, \beta) &= \binom{n}{k}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\frac{\Gamma(\beta+n-k)}{\Gamma(\beta)} \\
&= \frac{n!}{k!(n-k)!}\frac{\alpha(\alpha+1)\cdots(\alpha+k-1)\cdot\beta(\beta+1)\cdots(\beta+n-k-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+n-1)} \\
&= \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{k!}\frac{\beta(\beta+1)\cdots(\beta+n-k-1)}{(n-k)!} \\
&\times \frac{n!}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+n-1)} \\
&= \binom{\alpha+k-1}{k}\binom{\beta+(n-k)-1}{n-k}\Big/\binom{\alpha+\beta+n-1}{n}
\end{aligned}$$

さらに,

$$\begin{aligned}
(-1)^k\binom{-a}{k} &= (-1)^k\frac{(-a)(-a-1)\cdots(-a-k+1)}{k!} \\
&= \frac{a(a+1)\cdots(a+k-1)}{k!} = \binom{a+k-1}{k}
\end{aligned}$$

なので,

$$P(k|n, \alpha, \beta) = \binom{-\alpha}{k}\binom{-\beta}{n-k}\Big/\binom{-\alpha-\beta}{n}.$$

解説終

7.6 超幾何分布, 二項分布, ベータ二項分布の統一的な理解

この節の内容は本質的に [Pólya urn model (https://www.google.com/search?q=P%C3%B3lya+urn+model)](https://www.google.com/search?q=P%C3%B3lya+urn+model) の話である. この節では α と β は 0 以上の整数であり, $\alpha+\beta > 0$ であるとし, δ は -1 以上の整数であると仮定する.

最初の状態では袋の中に赤い球が α 個, 白い球が β 個入っているとし, 以下の操作を n 回繰り返そう:

1. 袋の中から玉を無作為に取り出し, その色を記録する.
2. その玉が赤いならば袋の中に赤い球を $\delta+1$ 個戻し(袋の中の赤い球は δ 個増える($\delta=-1$ ならば 1 個減る)), 白いならば袋の中に白い球を $\delta+1$ 個戻す(袋の中の白い球は δ 個増える($\delta=-1$ ならば 1 個減る)).

このとき, n 回中 k 回赤い球が取り出される確率は次のように書ける:

$$P(k|n, \alpha, \beta, \delta) = \binom{n}{k}\frac{\alpha(\alpha+\delta)\cdots(\alpha+(k-1)\delta)\cdot\beta(\beta+\delta)\cdots(\beta+(n-k-1)\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)\cdots(\alpha+\beta+(n-1)\delta)} \quad (k=0,1,\ldots,n).$$

注意: $\delta=-1$ のときであっても, $k > \alpha$ ならば

$$\alpha(\alpha+\delta)\cdots(\alpha+(k-1)\delta) = \alpha(\alpha-1)\cdots(\alpha-\alpha)\cdots(\alpha-(k-1)) = 0$$

となるので $p(k|n,\alpha,\beta,-1)=0$ となり, 同様にして $n-k>\beta$ ならば $p(k|n,\alpha,\beta,-1)=0$ となることがわかる. これより, $\delta=-1$ であっても $p(k|n,\alpha,\beta,-1)$ $(k=0,1,\dots,n)$ が負の値にならないことがわかる.

注意: $\delta>0$ ならば分子分母を $\delta^k\delta^{n-k}=\delta^n$ で割ることによって形式的に δ の部分を 1 にでき, α,β を 0 以上の正の有理数にできるので, $\delta=1$ の場合が本質的だと感が得られる. そして, その場合の極限として, α,β が 0 以上の任意の実数で $\delta=1$ の場合も理解できる.

注意: α,β を球の個数とみなす解釈を止めて, α,β,δ は任意の 0 以上の実数で $\alpha+\beta>0$ を満たすという設定にしても, 上の $p(k|n,\alpha,\beta,\delta)$ の定義式は確率質量函数を与える. その場合には, 最初の状態ではランダムに赤い玉が確率 $\alpha/(\alpha+\beta)$ で, 白い球が確率 $\beta/(\alpha+\beta)$ で出て来て, 赤い玉が出たらその次に赤い玉の出る確率が $(\alpha+\delta)/(\alpha+\beta+\delta)$ に増加し, 白い玉が出たらその次に白い玉の出る確率が $(\beta+\delta)/(\alpha+\beta+\delta)$ に増加し, その後も同様とすると考えればよい. このような設定の場合にも, 上の注意で述べたのと同じ理由で, $\delta>0$ の場合は $\delta=1$ の場合に帰着される. 数学的な設定としてなら, 袋の中に非整数個の赤い球と白い球が入っており, 取り出される確率がそれぞれの個数に比例すると考えてもよいだろう.

袋に取り出した玉を戻さない非復元抽出の場合 ($\delta=-1$): n 回中 k 回赤い球が取り出される確率の分布は超幾何分布 $\mathrm{Hypergeometric}(\alpha,\beta,n)$ になる:

$$\begin{aligned}P(k|n,\alpha,\beta,-1)&=\binom{n}{k}\frac{\alpha(\alpha-1)\cdots(\alpha-(k-1))\cdot\beta(\beta-1)\cdots(\beta-(n-k-1)}{(\alpha+\beta)(\alpha+\beta-1)\cdots(\alpha+\beta-(n-1))}\\&=\binom{\alpha}{k}\binom{\beta}{n-k}\Big/\binom{\alpha+\beta}{n}.\end{aligned}$$

これを超幾何分布 $\mathrm{Hypergeometric}(\alpha,\beta,n)$ の定義だと考えてよい.

袋に取り出した玉をそのまま玉を戻す復元抽出の場合 ($\delta=0$): n 回中 k 回赤い球が取り出される確率の分布は二項分布 $\mathrm{Binomial}(n,\alpha/(\alpha+\beta))$ になる:

$$P(k|n,\alpha,\beta,0)=\binom{n}{k}\frac{\alpha^k\beta^{n-k}}{(\alpha+\beta)^n}=\binom{n}{k}\left(\frac{\alpha}{\alpha+\beta}\right)^k\left(1-\frac{\alpha}{\alpha+\beta}\right)^{n-k}.$$

袋に取り出した玉と同じ色の玉を2個戻して袋の中のその色の玉が1個増えるようにした場合 ($\delta=0$): n 回中 k 回赤い球が取り出される確率の分布はベータ二項分布 $\mathrm{BetaBinomial}(n,\alpha,\beta)$ になる:

$$\begin{aligned}P(k|n,\alpha,\beta,1)&=\binom{n}{k}\frac{\alpha(\alpha+1)\cdots(\alpha+(k-1))\cdot\beta(\beta+1)\cdots(\beta+(n-k-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+(n-1))}\\&=\binom{n}{k}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\frac{\Gamma(\beta+n-k)}{\Gamma(\beta)}\\&=\binom{n}{k}\frac{B(\alpha+k,\beta+n-k)}{B(\alpha,\beta)}.\end{aligned}$$

袋の中の赤い球の個数と白い球の個数の初期値である α,β が n よりも非常に大きい場合には上の3つの場合の確率はほぼ同じになる.

ベータ二項分布の場合には, 赤い球を引けた人はその次に赤い球を引く確率が少し上がる. この意味でベータ二項分布は "The rich get richer and the poor get poorer. (https://www.google.com/search?q=the+rich+get+richer+and+the+poor+get+poorer)" (「金持ちはより金持ちに、貧乏人はより貧乏に」)という状況のモデル化になっていることがわかる.

In [70]:
```
function plot_hg_bin_bb(n, α, β; xtick=0:n)
    @assert n < α + β "n must be less than α + β."
    hg  = Hypergeometric(α, β, n)
    bin = Binomial(n, α/(α+β))
    bb  = BetaBinomial(n, α, β)
    μ, σ = mean(bb), std(bb)
    x = range(max(-0.5, μ-4σ), min(n+0.5, μ+4σ), 1000)
    plot(; xtick)
    plot!(x, x -> pdf(hg,  round(Int, x)); label="Hypergeom")
    plot!(x, x -> pdf(bin, round(Int, x)); label="Binomial", ls=:dash)
    plot!(x, x -> pdf(bb,  round(Int, x)); label="BetaBin",  ls=:dashdot)
    title!("n=$n, α=$α, β=$β")
end
```

Out[70]:
```
plot_hg_bin_bb (generic function with 1 method)
```

以下は $\alpha/(\alpha+\beta)=1/4$ を保ったまま α,β を大きくして行った場合の超幾何分布, 二項分布, ベータ二項分布の比較である. α,β を大きくして行くと, 3つの分布は一致して行く(直観的には当然そうなるべき結果).

In [71]:
```
plot(plot_hg_bin_bb(10, 3, 9), plot_hg_bin_bb(10, 10, 30); size=(800, 250))
```

Out[71]:

In [72]:
```
plot(plot_hg_bin_bb(10, 30, 90), plot_hg_bin_bb(10, 100, 300); size=(800, 250))
```

Out[72]:

7.7 問題: 超幾何分布, 二項分布, ベータ二項分布の期待値と分散の統一的な公式

前節の確率質量函数 $P(k|n,\alpha,\beta,\delta)$ を持つ確率変数 K について次が成立することを示せ:

$$E[K] = \frac{n\alpha}{\alpha+\beta}, \quad \mathrm{var}(K) = \frac{\alpha\beta n(\alpha+\beta+n\delta)}{(\alpha+\beta)^2(\alpha+\beta+\delta)}.$$

注意: 特にこれの $\delta = -1$, $\alpha = s$, $\beta = f$ の場合より, 超幾何分布 Hypergeometric(s,f,n) の期待値と分散が次のようになることがわかる: 確率変数 $K \sim$ Hypergeometric(s,f,n) について

$$E[K] = \frac{ns}{s+f}, \quad \mathrm{var}(K) = \frac{sfn(s+f-n)}{(s+f)^2(s+f-1)}.$$

解答例: 問題を一般化して, 前節の確率質量函数

$$P(k|n,\alpha,\beta,\delta) = \binom{n}{k}\frac{\alpha(\alpha+\delta)\cdots(\alpha+(k-1)\delta)\cdot\beta(\beta+\delta)\cdots(\beta+(n-k-1)\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)\cdots(\alpha+\beta+(n-1)\delta)}.$$

を持つ確率変数 K について考える. このとき,

$$\begin{aligned}
&E[K(K-1)\cdots(K-m+1)] \\
&= \sum_k k(k-1)\cdots(k-m+1)\binom{n}{k}\frac{\alpha(\alpha+\delta)\cdots(\alpha+(k-1)\delta)\cdot\beta(\beta+\delta)\cdots(\beta+(n-k-1)\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)\cdots(\alpha+\beta+(n-1)\delta)} \\
&= n(n-1)\cdots(n-m+1)\frac{\alpha(\alpha+\delta)\cdots(\alpha+(m-1)\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)\cdots(\alpha+\beta+(m-1)\delta)} \\
&\quad\times\underbrace{\sum_k\binom{n-m}{k-m}\frac{(\alpha+m\delta)(\alpha+(m+1)\delta)\cdots(\alpha+(k-1)\delta)\cdot\beta(\beta+\delta)\cdots(\beta+(n-k-1)\delta)}{(\alpha+\beta+m)(\alpha+\beta+(m+1)\delta)\cdots(\alpha+\beta+(n-1)\delta)}}_{=\sum_k P(k|n-m,\alpha+m,\beta)=1} \\
&= n(n-1)\cdots(n-m+1)\frac{\alpha(\alpha+\delta)\cdots(\alpha+(m-1)\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)\cdots(\alpha+\beta+(m-1)\delta)}.
\end{aligned}$$

例えば,

$$E[K] = \frac{n\alpha}{\alpha+\beta}, \quad E[K(K-1)] = \frac{n(n-1)\alpha(\alpha+\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)}.$$

これを使って少し計算すると,

$$\mathrm{var}(K) = E[K^2] - E[K]^2 = E[K(K-1)] + E[K] - E[K]^2 = \frac{\alpha\beta n(\alpha+\beta+n\delta)}{(\alpha+\beta)^2(\alpha+\beta+\delta)}$$

となることがわかる(以下のセルでのコンピュータによる代数計算も参照せよ).

解答終

In [73]:
```
@vars a b d n
var"E[K]" = n*a/(a+b)
var"E[K(K-1)]" = n*(n-1)*a*(a+d)/((a+b)*(a+b+d))
var"var(K)" = var"E[K(K-1)]" + var"E[K]" - var"E[K]"^2 ▷ factor
```

Out[73]:

$$\frac{abn\,(a+b+dn)}{(a+b)^2\,(a+b+d)}$$

7.8 問題: ベータ二項分布の極限として負の二項分布が得られる

$\alpha, \theta > 0$ であるとし, ベータ二項分布 $\mathrm{BetaBinomial}(L, \alpha, \beta)$ の確率質量函数

$$P(m|L,\alpha,\beta) = \binom{L}{m}\frac{B(\alpha+m, \beta+L-m)}{B(\alpha,\beta)} \qquad (m = 0, 1, \dots, L)$$

は, $\beta = L/\theta$ とおいて, $L \to \infty$ とすると,

$$P(m|L,\alpha,L/\theta) \to \binom{\alpha+m-1}{m}\left(\frac{1}{1+\theta}\right)^{\alpha}\left(1-\frac{1}{1+\theta}\right)^{m}$$

と負の二項分布 $\mathrm{NegativeBinomial}(\alpha, 1/(1+\theta))$ の確率質量函数に収束することを示せ.

解答例1: $n \to \infty$ で $n^{\alpha}B(\alpha, n+b) \to \Gamma(\alpha)$ となることを使うと, $L \to \infty$ のとき,

$$\begin{aligned}
P(m|L,\alpha,L/\theta) &= \binom{L}{m}\frac{B(\alpha+m, L/\theta+L-m)}{B(\alpha,L/\theta)} \\
&= \frac{L(L-1)\cdots(L-m+1)}{m!}\frac{(L/\theta)^{\alpha}}{((1+\theta)/\theta)L)^{\alpha+m}} \\
&\quad \times \frac{(((1+\theta)/\theta)L)^{\alpha+m}B(\alpha+m, ((1+\theta)/\theta)L-m)}{(L/\theta)^{\alpha}B(\alpha,L/\theta)} \\
&= \underbrace{\frac{L(L-1)\cdots(L-m+1)}{L^m}}_{\to 1} \\
&\quad \times \underbrace{\frac{(((1+\theta)/\theta)L)^{\alpha+m}B(\alpha+m, ((1+\theta)/\theta)L-m)}{m!\,(L/\theta)^{\alpha}B(\alpha,L/\theta)}}_{\to\Gamma(\alpha+m)/(m!\Gamma(\alpha))}\left(\frac{1}{1+\theta}\right)^{\alpha}\left(\frac{\theta}{1+\theta}\right)^{m} \\
&\to \frac{\Gamma(\alpha+m)}{m!\Gamma(\alpha)}\left(\frac{1}{1+\theta}\right)^{\alpha}\left(\frac{\theta}{1+\theta}\right)^{m} \\
&= \binom{\alpha+m-1}{m}\left(\frac{1}{1+\theta}\right)^{\alpha}\left(1-\frac{1}{1+\theta}\right)^{m}.
\end{aligned}$$

解答終

解答例2: せっかくなので, 積を取って積分してベータ二項分布の確率質量函数を作る前の二項分布の確率質量函数とベータ分布の確率密度函数の連続時間極限との整合性も確認してしまおう.

ベータ分布 $\mathrm{Beta}(\alpha, \beta)$ の確率密度函数 $\times dp$ については,

$$\beta = L/\theta, \quad p = \lambda/L, \quad dp = d\lambda/L$$

とおくと, $L \to \infty$ のとき以下が成立している:

$$
\begin{aligned}
\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\,dp &= \frac{(\lambda/L)^{\alpha-1}(1-\lambda/L)^{L/\theta-1}}{B(\alpha,L/\theta)}\frac{d\lambda}{L} \\
&= \frac{\lambda^{\alpha-1}(1-\lambda/L)^{L/\theta-1}}{L^{\alpha}B(\alpha,L/\theta)}\,d\lambda \\
&= \frac{1}{\theta^{\alpha}\underbrace{(L/\theta)^{\alpha}B(\alpha,L/\theta)}_{\to\Gamma(\alpha)}}\underbrace{\left(1-\frac{\lambda}{L}\right)^{L/\theta-1}}_{\to\exp(-\lambda/\theta)}\lambda^{\alpha-1}\,d\lambda \\
&\to \frac{1}{\theta^{\alpha}\Gamma(\alpha)}e^{-\lambda/\theta}\lambda^{\alpha-1}\,d\lambda.
\end{aligned}
$$

$L\to\infty$ のとき, ベータ分布はこのように λ に関するガンマ分布 $\mathrm{Gamma}(\alpha,\theta)$ に近付く. p は 0 と 1 のあいだを動くが, $\lambda = Lp$ なので $L\to\infty$ の極限で λ は正の実数全体を動くことになる.

二項分布 $\mathrm{Binomial}(n,p)$ の確率質量函数については,

$$
n = L, \quad p = \lambda/L, \quad k = m
$$

とおくと, $L\to\infty$ のとき以下が成立している:

$$
\begin{aligned}
\binom{n}{k}p^k(1-p)^{n-k} &= \binom{L}{m}\left(\frac{\lambda}{L}\right)^m\left(1-\frac{\lambda}{L}\right)^{L-m} \\
&= \underbrace{\frac{L(L-1)\cdots(L-m+1)}{L^m}}_{\to 1}\underbrace{\left(1-\frac{\lambda}{L}\right)^{L-m}}_{\to\exp(-\lambda)}\frac{\lambda^m}{m!} \\
&\to e^{-\lambda}\frac{\lambda^m}{m!}.
\end{aligned}
$$

$L\to\infty$ のとき, 二項分布はこのように m に関するPoisson分布 $\mathrm{Poisson}(\lambda)$ に近付く.

したがって, ベータ二項分布 $\mathrm{Beta}(L,\alpha,L/\theta)$ の確率質量函数

$$
P(m|L,\alpha,L/\theta) = \int_0^1 \binom{L}{m}p^k(1-p)^{L-m}\frac{p^{\alpha-1}(1-p)^{L/\theta-1}}{B(\alpha,L/\theta)}\,dp
$$

について, $p=\lambda/L$ とおくことによって, $L\to\infty$ で以下が成立することがわかる:

$$
\begin{aligned}
P(m|L,\alpha,\beta) &\to \int_0^\infty e^{-\lambda}\frac{\lambda^m}{m!}\frac{e^{-\lambda/\theta}\lambda^{\alpha-1}}{\theta^{\alpha}\Gamma(\alpha)}\,d\lambda \\
&= \frac{1}{m!\theta^{\alpha}\Gamma(\alpha)}\int_0^\infty e^{-((1+\theta)/\theta)\lambda}\lambda^{\alpha+m-1}\,d\lambda \\
&= \frac{1}{m!\theta^{\alpha}\Gamma(\alpha)}\left(\frac{\theta}{1+\theta}\right)^{\alpha+m}\Gamma(\alpha+m) \\
&= \frac{\Gamma(\alpha+m)}{m!\Gamma(\alpha)}\left(\frac{1}{1+\theta}\right)^{\alpha}\left(\frac{\theta}{1+\theta}\right)^{m} \\
&= \binom{\alpha+m-1}{m}\left(\frac{1}{1+\theta}\right)^{\alpha}\left(1-\frac{1}{1+\theta}\right)^{m}.
\end{aligned}
$$

$L\to\infty$ のとき, ベータ二項分布はこのように m に関する負の二項分布 $\mathrm{NegativeBinomial}(\alpha,1/(1+\theta))$ に近付く.

解答終

In [74]:

```julia
function plot_bbnb(L, α, θ; kwargs...)
    bb = BetaBinomial(L, α, L/θ)
    nb = NegativeBinomial(α, 1/(1+θ))
    μ, σ = mean(nb), std(nb)
    m = range(max(-1, μ-4σ), μ+4σ, 1000)
    plot(m, m -> pdf(bb, round(m)); label="BetaBin(L,α,L/θ)")
    plot!(m, m -> pdf(nb, round(m)); label="NegBin(α,1/(1+θ))", ls=:dash)
    title!("L=$L, α=$α, θ=$θ, p=1/(1+θ)=1/$(1+θ)")
    plot!(; kwargs...)
end
```

Out[74]:

```
plot_bbnb (generic function with 1 method)
```

In [75]:
```
plot(plot_bbnb.((10, 30), 3, 2)...; size=(800, 250))
```

Out[75]:

In [76]:
```
plot(plot_bbnb.((100, 300), 3, 2)...; size=(800, 250))
```

Out[76]:

7.9 問題: 二項分布とベータ二項分布のモーメント母函数

(1) 二項分布のモーメント母函数: 次が成立することを示せ:

$$K \sim \mathrm{Binomial}(n, p) \implies E[e^{tK}] = (1 - (1 - e^t)p)^n.$$

(2) ベータ二項分布のモーメント母函数が本質的にGaussの超幾何函数になっていることを示せ. すなわち, 次が成立することを示せ:

$$K \sim \mathrm{BetaBonomial}(n, \alpha, \beta) \implies E[e^{tK}] = {}_2F_1(-n, \alpha; \alpha + \beta; 1 - e^t).$$

ここで, ${}_2F_1(a, b; c; z)$ はGaussの超幾何函数で次のように定義される:

$${}_2F_1(a, b; c; z) = \frac{1}{B(b, c-b)} \int_0^1 (1 - zt)^{-a} t^{b-1} (1 - t)^{c-b-1}\, dt. \qquad (\mathrm{Re}\, c > \mathrm{Re}\, b > 0)$$

解答例: (1) $K \sim \mathrm{Binomial}(n, p)$ のとき, 二項定理より,

$$E[e^{tK}] = \sum_{k=0}^{n} \binom{n}{k} e^{tk} p^k (1 - p)^{n-k} = (e^t p + 1 - p)^n = (1 - (1 - e^t)p)^n.$$

(2) $K \sim \mathrm{BetaBonomial}(n, \alpha, \beta)$ と仮定する. 負の二項分布の確率質量函数の

$$P(k|n, \alpha, \beta) = \int_0^1 \binom{n}{k} p^k (1 - p)^{n-k} \frac{p^{\alpha-1}(1 - p)^{\beta-1}}{B(\alpha, \beta)}\, dp$$

という表示を使うと, (1)の計算より,

$$
\begin{aligned}
E[e^{tK}] &= \sum_{k=0}^{n} e^{tk} \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\, dp \\
&= \int_0^1 \left(\sum_{k=0}^{n} \binom{n}{k} e^{tk} p^k (1-p)^{n-k} \right) \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\, dp \\
&= \int_0^1 (1-(1-e^t)p)^n \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\, dp \\
&= \frac{1}{B(\alpha,\beta)} \int_0^1 (1-(1-e^t)p)^n p^{\alpha-1}(1-p)^{\beta-1}\, dp \\
&= {}_2F_1(-n, \alpha; \alpha+\beta; 1-e^t).
\end{aligned}
$$

解答終

注意: 二項分布もベータ分布も統計学における最も基本的な確率分布である. それらを組み合わせて作ったベータ二項分布から, モーメント母函数の形式でGaussの超幾何函数の積分表示式が自然に出て来ることはちょっと面白い. 超幾何函数のような特殊函数は主に微分方程式との関連で自然に現れて来ることが多いのだが, この場合には微分方程式と全く無関係に超幾何函数が自然に出て来ている. 統計学によく出て来る特殊函数達に関する話題は「微分方程式抜きの特殊函数論」というような趣がある.

文献: 超幾何函数から作られる確率分布については以下の文献を参照せよ:

- 清水良一. 一般超幾何分布. 統計数理研究所彙報, 第16巻, 第2号, 147-165, 1968-12. [link (https://ismrepo.ism.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=32361&item_no=1&page_id=13&block_id=
- Shibuya, Masaaki and Shimizu, Ryoichi. Classification of the Generalized Hypergeometric family of distributions. Keio Science and Technology Reports 34 (1), 1-38, 1981-09. [Google Scholar (https://scholar.google.co.jp/scholar?cluster=14585775532723000776)]
- Shibuya, Masaaki and Shimizu, Ryoichi. The generalized hypergeometric family of distributions. Annals of the Institute of Statistical Mathematics, volume 33, pages 177–190 (1981). [Google Scholar (https://scholar.google.co.jp/scholar?cluster=553693145216038105 7)]

8 ベータ負の二項分布とベータプライム分布

この節ではベータ負の二項分布 $\mathrm{BetaNegativeBinomial}(k,\alpha,\beta)$ という離散分布を定義し, その連続時間極限としてベータプライム分布が現れることを説明する.

8.1 ベータ負の二項分布の定義

$k,\alpha,\beta > 0$ と仮定する. $m = 0,1,2,\ldots$ に関する確率質量函数

$$
P(m|k,\alpha,\beta) = \binom{k+m-1}{m} \frac{B(\alpha+k,\beta+m)}{B(\alpha,\beta)}
$$

で定義される離散分布を **ベータ負の二項分布** と呼び,

$$
\mathrm{BetaNegativeBinomial}(k,\alpha,\beta)
$$

と書く. ベータ負の二項分布の確率質量函数は次のようにも表される:

$$
P(m|k,\alpha,\beta) = \int_0^1 \binom{k+m-1}{m} p^k (1-p)^m \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\, dp.
$$

8.2 問題: ベータ負の二項分布の期待値と分散

$M \sim \mathrm{BetaNegativeBinomial}(k,\alpha,\beta)$ のとき, 次が成立することを示せ:

$$
E[M] = \frac{k\beta}{\alpha-1}, \qquad \mathrm{var}(M) = \frac{\alpha(\alpha+k-1)\beta(\alpha+\beta-1)}{(\alpha-1)^2(\alpha-2)}.
$$

解答例: 負の二項分布に従う確率変数 $M \sim \mathrm{NegativeBinomial}(k,p)$ の函数 $f(M)$ の期待値を $E[M|k,p]$ と書くとき,

$$
E[M(M-1)\cdots(M-i+1)|k,p] = k(k+1)\cdots(k+i-1)\frac{(1-p)^i}{p^i}
$$

が成立するのであった. ベータ負の二項分布はパラメータ p がベータ分布 $\mathrm{Beta}(\alpha,\beta)$ に従っている場合の負の二項分布に等しいので, $M \sim \mathrm{BetaNegativeBinomial}(k,\alpha,\beta)$ のとき,

$$E[f(M)] = \int_0^1 E[f(M)|k,p]\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\,dp$$

が成立している. これと上を合わせると以下を得る:

$$\begin{aligned}
&E[M(M-1)\cdots(M-i+1)] \\
&= k(k+1)\cdots(k+i-1)\frac{B(\alpha-i,\beta+i)}{B(\alpha,\beta)} \\
&= k(k+1)\cdots(k+i-1)\frac{\Gamma(\alpha-i)\Gamma(\beta+i)}{\Gamma(\alpha+\beta)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \\
&= k(k+1)\cdots(k+i-1)\frac{\beta(\beta+1)\cdots(\beta+i-1)}{(\alpha-1)(\alpha-2)\cdots(\alpha-i)}.
\end{aligned}$$

特に,

$$E[M] = \frac{k\beta}{\alpha-1}, \quad E[M(M-1)] = \frac{k(k+1)\beta(\beta+1)}{(\alpha-1)(\alpha-2)}.$$

ゆえに,

$$\begin{aligned}
\mathrm{var}(M) &= E[M(M-1)] + E[M] - E[M]^2 \\
&= \frac{k(k+1)\beta(\beta+1)}{(\alpha-1)(\alpha-2)} + \frac{k\beta}{\alpha-1} - \frac{k^2\beta^2}{(\alpha-1)^2} \\
&= \frac{k\beta}{(\alpha-1)^2(\alpha-2)}((k+1)(\beta+1)(\alpha-1) + (\alpha-1)(\alpha-2) - k\beta(\alpha-1)) \\
&= \frac{k\beta}{(\alpha-1)^2(\alpha-2)}((\alpha-1)^2 + (k+\beta)(\alpha-1) + k\beta) \\
&= \frac{k\beta}{(\alpha-1)^2(\alpha-2)}(\alpha-1+k)(\alpha-1+\beta) = \frac{k(k+\alpha-1)\beta(\alpha+\beta-1)}{(\alpha-1)^2(\alpha-2)}.
\end{aligned}$$

解答終

In [77]:
```
@vars k α β
var"E[M]" = k*β/(α-1)
var"E[M(M-1)]" = k*(k+1)*β*(β+1)/((α-1)*(α-2))
var"var(M)" = var"E[M(M-1)]" + var"E[M]" - var"E[M]"^2 ▷ factor
```

Out[77]:

$$\frac{k\beta\,(k+\alpha-1)\,(\alpha+\beta-1)}{(\alpha-2)\,(\alpha-1)^2}$$

8.3 負の二項分布とベータ負の二項分布の期待値と分散の比較

負の二項分布と期待値と分散を比較しよう. $M' \sim \mathrm{NegativeBinomial}(k,p)$ のとき,

$$E[M'] = \frac{k(1-p)}{p}, \quad \mathrm{var}(M') = \frac{k(1-p)}{p^2}.$$

これに,

$$p = \frac{\alpha-1}{\alpha+\beta-1}, \quad 1-p = \frac{\beta}{\alpha+\beta-1}$$

を代入すると,

$$E[M'] = \frac{k\beta}{\alpha-1} = E[M], \quad \mathrm{var}(M') = \frac{k\beta(\alpha+\beta-1)}{(\alpha-1)^2}.$$

ゆえに, $k > 0$ より,

$$E[M] = \frac{\alpha+k-1}{\alpha-2}E[M'] > \frac{\alpha-1}{\alpha-2}E[M'] > E[M'].$$

ベータ負の二項分布の期待値と負の二項分布の期待値を上のようにして合わせると, ベータ負の二項分布の分散は負の二項分布の分散よりも大きくなる.

8.4 負の二項分布の連続時間極限の復習

任意に実数 b を取り,

$$b + m = Lt, \quad p = \frac{\lambda}{L}$$

とおいて, $L \to \infty$ とすると,

$$\begin{aligned}
\binom{k+m-1}{m} p^k (1-p)^m &= \frac{1}{(k+m)B(k, m+1)} p^k (1-p)^m \, dm \\
&= \frac{1}{(Lt+k-b)\, B(k, Lt-b+1)} \frac{\lambda^k}{L^k} \left(1 - \frac{\lambda}{L}\right)^{Lt-b} L \, dt \\
&= \frac{Lt}{Lt+k-b} \frac{(Lt^{k-1})}{(Lt)^k B(k, Lt-b+1)} \frac{\lambda^k}{L^{k-1}} \left(1 - \frac{\lambda}{L}\right)^{Lt-b} dt \\
&= \frac{Lt}{Lt-b} \frac{\lambda^k}{(Lt)^k B(k, Lt-b+1)} \left(1 - \frac{\lambda}{L}\right)^{Lt-b} t^{k-1} \, dt \\
&\to \frac{\lambda^k}{\Gamma(k)} e^{-\lambda t} t^{k-1} \, dt.
\end{aligned}$$

すなわち, $L \to \infty$ のとき,

$$\frac{b + \mathrm{NegativeBinomial}(k, \lambda/L)}{L} \to \mathrm{Gamma}(k, 1/\lambda).$$

注意: $\mathrm{NegativeBinomial}(k, p)$ の期待値は $k(1-p)/p$ なので, $k + \mathrm{NegativeBinomial}(k, p)$ の期待値は k/p になる. 一方, $\mathrm{Gamma}(k, 1/\lambda)$ の期待値は k/λ である. ゆえに $(k + \mathrm{NegativeBinomial}(k, \lambda/L))/L$ の期待値は $\mathrm{Gamma}(k, 1/\lambda)$ の期待値は一致する. だから, 上の連続時間極限では $b = k$ とすると収束先と期待値を一致させたままで極限を取ることができる.

注意: $\mathrm{NegativeBinomial}(k, p)$ の分散は $k(1-p)/p^2$ なので, $(k + \mathrm{NegativeBinomial}(k, \lambda/L))/L$ の分散は $(k/\lambda^2)(1 - \lambda/L)$ になる. 一方, $\mathrm{Gamma}(k, 1/\lambda)$ の分散は k/λ^2 になるので, 上の連続極限の過程は分散が単調増加しながらの収束になっている.

In [78]:
```
k, λ, L = 3, 4, 40
@show mean((k + NegativeBinomial(k, λ/L))/L), mean(Gamma(k, 1/λ))
@show var((k + NegativeBinomial(k, λ/L))/L), var(Gamma(k, 1/λ));
```

```
(mean((k + NegativeBinomial(k, λ / L)) / L), mean(Gamma(k, 1 / λ))) = (0.75, 0.75)
(var((k + NegativeBinomial(k, λ / L)) / L), var(Gamma(k, 1 / λ))) = (0.16875, 0.1875)
```

8.5 ベータ分布の極限の復習

$$p = \frac{\lambda}{L}, \quad \beta = \frac{L}{\theta}$$

とおいて, $L \to \infty$ とすると,

$$\begin{aligned}
\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)} dp &= \frac{L^\alpha}{\theta^\alpha (L/\theta)^\alpha B(\alpha, L/\theta)} \frac{\lambda^{\alpha-1}}{L^{\alpha-1}} \left(1 - \frac{\lambda}{L}\right)^{L/\theta-1} \frac{d\lambda}{L} \\
&= \frac{1}{\theta^\alpha (L/\theta)^\alpha B(\alpha, L/\theta)} \left(1 - \frac{\lambda}{L}\right)^{L/\theta-1} \lambda^{\alpha-1} \, d\lambda \\
&\to \frac{1}{\theta^\alpha \Gamma(\alpha)} e^{-\lambda/\theta} \lambda^{\alpha-1} \, d\lambda.
\end{aligned}$$

すなわち, $L \to \infty$ のとき,

$$L \, \mathrm{Beta}(\alpha, L/\theta) \to \mathrm{Gamma}(\alpha, \theta).$$

注意: $\mathrm{Beta}(\alpha, \beta)$ の期待値は $\alpha/(\alpha+\beta)$ なので $L \, \mathrm{Beta}(\alpha, L/\theta)$ の期待値は $\alpha\theta/(1 + \alpha\theta/L)$ になる. これは $L \to \infty$ で $\mathrm{Gamma}(\alpha, \theta)$ の期待値 $\alpha\theta$ に単調増加しながら収束する.

注意: $\mathrm{Beta}(\alpha, \beta)$ の分散は $\alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$ なので $L \, \mathrm{Beta}(\alpha, L/\theta)$ の分散は $\alpha\theta^2/((1+\alpha\theta/L)^2(1+(\alpha+1)\theta/L))$ になる. これは $L \to \infty$ で $\mathrm{Gamma}(\alpha, \theta)$ の分散 $\alpha\theta^2$ に単調増加しながら収束する.

In [79]:
```
α, θ, L = 3, 4, 2^10
@show mean(L*Beta(α, L/θ)), mean(Gamma(α, θ))
@show var(L*Beta(α, L/θ)), var(Gamma(α, θ));
```

```
(mean(L * Beta(α, L / θ)), mean(Gamma(α, θ))) = (11.86100386100386, 12.0)
(var(L * Beta(α, L / θ)), var(Gamma(α, θ))) = (46.17301746568156, 48.0)
```

8.6 ベータ負の二項分布の連続極限

任意に実数 b を取り,

$$b + m = Lt, \quad p = \frac{\lambda}{L}, \quad \beta = \frac{L}{\theta}$$

とおく. このとき,

$$\beta + m = \frac{L}{\theta} + Lt - b = L\left(\frac{1}{\theta} + t\right) - b = L\frac{1+\theta t}{\theta} - b$$

直接的に極限を取ると,

$$\begin{aligned}
&\binom{k+m-1}{m}\frac{B(\alpha+k,\beta+m)}{B(\alpha,\beta)}\,dm \\
&= \frac{1}{(k+m)B(k,m+1)}\frac{B(\alpha+k,\beta+m)}{B(\alpha,\beta)}\,dm \\
&= \frac{1}{(Lt-b+k)B(k,Lt-b+1)}\frac{B(\alpha+k,L/\theta+Lt-b)}{B(\alpha,L/\theta)}\,dm \\
&= \frac{Lt}{Lt-b+k}\frac{(Lt)^{k-1}}{(Lt)^k B(k,Lt-b+1)}\frac{(L(1+\theta t)/\theta)^{\alpha+k}B(\alpha+k,L(1+\theta t)/\theta-b)}{(L/\theta)^\alpha B(\alpha,L/\theta)}\frac{(L/\theta)^\alpha}{(L(1+\theta t)/\theta)^{\alpha+k}}L\,dt \\
&= \frac{Lt}{Lt-b+k}\frac{(L(1+\theta t)/\theta)^{\alpha+k}B(\alpha+k,L(1+\theta t)/\theta-b)}{(Lt)^k B(k,Lt-b+1)(L/\theta)^\alpha B(\alpha,L/\theta)}\frac{(\theta t)^k}{(1+\theta t)^{k+\alpha}}\frac{dt}{t} \\
&\to \frac{\Gamma(\alpha+k)}{\Gamma(k)\Gamma(\alpha)}\frac{(\theta t)^k}{(1+\theta t)^{k+\alpha}}\frac{dt}{t} = \frac{1}{B(k,\alpha)}\frac{(\theta t)^k}{(1+\theta t)^{k+\alpha}}\frac{dt}{t}.
\end{aligned}$$

これは, 負の二項分布に従う確率変数

$$M \sim \mathrm{BetaNegativeBinomial}(k,\alpha,L/\theta)$$

から定まる確率変数

$$T = \frac{b+M}{L}$$

が $L \to \infty$ の極限で

$$\theta T \sim \mathrm{BetaPrime}(k,\alpha), \quad \text{i.e.} \quad T \sim \frac{1}{\theta}\mathrm{BetaPrime}(k,\alpha)$$

を満たすことを意味する. すなわち, $L \to \infty$ のとき,

$$\frac{b + \mathrm{BetaNegativeBinomial}(k,\alpha,L/\theta)}{L} \to \frac{1}{\theta}\mathrm{BetaPrime}(k,\alpha).$$

負の二項分布とベータ分布の極限を取った後に積分すると,

$$\begin{aligned}
&\int_0^\infty \frac{\lambda^k}{\Gamma(k)}e^{-\lambda t}t^{k-1}\frac{1}{\theta^\alpha\Gamma(\alpha)}e^{-\lambda/\theta}\lambda^{\alpha-1}\,d\lambda \\
&= \frac{t^{k-1}}{\Gamma(k)\theta^\alpha\Gamma(\alpha)}\int_0^\infty e^{-\lambda(1+\theta t)/\theta}\lambda^{\alpha+k-1}\,d\lambda \\
&= \frac{t^{k-1}}{\Gamma(k)\theta^\alpha\Gamma(\alpha)}\frac{\theta^{\alpha+k}\Gamma(\alpha+k)}{(1+\theta t)^{\alpha+k}} = \frac{1}{B(k,\alpha)}\frac{(\theta t)^k}{(1+\theta t)^{k+\alpha}}\frac{dt}{t}.
\end{aligned}$$

当然だが, 上と同じ結果が得られる.

注意: $\mathrm{BetaNegativeBinomial}(k,\alpha,\beta)$ の期待値は $k\beta/(\alpha-1)$ なので, $\mathrm{BetaNegativeBinomial}(k,\alpha,L/\theta)/L$ の期待値は $k/(\theta(\alpha-1))$ になる. 一方, $\mathrm{BetaPrime}(k,\alpha)$ の期待値は $k/(\alpha-1)$ なので $\mathrm{BetaPrime}(k,\alpha)/\theta$ の期待値も $k/(\theta(\alpha-1))$ になる. ゆえに $b=0$ とすれば, 上の連続時間極限を期待値を収束先に一致させた状態での極限にできる.

注意: $\mathrm{BetaNegativeBinomial}(k,\alpha,\beta)$ の分散は $k(\alpha+k-1)\beta(\alpha+\beta-1)/((\alpha-1)^2(\alpha-2))$ なので, $\mathrm{BetaNegativeBinomial}(k,\alpha,L/\theta)/L$ の分散は

$$\frac{k(k+\alpha-1)}{\theta^2(\alpha-1)^2(\alpha-2)}\left(1+\frac{(\alpha-1)\theta}{L}\right)$$

になる. 一方, $\mathrm{BetaPrime}(k,\alpha)/\theta$ の分散は

$$\frac{k(k+\alpha-1)}{\theta^2(\alpha-1)^2(\alpha-2)}$$

である. ゆえに上の連続時間極限は分散が収束先の値に単調減少しながら収束する.

In [80]:

```
mean_betanegbin(k, α, β) = k*β/(α - 1)
var_betanegbin(k, α, β) = k*(α+k-1)*β*(α+β-1)/((α-1)^2*(α-2))
std_betanegbin(k, α, β) = √var_betanegbin(k, α, β)

k, α, θ, L = 3, 4, 2, 100
@show mean_betanegbin(k, α, L/θ)/L, mean(BetaPrime(k, α)/θ)
@show var_betanegbin(k, α, L/θ)/L^2, var(BetaPrime(k, α)/θ);
```

```
(mean_betanegbin(k, α, L / θ) / L, mean(BetaPrime(k, α) / θ)) = (0.5, 0.5)
(var_betanegbin(k, α, L / θ) / L ^ 2, var(BetaPrime(k, α) / θ)) = (0.265, 0.25)
```

In [81]:

```
function pdf_betanegbin(x, k, α, β)
    m = round(x)
    m < 0 && return 0.0
    exp(loggamma(k+m) - loggamma(m+1) - loggamma(k) +
        logbeta(α+k, β+m) - logbeta(α, β))
end

@show k, α, β = 3, 8, 12
@show mmax = 10^6
@show sum(pdf_betanegbin(m, k, α, β) for m in 0:mmax)
@show sum(m*pdf_betanegbin(m, k, α, β) for m in 0:mmax)
@show μ = mean_betanegbin(k, α, β)
@show sum((m - μ)^2*pdf_betanegbin(m, k, α, β) for m in 0:mmax)
@show var_betanegbin(k, α, β)
@show √sum((m - μ)^2*pdf_betanegbin(m, k, α, β) for m in 0:mmax)
@show σ = std_betanegbin(k, α, β)
plot(x -> pdf_betanegbin(x, k, α, β), -1, μ+4σ; label="")
title!("BetaNegBin(k=$k, α=$α, β=$β)")
```

```
(k, α, β) = (3, 8, 12) = (3, 8, 12)
mmax = 10 ^ 6 = 1000000
sum((pdf_betanegbin(m, k, α, β) for m = 0:mmax)) = 0.9999999999999972
sum((m * pdf_betanegbin(m, k, α, β) for m = 0:mmax)) = 5.142857142857044
μ = mean_betanegbin(k, α, β) = 5.142857142857143
sum(((m - μ) ^ 2 * pdf_betanegbin(m, k, α, β) for m = 0:mmax)) = 23.265306122447598
var_betanegbin(k, α, β) = 23.26530612244898
√(sum(((m - μ) ^ 2 * pdf_betanegbin(m, k, α, β) for m = 0:mmax))) = 4.823412290323895
σ = std_betanegbin(k, α, β) = 4.823412290324038
```

Out[81]:

In [82]:

```
function plot_bnb_bp(k, α, θ, L;
        bp = BetaPrime(k, α)/θ, μ_bp = mean(bp), σ_bp = std(bp),
        tlim = (max(-1/L, μ_bp - 4σ_bp), μ_bp + 4σ_bp))
    β = L/θ
    μ_bnb, σ_bnb = mean_betanegbin(k, α, β), std_betanegbin(k, α, β)
    @show μ_bnb/L, μ_bp
    @show σ_bnb/L, σ_bp
    plot(t -> pdf_betanegbin(L*t, k, α, β)*L, tlim...;
        label="BetaNegBin(k,α,L/θ)/L")
    plot!(bp, tlim...; label="BetaPrime(k,α)/θ")
    title!("k=$k, α=$α, θ=$θ, L=$L")
end
```

Out[82]: plot_bnb_bp (generic function with 1 method)

In [83]:
```
plot(plot_bnb_bp.(2, 3, 10, (10, 30, 100, 300); tlim=(-0.1, 0.7))...;
    size=(800, 500), layout=(2, 2))
```

```
(μ_bnb / L, μ_bp) = (0.1, 0.1)
(σ_bnb / L, σ_bp) = (0.2449489742783178, 0.14142135623730953)
(μ_bnb / L, μ_bp) = (0.1, 0.1)
(σ_bnb / L, σ_bp) = (0.18257418583505539, 0.14142135623730953)
(μ_bnb / L, μ_bp) = (0.1, 0.1)
(σ_bnb / L, σ_bp) = (0.15491933384829668, 0.14142135623730953)
(μ_bnb / L, μ_bp) = (0.1, 0.1)
(σ_bnb / L, σ_bp) = (0.1460593486680443, 0.14142135623730953)
```

Out[83]:

8.7 ベータ負の二項分布の解釈 (Pólyaの壺)

$k = 1, 2, 3, \ldots$ であると仮定する.

このとき, ベータ負の二項分布の確率質量函数は次のように書き直される:

$$P(m|k, \alpha, \beta) = \binom{k+m-1}{m} \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)\cdot\beta(\beta+1)\cdots(\beta+m-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+k+m-1)}.$$

次のように当たりとはずれが出るくじを引き続ける状況を考える. 当たりが出る確率 p は2つの正の値のパラメータ達 (α, β) から $p = \alpha/(\alpha+\beta)$ で決まっている. そして, くじを引くごとにそれらのパラメータの値は次のように変化して行く:

(1) 当たりが出たら, パラメータの値を $(\alpha, \beta) \mapsto (\alpha+1, \beta)$ と更新する. (これによって次に当たりが出る確率は $(\alpha+1)/(\alpha+\beta+1)$ に上昇する.

(2) はずれが出たら, パラメータの値を $(\alpha, \beta) \mapsto (\alpha, \beta+1)$ と更新する. (これによって次にはずれが出る確率は $(\beta+1)/(\alpha+\beta+1)$ に上昇する.

上の式を見ると, ベータ負の二項分布はこのようなくじを引き続けて, 当たりが k 回出るまでに出たはずれの回数 m の分布になっていることがわかる. ちょうどこれの二項分布の場合の類似がベータ二項分布の解釈になることはすでに説明した.

注意: 上のくじの設定は次のようにも解釈される. 壺の中に当たりとはずれの玉がそれぞれ α 個, β 個入っており, 壺から無作為に当たりの玉が取り出されたら壺に当たりの玉を2個戻し, はずれの玉が取り出されたら壺にはずれの玉を2個戻す. これを繰り返す. この設定の壺を **Pólyaの壺** (Pólya's urn) と呼んだりする. この設定では α, β は整数でなければいけないが, 非整数の場合にもそれらが正の値であれば離散確率分布としてwell-definedになる.

8.8 Pólyaの壺の確率分布の定義

$\alpha, \beta > 0$ であると仮定する. 試行回数 n の **Pólyaの壺の確率分布** は次の確率質量函数によって定義される: $x_i = 1, 0$ でかつ $x_1 + \cdots + x_n = k$ のとき,

$$P(x_1, \ldots, x_n | n, \alpha, \beta) = \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)\cdot\beta(\beta+1)\cdots(\beta+n-k-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+n-1)}.$$

1 は当たりを 0 ははずれを意味し, $x_1 + \cdots + x_n = k$ は x_i 達の中に含まれる 1 の個数(当たりの個数)を意味する. 試行回数 n のPólyaの壺の確率分布を

$$\text{Pólya}(n, \alpha, \beta)$$

と書くことにする.

Pólyaの壺での確率が実際に上のように表されることは以下のように考えればわかる.

例えば, $n = 5$ のPólyaの壺の出目が $(1, 0, 1, 1, 0)$ になる確率は

$$\frac{\alpha}{\alpha+\beta}\frac{\beta}{\alpha+\beta+1}\frac{\alpha+1}{\alpha+\beta+2}\frac{\alpha+2}{\alpha+\beta+3}\frac{\beta+1}{\alpha+\beta+4}$$

になり, 出目が $(0, 1, 1, 0, 1)$ になる確率は

$$\frac{\beta}{\alpha+\beta}\frac{\alpha}{\alpha+\beta+1}\frac{\alpha+1}{\alpha+\beta+2}\frac{\beta+1}{\alpha+\beta+3}\frac{\alpha+2}{\alpha+\beta+4}$$

になる. これらは共通の分子と分母を持つので互いに等しい.

この例から一般に n 回の試行で壺から当たりが出た回数が同じであれば, それらが出た順序に確率がよらず, 試行回数 n のPólyaの壺の確率分布の確率質量函数の定義通りの式で表されることがわかる.

Pólyaの壺の試行の $i + j - 1$ 回目で i 回目の当たりが出たとき, 壺のパラメータ値は $(\alpha + i - 1, \beta + j - 1)$ の形をしており, 確率の計算では $(\alpha + i - 1)/(\alpha + \beta + i + j - 2)$ がかけられる. $i + j - 1$ 回目で j 回目のはずれが出たとき, 壺のパラメータ値は $(\alpha + i - 1, \beta + j - 1)$ の形をしており, 確率の計算では $(\beta + j - 1)/(\alpha + \beta + i + j - 2)$ がかけられる. したがって, $x_1, \ldots, x_n$ の中に当たりが k 個含まれているとき, n 回の試行で $(x_1, \ldots, x_n)$ が生じる確率は, 分子に $\alpha + i - 1$ $(i = 1, 2, \ldots, k)$ と $\beta + j - 1$ $(j = 1, 2, \ldots, n - k)$が並び, 分母に $\alpha + \beta + r - 1$ $(r = 1, 2, \ldots, n)$ を並べてできる商の値になる.

試行回数 n のPólyaの壺の確率分布の確率質量函数は以下のように表される:

$$\begin{aligned}
P(x_1, \ldots, x_n | n, \alpha, \beta) &= \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)\cdot\beta(\beta+1)\cdots(\beta+n-k-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+n-1)} \\
&= \frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\frac{\Gamma(\beta+n-k)}{\Gamma(\beta)}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)} \\
&= \frac{B(\alpha+k, \beta+n-k)}{B(\alpha, \beta)} \\
&= \int_0^1 p^k (1-p)^{n-k} \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}\, dp \\
&= \int_0^1 \prod_{i=1}^n \left(p^{x_i}(1-p)^{1-x_i}\right) \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}\, dp.
\end{aligned}$$

ここで $\prod_{i=1}^n \left(p^{x_i}(1-p)^{1-x_i}\right)$ は通常のBernoulli試行の確率質量函数なので, Pólyaの壺の確率分布に従う試行は成功確率 p がベータ分布 $\text{Beta}(\alpha, \beta)$ に従って確率的に揺らいでいるようなBernoulli試行になっていることがわかった.

注意: この結果は, Pólyaの壺の確率分布が共役事前分布 $p \sim \text{Beta}(\alpha, \beta)$ に関するBernoulli試行 $\text{Bernoulli}(p)^n$ のBayes版の統計モデルそのものになっていることを意味している.

In [ ]:
```
1
```

標本分布について

- 黒木玄

- 2022-04-11〜2022-05-21

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)

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.

このノートはかなり長大なものになったが, 内容的には資料として役に立てるための視覚化の部分が多く, 実質的な分量はそう多くはない. 他であまり見ないグラフも載せたつもりなので, 全体に目を通して頂ければとてもうれしい.

目次

- 1 標本分布とその使い方
  - 1.1 同時確率質量函数と同時確率密度函数
  - 1.2 確率変数の独立性の定義
    - 1.2.1 独立な確率変数達の同時確率質量函数
    - 1.2.2 独立な確率変数達の同時確率密度函数
    - 1.2.3 独立性に関する大雑把なまとめ
    - 1.2.4 分散を0に近付けたときの正規分布について
  - 1.3 独立同分布(i.i.d., iid)の定義
  - 1.4 標本分布の定義
  - 1.5 確率分布の積と n 乗の定義
  - 1.6 重要: 分布 D の標本分布の主な使用用途 (統計モデル, 統計量)
  - 1.7 試行回数 n のBernoulli試行の分布はBernoulli分布の標本分布
  - 1.8 二項分布による推定の確率的揺らぎの記述
  - 1.9 必修問題: 大阪都構想に関する住民投票の結果について
    - 1.9.1 WolframAlphaによる計算の例
    - 1.9.2 Julia言語による計算の例
    - 1.9.3 Clopper-Pearsonの信頼区間とそれを与えるP値
    - 1.9.4 信頼区間よりも情報量が大きなP値函数のプロット
    - 1.9.5 Sterneの信頼区間とそれを与えるP値函数
    - 1.9.6 Sterneの信頼区間を与えるP値函数の実装例
  - 1.10 対ごとに独立であっても全体が独立であるとは限らない
  - 1.11 確率変数の独立性の現実における解釈に関する重大な注意
- 2 確率変数達の共分散と相関係数と無相関性
  - 2.1 確率変数達の共分散と相関係数の定義
  - 2.2 確率変数達の対ごとの無相関性の定義
  - 2.3 問題: 確率変数の相関係数の計算例
  - 2.4 問題: Cauchy-Schwarzの不等式
  - 2.5 問題: 等確率有限離散分布の相関係数と $\cos\theta$ の関係
  - 2.6 問題: 相関係数の計算
  - 2.7 問題: 共分散が 0 に近くても相関係数が 0 から遠い場合がある
  - 2.8 問題: 独立ならば無相関である (実質1行で解ける)
  - 2.9 問題: 無相関でも独立とは限らない
  - 2.10 問題: 対ごとに無相関な確率変数達の和の分散はそれぞれの分散の和になる
  - 2.11 問題(メタアナリシスの出発点): 共通の期待値と異なる分散を持つ確率変数の荷重平均
  - 2.12 問題: 二項分布と負の二項分布の平均と分散のBernoulli分布と幾何分布の場合への帰着
  - 2.13 問題: 番号が異なる確率変数達が対ごとに無相関なときの確率変数の和の共分散
  - 2.14 分散共分散行列とその半正定値性
- 3 標本(サンプル, データ)の平均と分散と共分散と相関係数
  - 3.1 標本平均の定義

- 3.2 問題: 対ごとに無相関な確率変数達の標本平均の分散
- 3.3 標本分散と不偏分散の定義
- 3.4 不偏推定量について: 不偏分散の定義ではどうして n ではなく $n-1$ で割るか
- 3.5 データの共分散の定義
- 3.6 問題: 標本平均達の共分散
- 3.7 問題: 不偏共分散の定義ではどうして n ではなく $n-1$ で割るか
- 3.8 データの相関係数の定義 (以上の定義のまとめにもなっている)
- 4 最小二乗法による線形回帰
  - 4.1 問題: 最小二乗法による線形回帰の公式の導出
    - 4.1.1 解答例1: 単回帰の公式の素朴な導出
    - 4.1.2 解答例2: データの5つの要約値だけで平均二乗残差を表示
    - 4.1.3 コンピュータによる最小二乗法の計算例
  - 4.2 必修問題: 最小二乗法の計算例 (Anscombe's quartet)
    - 4.2.1 WaolframAlphaでAnscombeの例1を扱う方法
    - 4.2.2 Julia言語でAnscombeの例を扱う方法
  - 4.3 Galton (1886)が描いた2つの回帰直線
    - 4.3.1 Galtonのグラフの再現
    - 4.3.2 グラフ中の楕円と2変量正規分布モデルや回帰直線の関係
    - 4.3.3 楕円と回帰直線の関係の視覚化
- 5 モーメントとキュムラントと歪度と尖度
  - 5.1 モーメントとその母函数と特性函数とキュムラント母函数の定義
  - 5.2 特性函数による期待値の表示
  - 5.3 問題: 分布のアフィン変換のキュムラント
  - 5.4 問題: 標準正規分布のモーメント母函数と特性函数とキュムラント母函数
  - 5.5 確率変数の標準化と標準化キュムラントと歪度と尖度
  - 5.6 問題: 正規分布の歪度と尖度
  - 5.7 問題: 一様分布のキュムラント母函数と歪度と尖度
  - 5.8 問題: 独立な確率変数達の和のモーメント母函数と特性函数とキュムラント母函数
  - 5.9 問題: ガンマ分布のキュムラント母函数と歪度と尖度
  - 5.10 問題: 二項分布のキュムラント母函数と歪度と尖度
- 6 独立同分布な確率変数達の不偏分散の分散
  - 6.1 期待値 0, 分散 1 の場合への帰着
  - 6.2 標本平均と不偏分散の共分散の計算
  - 6.3 不偏分散の分散の計算
  - 6.4 歪度と尖度に関する不等式
  - 6.5 問題: 不偏分散と不偏補正されていない標本分散の平均二乗誤差の比較
- 7 標本分布における標本平均と不偏分散の同時分布の視覚化
  - 7.1 正規分布の標本分布における $\bar{X}, S^2$ の同時分布
  - 7.2 一様分布の標本分布における$\bar{X}, S^2$ の同時分布
  - 7.3 単峰型のガンマ分布の標本分布における $\bar{X}, S^2$ の同時分布
  - 7.4 指数分布の標本分布における $\bar{X}, S^2$ の同時分布
  - 7.5 非対称なベータ分布の標本分布における $\bar{X}, S^2$ の同時分布
  - 7.6 非対称な2つ山の混合正規分布の標本分布における $\bar{X}, S^2$ の同時分布
  - 7.7 対称なBernoulli分布の標本分布における $\bar{X}, S^2$ の同時分布
  - 7.8 非対称なBernoulli分布の標本分布における $\bar{X}, S^2$ の同時分布
  - 7.9 Poisson分布の標本分布における $\bar{X}, S^2$ の同時分布
  - 7.10 対数正規分布の標本分布における $\bar{X}, S^2$ の同時分布
- 8 正規分布の標本分布の場合
  - 8.1 準備: $\sqrt{n}\,\bar{x}$ を含む正規直交座標系
  - 8.2 準備 $n-2$ 次元単位球面上の微小面積要素
  - 8.3 正規分布の標本分布における $\bar{X}, S^2$ の同時分布の計算
  - 8.4 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
  - 8.5 正規分布モデルの使用に関する重要な注意: t 分布の使いどころ
  - 8.6 必修問題: データの t 値の計算例
    - 8.6.1 WolframAlphaによる t 値の計算例
    - 8.6.2 Julia言語による t 値の計算例
    - 8.6.3 問題用の擬似データの生成の仕方
- 9 標本分布における T の分布の視覚化
  - 9.1 正規分布の標本分布における T の分布
  - 9.2 一様分布の標本分布における T の分布
  - 9.3 単峰型のガンマ分布の標本分布における T の分布
  - 9.4 指数分布の標本分布における R の分布
  - 9.5 非対称なベータ分布の標本分布における T の分布
  - 9.6 非対称な2つ山の混合正規分布の標本分布における T の分布
  - 9.7 対称なBernoulli分布の標本分布における T の分布

9.8 非対称なBernoulli分布の標本分布における T の分布
9.9 Poisson分布の標本分布における T の分布
9.10 対数正規分布の標本分布における T の分布

In [1]:
```
ENV["LINES"], ENV["COLUMNS"] = 100, 100
using BenchmarkTools
using Distributions
using LinearAlgebra
using Optim
using Printf
using QuadGK
using Random
Random.seed!(4649373)
using Roots
using SpecialFunctions
using StaticArrays
using StatsBase
using StatsFuns
using StatsPlots
default(fmt = :png, titlefontsize = 10, size = (400, 250))
using SymPy
```

In [2]:
```
# Override the Base.show definition of SymPy.jl:
# https://github.com/JuliaPy/SymPy.jl/blob/29c5bfd1d10ac53014fa7fef468bc8deccadc2fc/src/types.

@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::SymbolicObject)
    print(io, as_markdown("\\displaystyle " * sympy.latex(x, mode="plain", fold_short_frac=fal
end
@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::AbstractArray{Sym})
    function toeqnarray(x::Vector{Sym})
        a = join(["\\displaystyle " * sympy.latex(x[i]) for i in 1:length(x)], "\\\\")
        """\\left[ \\begin{array}{r}$a\\end{array} \\right]"""
    end
    function toeqnarray(x::AbstractArray{Sym,2})
        sz = size(x)
        a = join([join("\\displaystyle " .* map(sympy.latex, x[i,:]), "&") for i in 1:sz[1]],
        "\\left[ \\begin{array}{" * repeat("r",sz[2]) * "}" * a * "\\end{array}\\right]"
    end
    print(io, as_markdown(toeqnarray(x)))
end
```

1 標本分布とその使い方

1.1 同時確率質量函数と同時確率密度函数

確率変数達 $X_1, \dots, X_n$ が同時確率質量函数 $P(x_1, \dots, x_n)$ を持つとは, $P(x_1, \dots, x_n) \geq 0$ でかつ, $P(x_1, \dots, x_n)$ の総和が 1 であり,

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

が成立することである. このとき, X_i 単独の確率質量函数 $P(x_i)$ は

$$P(x_i) = \sum_{x_1} \cdots \widehat{\sum_{x_i}} \cdots \sum_{x_n} P(x_1, \cdots, x_i, \dots, x_n)$$

になる. ここで, $\widehat{\sum_{x_i}}$ は $\sum_{x_i}$ を除くという意味である. なぜならば,

$$\begin{aligned} E[f(X_i)] &= \sum_{x_1} \cdots \sum_{x_i} \cdots \sum_{x_n} f(x_i) P(x_1, \cdots, x_i, \dots, x_n) \\ &= \sum_{x_i} f(x_1) \left(\sum_{x_1} \cdots \widehat{\sum_{x_i}} \cdots \sum_{x_n} P(x_1, \cdots, x_i, \dots, x_n) \right). \end{aligned}$$

注意: $i \neq j$ のとき $P(x_i)$ と $P(x_j)$ は一般に異なる確率質量函数になることに注意せよ. 区別を明瞭にするためには $P(x_i)$ を $P_i(x_i)$ のように書くべきであるが, 統計学の世界での慣習に従ってこのような記号法上の省略を行った. これが気に入らない人は自分でノートを作るときに, 区別されるべきものに異なる記号を割り振るようにすればよいだろう. 以下についても同様である.

確率変数達 $X_1, \dots, X_n$ が同時確率密度函数 $p(x_1, \dots, x_n)$ を持つとは, $p(x_1, \dots, x_n) \geq 0$ でかつ, $p(x_1, \dots, x_n)$ の積分が 1 になり,

$$E[f(X_1,\dots,X_n)] = \int\cdots\int f(x_1,\dots,x_n)p(x_1,\dots,x_n)\,dx_1\cdots dx_n$$

が成立することである. ここで $\int$ は定積分を表す. このとき, X_i 単独の確率密度函数 $p(x_i)$ は

$$p(x_i) = \int\cdots\widehat{\int}\cdots\int p(x_1,\cdots,x_i,\dots,x_n)\,dx_1\cdots\widehat{dx_i}\cdots dx_n$$

になる. ここで, $\widehat{\int}$, $\widehat{dx_i}$ はそれらを除くことを意味する. なぜならば,

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

1.2 確率変数の独立性の定義

確率変数達 $X_1,\dots,X_n$ が与えられており, それらの函数の期待値 $E[f(X_1,\dots,X_n)]$ が定義されているとする. このとき, $X_1,\dots,X_n$ が **独立** (independent)であるとは,

$$E[f_1(X_1)\cdots f_n(X_n)] = E[f_1(X_1)]\cdots E[f_n(X_n)] \tag{1}$$

のように X_i ごと函数達 $f_i(X_i)$ の積の期待値が各々の $f_i(X_i)$ の期待値の積になることだと定める.

注意: 厳密には函数 f_i 達を動かす範囲を確定させる必要があるが, このノートではそのようなことにこだわらずに解説することにする. 厳密な理論の展開のためには測度論的確率論の知識が必要になるが, そのような方向はこのノートの目標とは異なる. 測度論的確率論に興味がある人は別の文献を参照して欲しい. ただし, 測度論的確率論の理解と統計学の理解は別の話題になってしまうことには注意して欲しい. 測度論的確率論と統計学では興味の方向が異なる.

1.2.1 独立な確率変数達の同時確率質量函数

確率変数達 $X_1,\dots,X_n$ が同時確率質量函数 $P(x_1,\dots,x_n)$ を持つとき, X_i 単独の確率質量函数を $P_i(x_i)$ と書くならば, 確率変数達 $X_1,\dots,X_n$ が独立であることと,

$$P(x_1,\dots,x_n) = P_1(x_1)\cdots P_n(x_n) \tag{2}$$

が成立することは同値である. すなわち, $(X_1,\dots,X_n)$ の函数の期待値が次のように表されることと同値である:

$$E[f(X_1,\dots,X_n)] = \sum_{x_1}\cdots\sum_{x_n} f(x_1,\dots,x_n)P_1(x_1)\cdots P_n(x_n)$$

注意: ここでは気分を変えて, $P_i(x_i)$ をずぼらに $P(x_i)$ と書く流儀をやめてみた.

証明: (2)が成立しているならば,

$$\begin{aligned}E[f_1(X_1)\cdots f_n(X_n)] &= \sum_{x_1}\cdots\sum_{x_n} f_1(x_1)\cdots f_n(x_n)P_1(x_1)\cdots P_n(x_n)\\ &= \sum_{x_1} f_1(x_1)P_1(x_1)\cdots\sum_{x_n} f_n(x_n)P_n(x_1)\\ &= E[f_1(X_1)]\cdots E[f_n(X_n)]\end{aligned}$$

と(1)が成立する.

逆に(1)が成立しているならば, $f_i(x_i)$ として $x_i = a_i$ のときにのみ 1 でそれ以外のときに 0 になる函数を取ると,

$$P(a_1,\dots,a_n) = E[f_1(X_1)\cdots f_n(X_n)] = E[f_1(X_1)]\cdots E[f_n(X_n)] = P_1(a_1)\cdots P_n(a_n)$$

なので, (2)が成立する.

1.2.2 独立な確率変数達の同時確率密度函数

確率変数達 $X_1,\dots,X_n$ が同時確率密度函数 $p(x_1,\dots,x_n)$ を持つとき, X_i 単独の確率密度函数を $p_i(x_i)$ と書くならば, 確率変数達 $X_1,\dots,X_n$ が独立であることと,

$$p(x_1,\dots,x_n) = p_1(x_1)\cdots p_n(x_n) \tag{3}$$

が成立することは同値である. すなわち, $(X_1,\dots,X_n)$ の函数の期待値が次のように表されることと同値である:

$$E[f(X_1, \dots, X_n)] = \int \cdots \int f(x_1, \dots, x_n) p_1(x_1) \cdots p_n(x_n)\, dx_1 \cdots dx_n.$$

ここで $\int \cdots dx_i$ は適切な範囲での定積分を表す.

注意: ここでは気分を変えて, $p_i(x_i)$ をずぼらに $p(x_i)$ と書く流儀をやめてみた.

(雑な)証明: (3)が成立しているならば,

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

と(1)が成立する.

簡単のため, 密度函数達 $p(x_1, \dots, p_n)$, $p_i(x_i)$ は連続函数になっていると仮定する(応用上は概ねこれで十分). このとき,

$$f_i(x_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i - a_i)^2}{2\sigma_i^2}\right)$$

とおくと, $\sigma_i \searrow 0$ のとき

$$\begin{aligned} E[f_i(X_i)] &= \int f_i(x_i) p_i(x_i)\, dx_i \to p_i(a_i), \\ E[f_1(X_1) \cdots f_n(X_n)] &= \int \cdots \int f_1(x_1) \cdots f_n(x_n) p(x_1, \dots, x_n)\, dx_1 \cdots dx_n \to p(a_1, \dots, a_n) \end{aligned}$$

となるので, 逆に(1)が成立しているならば

$$p(a_1, \dots, a_n) = p_1(a_1) \cdots p_n(a_n)$$

と(3)が成立する.

確率密度函数と確率質量函数が混ざっている場合も同様である.

1.2.3 独立性に関する大雑把なまとめ

大雑把に誤解を恐れずに言うと, 確率が積になるときに独立だという.

実際の計算では, 積の期待値が期待値の積になるという形式で確率変数の独立性を使う.

In [3]:

```
# 以下で使う図の準備

var"正規分布は分散が小さくなると一点に集中して行く." = plot()
for σ in (1.0, 0.5, 0.25, 0.1)
    plot!(Normal(0, σ); label="σ = $σ")
end
title!("Normal(μ=0, σ)");
```

1.2.4 分散を0に近付けたときの正規分布について

X_σ を平均 μ, 分散 σ^2 の正規分布に従う確率変数とすると, 有界な連続函数 $f(x)$ について

$$\lim_{\sigma \searrow 0} E[f(X_\sigma)] = f(\mu)$$

となる. 証明が知りたい人は次のリンク先の「総和核」に関する解説(特にGauss核の説明)を参照せよ:

- [12 Fourier解析 - 総和核](https://nbviewer.org/github/genkuroki/Calculus/blob/master/12%20Fourier%20analysis.ipynb#%E7%B7%8F%E5%92%8C%E6%

大雑把に言えば, これは $\sigma \searrow 0$ の極限で正規分布が一点 $x = \mu$ に集中することを意味している.

In [4]:
```
var"正規分布は分散が小さくなると一点に集中して行く."
```

Out[4]:

$E[f(X_0)] = f(\mu)$ を満たす確率変数 X_0 は実質定数 μ に等しく, 実質定数に等しい確率変数が従う分布を **Dirac分布** と呼ぶことがある.

1.3 独立同分布(i.i.d., iid)の定義

確率変数達 $X_1, \ldots, X_n$ が **独立同分布** (independent and identically distributed, **i.i.d.**, **iid**) であるとは, それらが独立でかつ X_i 達が従う分布が互いにすべて等しいことであると定める.

1.4 標本分布の定義

n 個の確率変数達 $X_1, \ldots, X_n$ は独立同分布であり, 各 X_i は共通の確率分布 D に従うと仮定する:

$$X_1, \ldots, X_n \sim D \quad \text{(independent)}$$

このとき, $X_1, \ldots, X_n$ を分布 D の **標本** または **サンプル** (sample)と呼び, $X_1, \ldots, X_n$ の同時確率分布を **分布 D のサイズ n の標本分布** (distribution of samples of size n from distribution D)と呼ぶことにする.

注意: 独立同分布な確率変数達 $X_1, \ldots, X_n$ (確率変数は数ではなく函数であった)そのものではなく, それらの実現値 $x_1, \ldots, x_n$ (函数としての確率変数の値達)を標本もしくはサンプルと呼ぶことがある. 1つひとつの値 x_i をサンプルと呼ぶのではなく, 数の列 $x_1, \ldots, x_n$ をサンプルと呼ぶことに注意せよ. その辺が紛らわしい場合には数の列 $x_1, \ldots, x_n$ を **データ** (data)と呼ぶことがある. $X_1, \ldots, X_n$ もデータと呼ぶことがある. この辺の用語の使い方はかなりイーカゲンになり易いので注意して欲しい.

確率変数 X_i 達が同一の確率密度函数 $p(x_i)$ を持つとき, $X_1, \ldots, X_n$ の同時確率密度函数が

$$p(x_1, \ldots, x_n) = p(x_1) \cdots p(x_n)$$

であることと, $X_1, \ldots, X_n$ の同時確率分布が確率密度函数 $p(x)$ が定める連続分布の標本分布であることは同値である.

確率変数 X_i 達が同一の確率質量函数 $P(x_i)$ を持つとき, $X_1, \ldots, X_n$ の同時確率質量函数が

$$P(x_1, \ldots, x_n) = P(x_1) \cdots P(x_n)$$

であることと, $X_1, \ldots, X_n$ の同時確率分布が確率質量函数 $P(x)$ が定める離散分布の標本分布であることは同値である.

1.5 確率分布の積と n 乗の定義

$X_1, \ldots, X_n$ が独立な確率変数達であり, X_i が従う分布が D_i であるとき, $(X_1, \ldots, X_n)$ が従う分布を

$$D_1 \times \cdots \times D_n$$

と書き, 分布 $D_1, \ldots, D_n$ の **積** (product, 直積, direct product)と呼ぶ:

$$(X_1, \ldots, X_n) \sim D_1 \times \cdots \times D_n \iff X_i \sim D_i \text{ for } i = 1, \ldots, n \text{ and } X_1, \ldots, X_n \text{ are independent.}$$

例えば, 分布 D_i が確率密度函数 $p_i(x_i)$ を持つとき, 積分布 $D_1 \times \cdots \times D_n$ は確率密度函数

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

によって定義される多変量連続分布になる. 確率密度函数が積の形になるので, 分布の積と呼ぶ.

D_i がすべて同一の分布 D であるとき, 積 $D_1 \times \cdots \times D_n$ を D^n と書き, 分布 D の **べき乗** (冪乗, power)もしくは **累乗** と呼ぶ. 前節で説明したように分布 D の n 乗 D^n は分布 D のサイズ n の標本分布と呼ばれる:

$$(X_1, \dots, X_n) \sim D^n \iff (X_1, \dots, X_n) \text{ is a sample of size } n \text{ from the distribution D.}$$

例: 試行回数 n, 成功確率 p のBernoulli試行の分布は成功確率 p のBernoulli分布の n 乗になる. 以下の節も参照せよ.

1.6 重要: 分布 D の標本分布の主な使用用途 (統計モデル, 統計量)

確率分布 D のサイズ n の標本分布は, 現実に得られる長さ n の数値列 $x_1, \dots, x_n$ の形式のデータの生成法則のモデル化としてよく使われている.

そのとき, 確率分布 D やその標本分布 D^n は **統計モデル** と呼ばれる.

ただし, 分布 D のサイズ n の標本分布 D^n は各々が分布 D に従う独立同分布な n 個の確率変数達 $X_1, \dots, X_n$ の同時確率分布のことだと定義されていたので, データ $x_1, \dots, x_n$ の生成法則が独立同分布という強い条件を近似的に満たしていなければ, そのようなモデル化は妥当ではなくなる. (独立同分布の仮定を緩めて統計モデルを作ることもよくある. しかし, 独立同分布の易しい場合についての理解抜きにその仮定を外した場合について理解することは無理である.)

使用している統計モデルの妥当性の判断をシンプルな一般論に従って完遂することは不可能である. 応用先の専門知識を徹底的に用いて, 個々の目的に合わせて, 統計モデルの適切さをそのユーザー側が独自に見積もる必要がある.

統計モデルとしての標本分布 D^n とそれに従う独立同分布な確率変数達 $X_1, \dots, X_n$ が与えられたとき, $X_1, \dots, X_n$ の函数 $f(X_1, \dots, X_n)$ (これも確率変数になる)を **統計量** (statistic)と呼ぶことがある.

例えば, 標本の平均

$$\bar{X} = \mathrm{mean}(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^{n} X_i$$

は統計量の典型的な例になっている. これ以外にも, 標本の分散, 標本の中央値, 標本の四分位数, 標本の最大値や最小値などはすべて以上の意味での統計量とみなされる.

統計モデルに従う確率変数達 $X_1, \dots, X_n$ は現実で得られるデータ $x_1, \dots, x_n$ の生成法則のモデル化になっているので, 統計量 $f(X_1, \dots, X_n)$ には現実から得たデータの数値 $x_1, \dots, x_n$ を代入して, 数値 $f(x_1, \dots, x_n)$ を計算できる.

例えば, 上の標本平均の定義式中の X_i に現実から得た数値 x_i を代入すれば, データの数値の標本平均

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

が得られる.

現実世界と統計モデルの比較は様々な統計モデル内での確率変数としての統計量 $f(X_1, \dots, X_n)$ 達と現実世界で得たデータから得られるそれに対応する数値 $f(x_1, \dots, x_n)$ 達を比較することによって行われる.

例えば, 統計モデル内の確率変数としての標本平均 $\bar{X}$ の振る舞い方と, 現実のデータの標本平均 $\bar{x}$ (これは数値になる)を比較することによって, 平均に関する統計分析が行われる.

統計モデルを現実から得たデータ(もしくはその生成法則)に適合させるためにモデルのパラメータを調節することを **パラメータの推定** (estimating parameters)という.

例えばモデルの平均パラメータ μ はデータの標本平均で推定することが多い.

標本平均は統計量(統計モデル内での確率変数)でもあった. モデルのパラメータ θ の推定に使われる統計量をよく $\hat{\theta}$ のように書いて, パラメータ θ の **推定量** (estimator)と呼ぶ. 平均を意味するパラメータ μ を標本平均で推定する場合には $\hat{\mu} = \bar{X}$ と書いたりする.

例: データ $x_1, \dots, x_n$ がコンピュータで rand() 函数を n 回繰り返して得られた乱数列とするとき, そのデータの生成法則は一様分布 $\mathrm{Uniform}(0, 1)$ のサイズ n の標本分布に近似的に従っているとみなされる. **例終**

例(基本!): S市の中学3年生男子の総人数は N 人であるとし, その身長の値全体を $a_1, a_2, \dots, a_N$ と書く. 等確率 $1/N$ で a_i 達のどれかの値が選ばれるという条件で定義される有限離散分布を Q と書き, S市の中3男子の身長の分布と呼ぶことにする. (全数調査前であるという理由で現時点で分布 Q は未知であると仮定する.)

$1, 2, \dots, N$ から n 個の数 $i_1, \dots, i_n$ を無作為に選んで, それに対応する N 人中 n 人分の中3男子の身長を測定してデータ $(x_1, \dots, x_N) = (a_{i_1}, \dots, a_{i_N})$ が得られたとする. このとき, データ $x_1, \dots, x_N$ の生成法則はS市の中3男子の身長の分布 Q のサイズ n の標本分布になっていると考えられる. (確率的揺らぎは無作為抽出の段階で生じている.)

未知の分布 Q についての推定・推測・推論をデータ $x_1, \dots, x_N$ を用いて行いたい.

そのときに, 未知の分布 Q のモデル化として平均 μ と分散 σ^2 をパラメータに持つ正規分布を設定し, データ $x_1, \dots, x_N$ の生成法則のモデル化として, 正規分布のサイズ n の標本分布 $\mathrm{Normal}(\mu, \sigma)^n$ を採用することが考えらえる. この場合には正規分布モデルを統計モデルとして採用したことになる.

モデルのパラメータ μ, σ はデータ $x_1, \dots, x_N$ を使って推定されることになる.

未知であるS市の中3男子の身長の分布 Q (よく **真の分布** と呼ばれる)が正規分布からかけ離れた形をしていると, 正規分布モデルの採用は妥当ではなくなる. 真の分布 Q が十分に正規分布に近いことが十分に確からしい場合には正規分布モデルの採用は適切になる.

統計モデル内で確率変数としての標本平均 $\bar{X}$ が従う分布とデータから得た数値としての標本平均 $\bar{x}$ の関係を調べて, 平均を意味するモデルのパラメータ μ について何が言えるかを考えることが平均に関する統計分析の出発点になる.

例終

例: 当たりが常に同じ確率で出続けると考えられるルーレットを n 回まわして, 当たりなら 1 を記録し, 外れなら 0 を記録することによって得られる $1, 0$ で構成された長さ n の数列 $x_1, \dots, x_n$ の生成法則は, 未知の成功確率 p を持つ試行回数 n のBernoulli試行の確率分布 $\mathrm{Bernoulli}(p)^n$ (これはBernoulli分布の標本分布に等しい)に近似的に従うと考えられる.
$(X_1, \dots, X_n) \sim \mathrm{Bernoulli}(p)^n$ のとき, 確率変数 $K = X_1 + \cdots + X_n$ は「n 回のBernoulli試行で 1 が出た回数」を意味する統計量になっている. 統計量 K はBernoulli試行モデル内部では二項分布に従う. 現実世界で「n 回中 k 回成功した」の形式のデータが得られたならば, 二項分布に従う統計量 K と現実世界で得たデータの数値 k を比較することができる. 以下の節ではそのための準備を行う. **例終**

1.7 試行回数 n のBernoulli試行の分布はBernoulli分布の標本分布

試行回数 n, 成功確率 p のBernoulli試行の確率質量函数は

$$P(x_1, \dots, x_n) = p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)} = \prod_{i=1}^{n}(p^{x_i}(1-p)^{1-x_i}) \quad (x_i = 1, 0)$$

であった. これは成功確率 p のBernoulli分布の確率質量函数

$$P(x_i) = p^{x_i}(1-p)^{1-x_i} \quad (x_i = 1, 0)$$

の積になっているので, 試行回数 n, 成功確率 p のBernoulli試行の確率分布は, 成功確率 p のBernoulli分布のサイズ n の標本分布になっている.

未知の確率 p で当たりが出るルーレットを n 回まわしたときの, 長さ n の当たりと外れからなる列をそのルーレットの出目のサイズ n の **標本** (**サンプル**, sample)と呼ぶ.

その意味でのルーレットの出目のサンプルの確率的揺らぎは, Bernoulli分布の標本分布(すなわちBernoulli試行の分布)でモデル化される.

例えば, 未知の確率 p で当たりが出るルーレットを $n = 1000$ 回まわしてサンプル(データ)を取得したら, 1000 回中当たりが 308 回で外れが 692 回ならば, そのルーレットで当たりが出る確率は3割程度だろうと推定できる. 実際にはルーレットを 1000 回まわし直すたびに当たりの回数は確率的に揺らぐので, 推定結果も確率的に揺らぐことになる. そのような揺らぎを数学的にモデル化するために標本分布は使用される.

言葉使いに関する重要な注意: 「標本」「サンプル」は一般に複数の数値の集まりになる. 上のルーレットの場合には当たりには 1 を対応させ, 外れを 0 に対応させると, サンプルは 1 と 0 からなる長さ n の列になる. 一つひとつの数値をサンプルと呼ぶのではなく, 複数の数値の集まりをサンプルと呼ぶ. この専門用語的言葉遣いは日常用語的なサンプルという言葉の使い方からはずれているので注意が必要である. この辺の言葉遣いで誤解を防ぎたい場合には「データ」と呼ぶこともある.

モデルと現実の区別: 「標本」「サンプル」「データ」という用語は統計モデル内での確率変数列についても使われることがある. 現実から得た数値列 $x_1, \dots, x_n$ と統計モデル内での確率変数列 $X_1, \dots, X_n$ は厳密に区別されなければいけない. しかし, 慣習的な言葉遣いはその区別を難しくするスタイルになっているので注意が必要である.

1.8 二項分布による推定の確率的揺らぎの記述

前節の設定を引き継ぐ.

Bernoulli分布のサイズ n の標本分布における 1 の個数の分布は二項分布になるのであった. 仮に

> n 回中 k 回の当たりが出たときに, 未知である当たりが出る確率は k/n に近いだろうと推定

することにしたときの, k/n の確率的な揺らぎは二項分布によって計算できる.

K は二項分布 $\mathrm{Binomial}(n, p)$ に従う確率変数であるとし, 確率変数 $\hat{p}$ を $\hat{p} = K/n$ と定める. この確率変数 $\hat{p} = K/n$ は上のルールで定めた未知の確率 p の推定の仕方の数学的記述になっている. このとき, $\hat{p}$ はパラメータ p の **推定量** (estimator)であるという. 確率変数 $\hat{p}$ は「n 回中 k 回当たりが出た」というデータに対応する確率変数 K の函数になっている.

$\hat{p} = K/n$ の期待値と分散は, 二項分布の期待値と分散がそれぞれ np と $np(1-p)$ であることより, 以下のように計算される:

$$E[\hat{p}] = E[K/n] = \frac{E[K]}{n} = \frac{np}{n} = p,$$
$$\mathrm{var}(\hat{p}) = \mathrm{var}(K/n) = \frac{\mathrm{var}(K)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

例えば, $n = 1000$, $p = 0.3$ のとき, $\hat{p}$ の標準偏差は

$$\mathrm{std}(\hat{p}) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.3 \cdot 0.7}{1000}} \approx 1.45\%.$$

モデル内の設定では大雑把に推定結果はこの2倍の $\pm 3\%$ 程度揺らぐと考えらえる.

未知の p を使わずに $\mathrm{SE} = \mathrm{std}(\hat{p})$ の値を推定するためには, その式の中の p に p の推定量 $\hat{p}$ を代入して得られる公式

$$\widehat{SE} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

を使えばよいだろう. ($\mathrm{SE} = \mathrm{std}(\hat{p})$ は **標準誤差** (standard error)と呼ばれる. $\widehat{SE}$ は標準誤差の推定量である. $\widehat{SE}$ のことも標準誤差と呼ぶことがある. この辺の言葉遣いもイーカゲンな場合が多いので注意が必要である. 何をどう呼ぶかよりも, それが正確には何を意味しているかが重要である.

このような統計分析の結果が現実において信頼できるかどうかは, Bernoulli分布の標本分布によるモデル化の現実における妥当性に依存する. モデルが現実において妥当である証拠が全然無ければ, このような推測結果も信頼できないことになる. モデルの現実における妥当性の証拠の提示は統計モデルのユーザー側が独自に行う必要がある.

In [5]:

```
n, p = 1000, 0.3
dist = Binomial(n, p)
L = 10^6
K = rand(dist, L)
p̂ = K/n
histogram(p̂; norm=true, alpha=0.3, label="p̂ = K/n", bin=100)
vline!([p]; label="p=$p", xlabel="p̂", xtick=0.30-0.12:0.03:0.30+0.12)
title!("n = $n, p = $p")
```

Out[5]:

確かにランダムに決まる $\hat{p} = K/n$ の値は, 推定先の値である $p = 0.30 = 30\%$ を中心に大雑把に $\pm 3\%$ 程度の範囲に分布している. (グラフを見ればわかるように小さな確率で $\pm 3\%$ よりも大きく外れることもあることに注意せよ.)

現実の統計分析では, データを上のグラフの例のように100万回取り直したりできないので, このようなグラフを描くことはできない. たった1つの「n 回中 k 回成功」(もしくは「n 人の中 k 人」)の型のデータだけを使って判断を下さなければいけなくなる.

真っ暗闇の中を手探りで進むような感じになる.

1.9 必修問題: 大阪都構想に関する住民投票の結果について

2015年と2020年の大阪都構想に関する住民投票の結果は

- 2015年: 賛成: 694,844 (49.6%)　反対: 705,585 (50.4%)
- 2020年: 賛成: 675,829 (49.4%)　反対: 692,996 (50.6%)

であった(検索 (https://www.google.com/search?q=%E5%A4%A7%E9%98%AA%E9%83%BD%E6%A7%8B%E6%83%B3++%E4%BD%8F%E6%B0%91%E6%8A%95%E7%A5%A

どちらでも僅差で反対派が勝利した. パーセントの数値を見ると大変な僅差であったようにも見える. この数値に二項分布モデルを適用したらどうなるかを計算するのがこの問題の内容である. (注意・警告: ただし, 二項分布モデルの適用がこの場合に妥当であるかどうかは度外視する. 妥当であるか否かは目的によって異なる.)

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

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

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

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

- $P(K \geq 694844) = 2.5\%$ になるようなパラメータ p の値 p_L,
- $P(K \leq 694844) = 2.5\%$ になるようなパラメータ p の値 p_U.

(4) $n = 675829 + 692996 = 1368825$ のとき, 以下を求めよ:

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

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

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

$$1.234 \times 10^{20} = 0.\underbrace{00000000000000000001}_{20}234$$

のように書かずに,

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

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

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

$$\begin{aligned} P(K \geq k) &= 1 - \mathrm{cdf}(\mathrm{Binomial}(n, p), k-1) = \mathrm{cdf}(\mathrm{Beta}(k, n-k+1), p), \\ P(K \leq k) &= \mathrm{cdf}(\mathrm{Binomial}(n, p), k) = 1 - \mathrm{cdf}(\mathrm{Beta}(k+1, n-k), p). \end{aligned}$$

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

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

WolframAlphaでは Beta は `BetaDistribution` と書く.

注意: この問題の内容を一般化するだけで **検定** (統計的仮説検定, statistical hypothesis testing)や **信頼区間** (confidence interval)の一般論が得られる. (1), (2)はP値(モデルのデータへの適合度の指標の一種)を求める問題になっており, (3), (4)は 95% 信頼区間 $[p_L, p_U]$ を求める問題になっている(ただし両方Clopper-Pearsonの信頼区間の場合).

解答例:

(1) $P(K \leq 694844) \approx 1.130\mathrm{e}{-19}$

(2) $P(K \leq 675829) \approx 9.687\mathrm{e}{-49}$

(3) $[p_L, p_U] \approx [0.4953, 0.4970]$

(4) $[p_L, p_U] \approx [0.4929, 0.4946]$

解答終

1.9.1 WolframAlphaによる計算の例

以下の例は一部分が間違っている可能性がある. 鵜呑みせずに自分で正しい方法を採用しているかどうかを調べてから使って欲しい. (3), (4)については検算も重要なので確認しておくこと.

(1) 2 cdf(BinomialDistribution(694844 + 705585, 0.5), 694844) (https://www.wolframalpha.com/input?i=2+cdf%28BinomialDistribution%28694844+%2B+705585%2C+0.5%29%2C+694844%29)

(2) 2 cdf(BinomialDistribution(675829 + 692996, 0.5), 675829) (https://www.wolframalpha.com/input?i=2+cdf%28BinomialDistribution%28675829+%2B+692996%2C+0.5%29%2C+675829%29)

(3) p_L : quantile(BetaDistribution(694844, 705585+1), 0.025) (https://www.wolframalpha.com/input?i=quantile%28BetaDistribution%28694844%2C+705585%2B1%29%2C+0.025%29), もしくは InverseBetaRegularized(0.025, 694844, 705585 + 1) (https://www.wolframalpha.com/input?i=InverseBetaRegularized%280.025%2C+694844%2C+705585+%2B+1%29&lang=ja). 検算: 1 - cdf(BinomialDistribution(694844 + 705585, 0.49533667058), 694843) (https://www.wolframalpha.com/input?i=1+-+cdf%28BinomialDistribution%28694844+%2B+705585%2C+0.49533667058%29%2C+694843%29&lang=ja)

p_U : quantile(BetaDistribution(694844+1, 705585), 0.975) (https://www.wolframalpha.com/input?i=quantile%28BetaDistribution%28694844%2B1%2C+705585%29%2C+0.975%29), もしくは InverseBetaRegularized(0.975, 694844+1, 705585) (https://www.wolframalpha.com/input?i=InverseBetaRegularized%280.975%2C+694844%2B1%2C+705585%29). 検算: cdf(BinomialDistribution(694844 + 705585, 0.496993553), 694844) (https://www.wolframalpha.com/input?i=cdf%28BinomialDistribution%28694844+%2B+705585%2C+0.496993553%29%2C+694844%29&lang=ja)

(4) p_L : quantile(BetaDistribution(675829, 692996 + 1), 0.025) (https://www.wolframalpha.com/input?i=quantile%28BetaDistribution%28675829%2C+692996+%2B+1%29%2C+0.025%29), もしくは InverseBetaRegularized(0.025, 675829, 692996 + 1) (https://www.wolframalpha.com/input?i=InverseBetaRegularized%280.025%2C+675829%2C+692996+%2B+1%29). 検算: 1 - cdf(BinomialDistribution(675829 + 692996, 0.49289139359), 675828) (https://www.wolframalpha.com/input?i=1+-+cdf%28BinomialDistribution%28675829+%2B+692996%2C+0.49289139359%29%2C+675828%29)

p_U : quantile(BetaDistribution(675829+1, 692996), 0.975) (https://www.wolframalpha.com/input?i=quantile%28BetaDistribution%28675829%2B1%2C+692996%29%2C+0.975%29), もしくは InverseBetaRegularized(0.975, 675829+1, 692996) (https://www.wolframalpha.com/input?i=InverseBetaRegularized%280.975%2C+675829%2B1%2C+692996%29). 検算: cdf(BinomialDistribution(675829 + 692996, 0.494567220), 675829) (https://www.wolframalpha.com/input?i=cdf%28BinomialDistribution%28675829+%2B+692996%2C+0.494567220%29%2C+675829%29&lang=ja)

用語の復習:

- cdf = cumulative distribution function = 累積分布函数 $F(x) = P(X \leq x)$
- quantile function = 分位点函数 = 累積分布函数の逆函数 $Q(p) = F^{-1}(p)$

確率分布に関するこれらの函数はコンピュータ上でも実装されており, 便利に使える. 確率分布をどのように使っているかがよくわからないブラックボックス的な統計ソフトのアウトプットを眺めるのではなく, 確率分布をどのように使ったかが分かるような道具を使うようにすると, 理解を深め易くなる.

1.9.2 Julia言語による計算の例

In [6]:
```
# 確率関係の数値計算は対数値を使わないと多くの場合に失敗する.
# 次は二項分布における確率質量函数の対数である.
logP(n, p, k) = logabsbinomial(n, k)[1] + k*log(p) + (n-k)*log(1-p)
```

Out[6]:
```
logP (generic function with 1 method)
```

In [7]:
```
# (1)
@show 2exp(logsumexp(logP(694844 + 705585, 0.5, k) for k in 0:694844))
@show 2cdf(Binomial(694844 + 705585, 0.5), 694844);
```

```
2 * exp(logsumexp((logP(694844 + 705585, 0.5, k) for k = 0:694844))) = 1.130422573100686e-19
2 * cdf(Binomial(694844 + 705585, 0.5), 694844) = 1.130422573435063e-19
```

In [8]:
```
# (2)
@show 2exp(logsumexp(logP(675829 + 692996, 0.5, k) for k in 0:675829))
@show 2cdf(Binomial(675829 + 692996, 0.5), 675829);
```

```
2 * exp(logsumexp((logP(675829 + 692996, 0.5, k) for k = 0:675829))) = 9.687442124914083e-49
2 * cdf(Binomial(675829 + 692996, 0.5), 675829) = 9.687442131513948e-49
```

In [9]:
```
# (3)
f(t) = ccdf(Binomial(694844 + 705585, t), 694844-1) - 0.025
g(t) =  cdf(Binomial(694844 + 705585, t), 694844)   - 0.025
@show p_L = find_zero(f, (0, 1))
@show p_U = find_zero(g, (0, 1));
```

```
p_L = find_zero(f, (0, 1)) = 0.4953366705819138
p_U = find_zero(g, (0, 1)) = 0.4969935526649652
```

In [10]:
```
# (3)
@show n = 694844 + 705585
@show k = 694844
@show α = 0.05
@show p_L = quantile(Beta(k, n-k+1), α/2)
@show p_U = quantile(Beta(k+1, n-k), 1 - α/2);
```

```
n = 694844 + 705585 = 1400429
k = 694844 = 694844
α = 0.05 = 0.05
p_L = quantile(Beta(k, (n - k) + 1), α / 2) = 0.4953366705819138
p_U = quantile(Beta(k + 1, n - k), 1 - α / 2) = 0.49699355266496514
```

In [11]:
```
# (3)
@show n = 694844 + 705585
@show k = 694844
@show α = 0.05
@show p_L = beta_inc_inv(k, n-k+1, α/2)[1]
@show p_U = beta_inc_inv(k+1, n-k, 1 - α/2)[1];
```

```
n = 694844 + 705585 = 1400429
k = 694844 = 694844
α = 0.05 = 0.05
p_L = (beta_inc_inv(k, (n - k) + 1, α / 2))[1] = 0.4953366705819138
p_U = (beta_inc_inv(k + 1, n - k, 1 - α / 2))[1] = 0.49699355266496514
```

In [12]:
```
# (4)
f(t) = ccdf(Binomial(675829 + 692996, t), 675829-1) - 0.025
g(t) =  cdf(Binomial(675829 + 692996, t), 675829)   - 0.025
@show p_L = find_zero(f, (0, 1))
@show p_U = find_zero(g, (0, 1));
```

```
p_L = find_zero(f, (0, 1)) = 0.49289139358630474
p_U = find_zero(g, (0, 1)) = 0.4945672196912542
```

In [13]:
```
# (4)
@show n = 675829 + 692996
@show k = 675829
@show α = 0.05
@show p_L = quantile(Beta(k, n-k+1), α/2)
@show p_U = quantile(Beta(k+1, n-k), 1 - α/2);
```

```
n = 675829 + 692996 = 1368825
k = 675829 = 675829
α = 0.05 = 0.05
p_L = quantile(Beta(k, (n - k) + 1), α / 2) = 0.49289139358630474
p_U = quantile(Beta(k + 1, n - k), 1 - α / 2) = 0.4945672196912543
```

In [14]:
```
# (4)
@show n = 675829 + 692996
@show k = 675829
@show α = 0.05
@show p_L = beta_inc_inv(k, n-k+1, α/2)[1]
@show p_U = beta_inc_inv(k+1, n-k, 1 - α/2)[1];
```

```
n = 675829 + 692996 = 1368825
k = 675829 = 675829
α = 0.05 = 0.05
p_L = (beta_inc_inv(k, (n - k) + 1, α / 2))[1] = 0.49289139358630474
p_U = (beta_inc_inv(k + 1, n - k, 1 - α / 2))[1] = 0.4945672196912543
```

1.9.3 Clopper-Pearsonの信頼区間とそれを与えるP値

(3)と(4)で計算した値から得られる区間 $[p_L, p_U]$ は **母比率に関する信頼度95%のClopper-Pearsonの信頼区間** として統計学ユーザーのあいだでよく知られている.

(1)と(2)での2倍する前の確率は **片側検定のP値** になっている. 2倍後の値は **両側検定のP値** (通常はこちらを使う)の一種になっており, Clopper-Pearsonの信頼区間を与える.

P値は採用した統計モデルとデータの整合性の指標である(P値が小さければ整合性が低い). (1)と(2)で求めたP値の値は極めて小さいということは, 成功確率 $p = 0.5$ の二項分布モデルと2015年と2020年の大阪都構想に関する住民投票の結果の整合性が極めて低いということを意味している. 2015年と2020年の大阪都構想に関する住民投票の結果については, 成功確率 $p = 0.5$ の二項分布モデルは捨て去る必要があると考えられる.

(3)と(4)で求めた区間 $[p_L, p_U]$ はデータから計算されるP値が 5% 以上になるパラメータ p の値全体の集合になっている. すなわち, P値に関する 5% の閾値(**有意水準** と呼ばれる)で整合性が低すぎるという理由で捨て去られずにすむ p の値全体が信頼度 $100\% - 5\% = 95\%$ の信頼区間になっている. このパターンは一般の場合にもそのまま通用する.

注意・警告: この 5% の閾値の設定は非常に普及しているが, 5% の閾値の設定に科学的に合理的な根拠があるわけではない. そのような閾値に頼る推論は科学的に不合理な推論になるので注意を要する. 非常に残念なことに, 統計学をかじってしまったせいで, 自分自身は科学的だと信じならが, そのような非科学的な推論をしてしまう人達がたくさんいる.

Clopper-Pearsonの信頼区間の効率的計算には, 二項分布の累積分布函数はベータ分布の累積分布函数で書けることが使われる: $K \sim \mathrm{Binomial}(n, p)$ のとき,

$$\begin{aligned} P(K \geq k) &= 1 - \mathrm{cdf}(\mathrm{Binomial}(n, p), k-1) = \mathrm{cdf}(\mathrm{Beta}(k, n-k+1), p), \\ P(K \leq k) &= \mathrm{cdf}(\mathrm{Binomial}(n, p), k) = 1 - \mathrm{cdf}(\mathrm{Beta}(k+1, n-k), p). \end{aligned}$$

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

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

ベータ分布の累積分布函数は **正則化された不完全ベータ函数** (regularized incomplete Beta function)になっている:

$$P(T \leq \theta) = I_\theta(\alpha, \beta) = \frac{\int_0^\theta t^{\alpha-1}(1-t)^{\beta-1}\,dt}{B(\alpha, \beta)} \quad \text{if} \quad T \sim \mathrm{Beta}(\alpha, \beta).$$

このことから, `quantile` 函数の代わりに $\theta \mapsto p = I_\theta(\alpha, \beta)$ の逆函数を直接使ってもよい:

$$p_L = \mathrm{beta_inc_inv}(k, n-k+1, \alpha/2)[1], \quad p_U = \mathrm{beta_inc_inv}(k+1, n-k, 1-\alpha/2)[1].$$

Julia言語の `SpecialFunctions.jl` では正則化された不完全ベータ函数とその逆函数はそれぞれ `using SpecialFunctions` した後に `p = beta_inc(α, β, θ)[1]` と `θ = beta_inc_inv(α, β, p)[1]` で使える.

Wolfram言語では `BetaRegularized` (https://reference.wolfram.com/language/ref/BetaRegularized.html) と `InverseBetaRegularized` (https://reference.wolfram.com/language/ref/InverseBetaRegularized.html) を使う.

Clopper-Pearsonの信頼区間を使うことのメリットは, 二項分布の累積分布函数を和で計算してからパラメータに関する方程式を解くという面倒な手続きを経由せずに, 基本特殊函数の1つである正則化された不完全ベータ函数の逆函数に帰着して効率的に計算できることである. 別の信頼区間の定義の仕方との比較でデメリットもある. Sterneの信頼区間(Sterne (1954) (https://www.jstor.org/stable/2333026))との相対比較では, Clopper-Pearsonの信頼区間の方が無駄に広くなってしまう場合が多い. np と $n(1-p)$ の両方が十分に大きな場合にはどちらを使っても実践的に意味のある差は出ない.

In [15]:

```
n, k = 16, 7

P1 = plot(p -> 1 - cdf(Binomial(n, p), k-1), 0, 1;
    label="1 - cdf(Binomial(n, p), k-1)", legend=:bottomright)
plot!(p -> cdf(Beta(k, n-k+1), p), 0, 1;
    label="cdf(Beta(k, n-k+1), p)", ls=:dash)
title!("n = $n,  k = $k"; xlabel="p")

P2 = plot(p -> cdf(Binomial(n, p), k), 0, 1;
    label="cdf(Binomial(n, p), k)", legend=:topright)
plot!(p -> 1 - cdf(Beta(k+1, n-k), p), 0, 1;
    label="1 - cdf(Beta(k+1, n-k), p)", ls=:dash)
title!("n = $n,  k = $k"; xlabel="p")

plot(P1, P2; size = (800, 250), bottommargin=4Plots.mm)
```

Out[15]:

上のグラフでは $1 - \mathrm{cdf}(\mathrm{Binomial}(n, p), k-1)$ と $\mathrm{cdf}(\mathrm{Beta}(k, n-k+1), p)$ がぴったり一致し, $\mathrm{cdf}(\mathrm{Binomial}(n, p), k)$ と $1 - \mathrm{cdf}(\mathrm{Beta}(k+1, n-k), p)$ がぴったり一致し, 区別がつかなくなっている. ここで $\mathrm{cdf}(D, x)$ は分布 D の累積分布函数を表す:

$$X \sim D \implies \mathrm{cdf}(D, x) = P(X \leq x).$$

1.9.4 信頼区間よりも情報量が大きなP値函数のプロット

二項分布モデルにおいて n, k が与えられたときに, パラメータ p に対してP値を対応させる函数を **P値函数** (p-value function)と呼ぶ. P値函数の値が有意水準 α 以上の p 全体の集合が信頼度 $1-\alpha$ の信頼区間になる. この意味でP値函数はすべての信頼度に関する信頼区間の情報をすべて持っており, 適当な条件の下ではすべての信頼度に関する信頼区間が与えられていればそこからP値函数を逆に作れる. この意味でP値函数と信頼区間達は表裏一体の関係になっている.

以下では上の問題の場合についてのP値函数をプロットしてみよう.

In [16]:

```julia
function plot_pvalue_function!(pvalue_func, n, k; label="", kwargs...)
    p̂ = k/n
    SE = √(p̂*(1 - p̂)/n)
    ps = range(p̂ - 4SE, p̂ + 4SE, 1000)
    plot!(ps, p -> pvalue_func(n, k, p); label, kwargs...)
end

function plot_pvalue_function(pvalue_func, n, k; label="", kwargs...)
    plot()
    plot_pvalue_function!(pvalue_func, n, k; label)
    title!("n = $n, k = $k")
    plot!(; ytick=0:0.1:1, xlabel="parameter p", ylabel="p-value")
    plot!(; kwargs...)
end

function pvalue_clopper_pearson(dist::DiscreteUnivariateDistribution, x)
    min(1, 2cdf(dist, x), 2ccdf(dist, x-1))
end
pvalue_clopper_pearson(n, k, p) = pvalue_clopper_pearson(Binomial(n, p), k)

# (3)
P1 = plot_pvalue_function(pvalue_clopper_pearson, 694844 + 705585, 694844;
    label="Clopper-Pearson", xtick=0:0.0005:1)

# (4)
P2 = plot_pvalue_function(pvalue_clopper_pearson, 675829 + 692996, 675829;
    label="Clopper-Pearson", xtick=0:0.0005:1)

plot(P1, P2; size=(800, 250),
    leftmargin=4Plots.mm, bottommargin=4Plots.mm, rightmargin=4Plots.mm)
```

Out[16]:

データ x を与えられているとき(上の場合には「n 人中 k 人が賛成」というデータを与えられているとき), 統計モデルのパラメータ θ に(上の場合には p に)対してP値を対応させる函数をP値函数と呼ぶ. 上のグラフは大阪都構想に関する住民投票の結果データとし, 統計モデルを二項分布にしたときのP値函数のプロットである.

P値が高いパラメータ値ほど与えられたデータに統計モデルがよくフィット(fit, 適合)している. P値が低い場合にはその逆になる. あまりにも低いP値を持つパラメータ値については, 統計モデルとそのパラメータ値の組が与えられたデータに全然適合していないことになる.

そのとき重要なことは, P値が低いパラメータ値について, そのパラメータ値のみについてデータに適合していないと考えるのではなく, 統計モデルとそのパラメータ値の組がデータにフィットしていないと考えなければいけないことである.

例えば, 上のグラフを見れば, グラフに描いた部分の横軸の目盛りをはみ出しているパラメータ値 $p = 0.5$ のP値はおそろしく低い. 二項分布モデル内で $p = 0.5$ は賛成派の比率がちょうど 50% であることを意味している. $p = 0.5$ のP値がおそろしく低いことを理由に「賛成派の比率が 50% であるという仮説は否定された」のように単純に考えてはいけない. そうではなく, 「賛成派の比率が 50% であることを意味するパラメータ値の二項分布モデルの現実における妥当性は疑わしい」と考えなければいけない. 二項分布モデルも現実における妥当性を疑う対象に入れる必要がある.

この手のことが大学学部生向けの教科書に書かれていないせいで, 大学で統計学の講義を受講した人達の多くが統計学の最も基本的な部分をひどく誤解したまま単位を取得しているように思われる.

誤解を修正するためには次の講義動画がお勧めである:

- 京都大学大学院医学研究科 聴講コース 臨床研究者のための生物統計学「仮説検定とP値の誤解」佐藤 俊哉 医学研究科教授. [YouTube (https://youtu.be/vz9cZnB1d1c)]

信頼区間に関する解説は40:22以降にある.

その講義で扱われているP値に関するASA声明の翻訳とその翻訳の経緯の解説を以下の場所で読める:

- 統計的有意性とP値に関するASA声明 [pdf (https://www.biometrics.gr.jp/news/all/ASA.pdf)]
- 佐藤俊哉, ASA声明と疫学研究におけるP値, 計量生物学, 2018年38巻2号, pp. 109-115. [link (https://www.jstage.jst.go.jp/article/jjb/38/2/38_109/_article/-char/ja/)]

1.9.5 Sterneの信頼区間とそれを与えるP値函数

Sterneの信頼区間を与えるP値函数の定義は, 二項分布の確率質量函数を

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

と書くとき,

$$\mathrm{pvalue}_{\mathrm{Sterne}}(k|n,p) = \sum_{j \text{ with } P(j|n,p) \leq P(k|n,p)} P(j|n,p)$$

と $P(k|n,p)$ 以下となるような $P(j|n,p)$ 達の和として定義される. すなわち, 二項分布 $\mathrm{Binomial}(n,p)$ においてその値が生じる確率がデータの数値 k が生じる確率以下になる確率としてSterneの信頼区間を与えるP値は定義される.

そして, 与えられた n, k について, Sterneの信頼区間はこのP値が α 以上になるパラメータ p の範囲として定義される. (実はその定義だと区間になるとは限らない場合が稀にあるので, その場合には適当に定義を訂正することになる.)

1.9.6 Sterneの信頼区間を与えるP値函数の実装例

以下はSterneの信頼区間を与えるP値函数の実装例である.

Clopper-Pearsonの信頼区間を与えるP値函数の実装(実質1行!)と比較すると相当に複雑になっている.

そして, 実装の仕方によって計算効率に大きな違いが生じていることにも注目せよ.

In [17]:

```julia
x ≲ y = x < y || x ≈ y

# Naive implementation is terribly slow.
function pvalue_sterne_naive(dist::DiscreteUnivariateDistribution, x; xmax = 10^6)
    Px = pdf(dist, x)
    Px == 0 && return Px
    ymin, maxdist = minimum(dist), maximum(dist)
    ymax = maxdist == Inf ? xmax : maxdist
    sum(pdf(dist, y) for y in ymin:ymax if 0 < pdf(dist, y) ≲ Px; init = 0.0)
end
pvalue_sterne_naive(n, k, p) = pvalue_sterne_naive(Binomial(n, p), k)

# Second implementation is very slow.
function pvalue_sterne_old(dist::DiscreteUnivariateDistribution, x)
    Px = pdf(dist, x)
    Px == 0 && return Px
    distmin, distmax = extrema(dist)
    m = mode(dist)
    Px ≈ pdf(dist, m) && return one(Px)
    if x < m
        y = m + 1
        while !(pdf(dist, y) ≲ Px)
            y += 1
        end
        cdf(dist, x) + ccdf(dist, y-1)
    else # k > m
        y = m - 1
        while !(pdf(dist, y) ≲ Px)
            y -= 1
        end
        cdf(dist, y) + ccdf(dist, x-1)
    end
end
pvalue_sterne_old(n, k, p) = pvalue_sterne_old(Binomial(n, p), k)

### The following implementation is efficient.

_pdf_le(x, (dist, y)) =  pdf(dist, x) ≲ y

function _search_boundary(f, x0, Δx, param)
    x = x0
    if f(x, param)
        while f(x - Δx, param) x -= Δx end
    else
        x += Δx
        while !f(x, param) x += Δx end
    end
    x
end

function pvalue_sterne(dist::DiscreteUnivariateDistribution, x)
    Px = pdf(dist, x)
    Px == 0 && return Px
    m = mode(dist)
    Px ≈ pdf(dist, m) && return one(Px)
    if x < m
        y = _search_boundary(_pdf_le, 2m - x, 1, (dist, Px))
        cdf(dist, x) + ccdf(dist, y-1)
    else # x > m
        y = _search_boundary(_pdf_le, 2m - x, -1, (dist, Px))
        cdf(dist, y) + ccdf(dist, x-1)
    end
end
pvalue_sterne(n, k, p) = pvalue_sterne(Binomial(n, p), k)
```

Out[17]:

```
pvalue_sterne (generic function with 2 methods)
```

In [18]:
```
n = 10
k = -1:11
p = 0.4
a = @time pvalue_sterne_naive.(n, k, p)
b = @time pvalue_sterne_old.(n, k, p)
c = @time pvalue_sterne.(n, k, p)
d = @time pvalue_clopper_pearson.(n, k, p)
@show a ≈ b ≈ c
[a b c d]
```

```
  0.180538 seconds (441.16 k allocations: 24.337 MiB, 99.60% compilation time)
  0.159275 seconds (252.91 k allocations: 13.412 MiB, 99.44% compilation time)
  0.123667 seconds (199.45 k allocations: 10.607 MiB, 99.51% compilation time)
  0.103345 seconds (133.13 k allocations: 7.216 MiB, 31.14% gc time, 99.18% compilation time)
a ≈ b ≈ c = true
```

Out[18]:
```
13×4 Matrix{Float64}:
 0.0          0.0          0.0          0.0
 0.00772434   0.00772434   0.00772434   0.0120932
 0.058652     0.058652     0.058652     0.0927148
 0.333528     0.333528     0.333528     0.33458
 0.749177     0.749177     0.749177     0.764561
 1.0          1.0          1.0          1.0
 0.534186     0.534186     0.534186     0.733793
 0.212596     0.212596     0.212596     0.332477
 0.101119     0.101119     0.101119     0.109524
 0.0183412    0.0183412    0.0183412    0.0245891
 0.00167772   0.00167772   0.00167772   0.00335544
 0.000104858  0.000104858  0.000104858  0.000209715
 0.0          0.0          0.0          0.0
```

In [19]:
```
# (3)の場合に
# pvalue_sterne_naive は pvalue_sterne_old よりも数百倍遅く,
# pvalue_sterne_old は pvalue_sterne よりも数百倍遅く,
# pvalue_sterne は pvalue_clopper_pearson よりも少し遅い.
n = 694844 + 705585
k = 694844
a = @btime pvalue_sterne_naive($n, $k, 0.5)
b = @btime pvalue_sterne_old($n, $k, 0.5)
c = @btime pvalue_sterne($n, $k, 0.5)
d = @btime pvalue_clopper_pearson($n, $k, 0.5)
@show a ≈ b ≈ c ≈ d
a, b, c, d
```

```
  159.315 ms (0 allocations: 0 bytes)
  616.100 μs (8 allocations: 1.88 KiB)
  1.370 μs (8 allocations: 1.88 KiB)
  863.492 ns (8 allocations: 1.88 KiB)
a ≈ b ≈ c ≈ d = true
```

Out[19]:
```
(1.1304225731963725e-19, 1.130422573435063e-19, 1.130422573435063e-19, 1.130422573435063e-19)
```

In [20]:
```
# 極端な場合
n = 694844 + 705585
k = 600000
b = @btime pvalue_sterne_old($n, $k, 0.5)
c = @btime pvalue_sterne($n, $k, 0.5)
d = @btime pvalue_clopper_pearson($n, $k, 0.5)
b, c, d
```

```
  107.511 ns (0 allocations: 0 bytes)
  102.817 ns (0 allocations: 0 bytes)
  180.805 ns (0 allocations: 0 bytes)
```

Out[20]:
```
(0.0, 0.0, 0.0)
```

In [21]:
```
# この場合には pvalue_sterne_naive はさらに遅い.
n = 100000
k = 49500:50500
a = @time pvalue_sterne_naive.(n, k, 0.5)
b = @time pvalue_sterne_old.(n, k, 0.5)
c = @time pvalue_sterne.(n, k, 0.5)
d = @time pvalue_clopper_pearson.(n, k, 0.5)
@show a ≈ b ≈ c ≈ d;
```

```
 14.064621 seconds (3 allocations: 8.094 KiB)
  0.033857 seconds (8.00 k allocations: 1.839 MiB)
  0.002084 seconds (8.00 k allocations: 1.839 MiB)
  0.001591 seconds (8.01 k allocations: 1.841 MiB)
a ≈ b ≈ c ≈ d = true
```

In [22]:
```
# 以上の実装は超幾何分布でも使える.
dist = Hypergeometric(9, 9, 9)
k = -1:10
a = @time pvalue_sterne_naive.(dist, k)
b = @time pvalue_sterne_old.(dist, k)
c = @time pvalue_sterne.(dist, k)
d = @time pvalue_clopper_pearson.(dist, k)
@show a ≈ b ≈ c ≈ d
[a b c d]
```

```
  0.123616 seconds (360.92 k allocations: 19.790 MiB, 18.91% gc time, 99.43% compilation time)
  0.077116 seconds (166.49 k allocations: 8.988 MiB, 83.49% compilation time)
  0.069360 seconds (150.24 k allocations: 8.068 MiB, 99.22% compilation time)
  0.048367 seconds (171.80 k allocations: 8.864 MiB, 98.92% compilation time)
a ≈ b ≈ c ≈ d = true
```

Out[22]:
```
12×4 Matrix{Float64}:
 0.0          0.0          0.0          0.0
 4.11353e-5   4.11353e-5   4.11353e-5   4.11353e-5
 0.0033731    0.0033731    0.0033731    0.0033731
 0.0566845    0.0566845    0.0566845    0.0566845
 0.346935     0.346935     0.346935     0.346935
 1.0          1.0          1.0          1.0
 1.0          1.0          1.0          1.0
 0.346935     0.346935     0.346935     0.346935
 0.0566845    0.0566845    0.0566845    0.0566845
 0.0033731    0.0033731    0.0033731    0.0033731
 4.11353e-5   4.11353e-5   4.11353e-5   4.11353e-5
 0.0          0.0          0.0          0.0
```

In [23]:
```
# 以上の実装は無限離散分布のPoisson分布でも使える.
dist = Poisson(4)
k = -1:10
a = @time pvalue_sterne_naive.(dist, k)
b = @time pvalue_sterne_old.(dist, k)
c = @time pvalue_sterne.(dist, k)
d = @time pvalue_clopper_pearson.(dist, k)
@show a ≈ b ≈ c
[a b c d]
```

```
  0.713966 seconds (434.31 k allocations: 24.235 MiB, 2.61% gc time, 22.22% compilation time)
  0.070172 seconds (225.54 k allocations: 12.067 MiB, 99.15% compilation time)
  0.085095 seconds (152.06 k allocations: 8.159 MiB, 76.92% compilation time)
  0.057104 seconds (150.05 k allocations: 8.135 MiB, 98.77% compilation time)
a ≈ b ≈ c = true
```

Out[23]:
```
12×4 Matrix{Float64}:
 0.0          0.0          0.0          0.0
 0.0396791    0.0396791    0.0396791    0.0366313
 0.202252     0.202252     0.202252     0.183156
 0.452973     0.452973     0.452973     0.476207
 1.0          1.0          1.0          0.86694
 1.0          1.0          1.0          1.0
 0.609266     0.609266     0.609266     0.742326
 0.306448     0.306448     0.306448     0.429739
 0.12899      0.12899      0.12899      0.221348
 0.0694493    0.0694493    0.0694493    0.102267
 0.0213634    0.0213634    0.0213634    0.0427269
 0.00813224   0.00813224   0.00813224   0.0162645
```

In [24]:

```
# (3)
P1 = plot_pvalue_function(pvalue_sterne, 694844 + 705585, 694844;
    label="Sterne", xtick=0:0.0005:1)

# (4)
P2 = plot_pvalue_function(pvalue_sterne, 675829 + 692996, 675829;
    label="Sterne", xtick=0:0.0005:1)

plot(P1, P2; size=(800, 250), leftmargin=4Plots.mm, bottommargin=4Plots.mm)
```

Out[24]:

In [25]:

```
# Clopper-Pearsonの信頼区間を与えるP値函数とSterneの信頼区間を与えるP値函数の比較
# (3)の場合にはほぼぴったり一致している.
n, k = 694844 + 705585, 694844
plot(title="p-value functions for n = $n, k = $k", ytick=0:0.1:1)
plot_pvalue_function!(pvalue_sterne, n, k; label="Sterne")
plot_pvalue_function!(pvalue_clopper_pearson, n, k; label="Clopper-Pearson", ls=:dash)
```

Out[25]:

このように数値的にぴったり一致する場合にはClopper-Pearsonの信頼区間を与えるP値函数とSterneの信頼区間のどちらを使うか悩む必要はないだろう. どちらを使っても(ほぼ)同じ結果が得られる.

In [26]:

```
# Clopper-Pearsonの信頼区間を与えるP値函数とSterneの信頼区間を与えるP値函数の比較
# n = 50, k = 15 の場合には違いが見える.
n, k = 50, 15
plot(title="p-value functions for n = $n, k = $k", ytick=0:0.1:1)
plot_pvalue_function!(pvalue_sterne, n, k; label="Sterne")
plot_pvalue_function!(pvalue_clopper_pearson, n, k; label="Clopper-Pearson", ls=:dash)
```

Out[26]:

Sterneの信頼区間を与えるP値函数の値はClopper-Pearsonの信頼区間を与えるP値函数の値よりも小さくなりことが多い. (常にそうなるわけではない.) その結果, 対応する信頼区間もSterneの側が狭くなってくれることが多い.

1.10 対ごとに独立であっても全体が独立であるとは限らない

確率変数達 X, Y, Z の互いに異なる任意の2つが独立であっても, X, Y, Z の全体が独立でない場合があることを具体的な例を作ることによって証明しよう.

確率変数 X, Y, Z の同時確率質量函数が $P(x, y, z)$ であるとき, X 単独の確率質量函数は $P(x) = \sum_{y,z} P(x, y, z)$ になり, (X, Y) の同時確率質量函数は $P(x, y) = \sum_z P(x, y, z)$ になる. 他の場合も同様である.

注意: 変数名で異なる函数を区別するスタイルを採用したので記号法の運用時に注意すること. このスタイルでは $P(x, y)$ と $P(x, z)$ は異なる函数になる. 値を代入する場合には $P(x = 1, y = 1)$ や $P(x = 1, z = 1)$ のように書いて区別できるようにする.

$P(x, y, z)$ で $P(x, y) = P(x)P(y)$, $P(x, z) = P(x)P(z)$, $P(y, z) = P(y)P(z)$ を満たすが $P(x, y, z) \neq P(x)P(y)P(z)$ となるものを具体的に構成すればよい.

例えば以下のような確率質量函数 $P(x, y, z)$ の例を作ることができる:

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

x, y, z はそれぞれ $1, 0$ を動くとする. このとき, $P(x, y)$, $P(x, z)$, $P(y, z)$ の値はすべて $1/4$ になり, $P(x)$, $P(y)$, $P(z)$ の値はすべて $1/2$ になることがわかる. たとえば, $P(x = 1, y = 1) = P(1, 1, 1) + P(1, 1, 0) = 1/4 + 0 = 1/4$ であり, $P(x = 1) = \sum_{y,z} P(1, y, z) = 1/4 + 0 + 0 + 1/4 = 1/2$.

だから, $P(x, y) = P(x)P(y)$, $P(x, z) = P(x)P(z)$, $P(y, z) = P(y)P(z)$ が成立するが, $P(x, y, z) \neq 1/8 = P(x)P(y)P(z)$ となっている.

このとき, X, Y, Z が同時確率質量函数 $P(x, y, z)$ を持つとすると, そのうちの任意の異なる2つは独立だが, X, Y, Z の全体は独立ではない.

1.11 確率変数の独立性の現実における解釈に関する重大な注意

上の例は具体的には次のような状況だと解釈可能である.

(1) $P(x) = 1/2$ の解釈: $X = 1$ は薬Aを与えたことを, $X = 0$ は薬Bを与えたことを意味する. 全員に確率 $1/2$ で薬Aまたは薬Bを与えた.

$x = 1 =$ 薬A	$P(x = 1) = 1/2$
$x = 0 =$ 薬B	$P(x = 0) = 1/2$

(2) $P(y) = 1/2$ の解釈: $Y = 1$ は薬に効果があったことを, $Y = 0$ は薬に効果が無かったことを意味する. 全体で見たら, $1/2$ の確率で薬には効果があった.

$y=1=$ 効果有	$y=0=$ 効果無
$P(y=1)=1/2$	$P(y=0)=1/2$

(3) $P(z)=1/2$ の解釈: $Z=1$ は女性であることをを, $Z=0$ は男性であることを意味する. 女性である確率と男性である確率は $1/2$ だった.

$z=1=$ 女性	$z=0=$ 男性
$P(z=1)=1/2$	$P(z=0)=1/2$

(4) 男女の区別をやめると, 薬Aも薬Bも効果がある確率は半々であり, 薬Aと薬Bのどちらを与えたかと効果があったかどうかは独立である. 男女を合わせた($z=1,0$ の場合の和を取って得られる)確率質量函数 $P(x,y)$ の表

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

は$x=1$ の薬Aの場合も $x=0$ の薬Bの場合も男女の区別をやめると薬Aと薬Bで効果に変わりがないことを意味している.

(5) 薬Aと薬Bのどちらを与えたかと男女のどちらであるかは独立である. そのことは効果の有無を意味する $y=1,0$ について和を取って得られる確率質量函数 $P(x,z)$ の表

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

からわかる.

(6) 薬Aと薬Bのどちらを与えたかを無視すると, 男女のどちらであるかと薬の効果の有無は独立である. そのことは薬Aと薬Bのどちらを与えたかを意味する $x=1,0$ について和を取って得られる確率質量函数 $P(y,z)$ の表

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

からわかる.

(7) しかし, 男女を区別すると全然違う結果が見えて来る. 薬Aは女性だけに効果があり, 薬Bは男性だけに効果がある. $z=1$ の女性の場合に制限した確率質量函数 $P(x,y,z)$ $P(x,y,z)$ の表

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

より, $x=1$ の薬Aの場合には $y=1$ の効果有の確率が正であるが, $x=0$ の薬Bの場合には $y=1$ の効果有の確率が0になっている. $z=0$ の男性の場合に制限した確率質量函数 $P(x,y,z)$ の表

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

より, $x=1$ の薬Aの場合には $y=1$ の効果有の確率が0で効果がないが, $x=0$ の薬Bの場合には $y=1$ の効果有の確率が正になっている.

このように現実世界において確率変数達が独立か否かは重大な意味を持ち得る. ある重要な条件(上の場合には女性か男性か)を無視して, 「XとYは独立である」(どちらの薬を与えても効果は同じである)と結論すると大変なことになってしまうかもしれない. XとYも, XとZも, YとZも独立であっても, XとYとZの全体は独立でないかもしれない.

2 確率変数達の共分散と相関係数と無相関性

分散と共分散と相関係数と無相関性に関する理論は線形代数における有限実ベクトル空間における内積の理論と本質的に同じものであると考えられる. 分散, 共分散, 相関係数, 無相関性はそれぞれベクトルのノルムの2乗, 内積, ベクトルのあいだの角度 θ に関する $\cos\theta$, ベクトルの直交に対応している.

確率変数	線形代数
分散	ノルムの2乗
共分散	内積
相関係数	$\cos\theta$
無相関	直交

2.1 確率変数達の共分散と相関係数の定義

確率函数 X, Y の期待値をそれぞれ $\mu_X = E[X]$, $\mu_Y = E[Y]$ と書くことにする.

確率変数 X, Y の **共分散** (covariance) $\sigma_{XY} = \mathrm{cov}(X, Y)$ を次のように定義する:

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

$X = Y$ のときこれは X の分散になる. 分散の場合と同様に

$$\begin{aligned}\mathrm{cov}(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\ &= E[XY] - \mu_X E[Y] - E[X]\mu_Y + \mu_X\mu_Y \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] \\ &= E[XY] - E[X]E[Y].\end{aligned}$$

$X = Y$ の場合から $\mathrm{var}(X) = E[X^2] - E[X]^2$ が得られる.

線形代数の言葉を使えば, 共分散 $\sigma_{XY} = \mathrm{cov}(X, Y)$ は内積に対応しており, 分散 $\sigma_X^2 = \mathrm{var}(X)$, $\sigma_Y^2 = \mathrm{var}(Y)$ はノルム(ベクトルの長さ)の二乗に対応しており, 標準偏差 $\sigma_X = \mathrm{std}(X)$, $\sigma_Y = \mathrm{std}(Y)$ はノルム(ベクトルの長さ)に対応している.

さらに確率変数 X, Y の **相関係数** (correlation coefficient) $\rho_{XY} = \mathrm{cov}(X, Y)$ はベクトルのあいだのなす角度 θ に対する $\cos\theta$ の対応物として次のように定義される:

$$\rho_{XY} = \mathrm{cor}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\mathrm{std}(X)\,\mathrm{std}(Y)} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

相関係数という言葉を見たら $\cos\theta$ を想像し, 分散, 標準偏差, 共分散という言葉を見たら, ベクトルの長さの2乗, ベクトルの長さ, ベクトルの内積を想像すればよい. $X - \mu_X$ や $Y - \mu_Y$ がベクトルに対応している.

非常に重要な注意・警告: 以上で定義したのは確率変数達の共分散と相関係数であり, 後で定義する標本(サンプル, データ)の共分散と相関係数とは異なる数学的対象になるので注意が必要である. この点は平均と分散の場合と同様である. 我々は複雑な現実世界に立ち向かうための道具として数学的に統計モデルを設定し, 統計モデル内の住人として確率変数を考える. そのとき, 確率変数達の平均(期待値), 分散, 共分散, 相関係数を計算することは, 統計モデル内の確率変数の特徴を調べていることになる. 統計モデルを数学的フィクションである現実世界のデータと比較することは以下のように行われる:

1. 数学的フィクションである統計モデル内に住んでる現実世界のデータのモデル化になっている確率変数の特徴を調べる.
2. 現実世界で取得したデータの特徴を調べる.
3. 統計モデル内の確率変数の特徴と現実世界で取得したデータの特徴を比較してみる.

上で定義したの確率変数達の共分散と相関係数は1の段階で使用される. 2の段階で使われる道具については後で説明する.

数学的な注意: 統計学的な位置付けとして, 確率変数達の平均, 分散, 共分散, 相関係数などと, 標本(サンプル, データ)平均, 分散, 共分散, 相関係数は異なる対象だとみなす必要があることはすぐ上で説明した. しかし, 純粋に数学的性質を調べる場合には, 後者は前者の特別な場合になっていることから, 前者の数学的性質を調べれば後者の数学的性質もわかるという仕組みになっている. 初学者が混乱しがちになる原因の1つは, 数学的に鋭い初学者が確率変数達の扱いとデータ(サンプル, 標本)の取り扱いが数学的に似ていることに気付いてしまうせいで, それらの位置付けを混同してしまいがちなことにもあるように思われる.

2.2 確率変数達の対ごとの無相関性の定義

X, Y の相関係数 $\rho_{XY} = \mathrm{cor}(X, Y)$ が 0 になるとき, 確率変数達 X, Y は **無相関** であるという. X と Y が無相関であることは $X - \mu_X$ と $Y - \mu_Y$ が「直交している」と考えてもよい.

注意: X, Y の相関係数 $\rho_{XY} = \mathrm{cor}(X, Y)$ が **近似的に 0 になっている場合** には($\cos\theta$ に対応する量が近似的に 0 になっている場合には), 確率変数達 X, Y は **近似的に無相関** であると考える. そのとき注意するべきことは共分散 $\sigma_{XY} = \mathrm{cov}(X, Y)$ が近似的に 0 になっていても(内積に当たる量が近似的に 0 になっていても), 相関係数が 0 からほど遠い値になっていて, 近似的に無相関とは言えない場合があることである. 下にある関連問題の節の内容を参照せよ.

確率変数達 $X_1, \dots, X_n$ が **対ごとに無相関** であるとは, そのうちの互いに異なる任意の2つが無相関であることだと定める.
$\mu_i = E[X_i]$, $\sigma_i^2 = \mathrm{var}(X_i)$ のとき, $X_1, \dots, X_n$ が対ごとに無相関であることは,

$$\mathrm{cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = \sigma_i^2\delta_{ij}$$

と書けることと同値である. ここで δ_{ij} は $i = j$ の場合にのみ 1 になり, それ以外の場合に 0 になるKroneckerのデルタである.

2.3 問題: 確率変数の相関係数の計算例

確率変数の組 (X, Y) は確率 $0.1, 0.2, 0.3, 0.4$ でそれぞれ $(10, -50), (-4, -10), (0, 20), (2, 0)$ という値になると仮定する:

probability	0.1	0.2	0.3	0.4
X	10	-4	0	2
Y	-50	-10	20	0

確率変数 X, Y の相関係数を小数点以下第2桁まで求めよ.

解答例: まず X, Y の期待値を求める:

$$\begin{aligned}\mu_X &= E[X] = 10 \cdot 0.1 + (-4) \cdot 0.2 + 0 \cdot 0.3 + 2 \cdot 0.4 = 1,\\ \mu_Y &= E[Y] = (-50) \cdot 0.1 + (-10) \cdot 0.2 + 20 \cdot 0.3 + 0 \cdot 0.4 = -1.\end{aligned}$$

ゆえに期待値を X, Y から引いた値の分布は次の表のようになる:

probability	0.1	0.2	0.3	0.4
$X - \mu_X$	9	-5	-1	1
$Y - \mu_Y$	-49	-9	21	1

この表を用いて確率変数達としての X, Y の分散と共分散を求める:

$$\begin{aligned}\mathrm{var}(X) &= 9^2 \cdot 0.1 + (-5)^2 \cdot 0.2 + (-1)^2 \cdot 0.3 + 1^2 \cdot 0.4 = 13.8,\\ \mathrm{var}(Y) &= (-49)^2 \cdot 0.1 + (-9)^2 \cdot 0.2 + 21^2 \cdot 0.3 + 1^2 \cdot 0.4 = 389,\\ \mathrm{cov}(X, Y) &= 9(-49) \cdot 0.1 + (-5)(-9) \cdot 0.2 + (-1)21 \cdot 0.3 + 1 \cdot 1 \cdot 0.4 = -41.\end{aligned}$$

相関係数を求めよう:

$$\mathrm{cor}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}} \approx -0.56.$$

解答終

In [27]:
```
@show E_X = 10*0.1 + (-4)*0.2 + 0*0.3 + 2*0.4
@show E_Y = (-50)*0.1 + (-10)*0.2 + 20*0.3 + 0*0.4
```

```
E_X = 10 * 0.1 + -4 * 0.2 + 0 * 0.3 + 2 * 0.4 = 1.0
E_Y = -50 * 0.1 + -10 * 0.2 + 20 * 0.3 + 0 * 0.4 = -1.0
```

Out[27]:
```
-1.0
```

In [28]:
```
@show var_X = 9^2*0.1 + (-5)^2*0.2 + (-1)^2*0.3 + 1^2*0.4
@show var_Y = (-49)^2*0.1 + (-9)^2*0.2 + 21^2*0.3 + 1^2*0.4
@show cov_XY = 9*(-49)*0.1 + (-5)*(-9)*0.2 + (-1)*21*0.3 + 1*1*0.4;
```

```
var_X = 9 ^ 2 * 0.1 + (-5) ^ 2 * 0.2 + (-1) ^ 2 * 0.3 + 1 ^ 2 * 0.4 = 13.8
var_Y = (-49) ^ 2 * 0.1 + (-9) ^ 2 * 0.2 + 21 ^ 2 * 0.3 + 1 ^ 2 * 0.4 = 389.0
cov_XY = 9 * -49 * 0.1 + -5 * -9 * 0.2 + -1 * 21 * 0.3 + 1 * 1 * 0.4 = -41.0
```

In [29]:
```
@show cor_XY = cov_XY/(√var_X * √var_Y);
```

```
cor_XY = cov_XY / (√var_X * √var_Y) = -0.5595894432479921
```

In [30]:

```
X = [10, -4, 0, 2]
Y = [-50, -10, 20, 0]
P = [0.1, 0.2, 0.3, 0.4]
@show E_X = sum(X[i]*P[i] for i in eachindex(P))
@show E_Y = sum(Y[i]*P[i] for i in eachindex(P))
@show var_X = sum((X[i] - E_X)^2*P[i] for i in eachindex(P))
@show var_Y = sum((Y[i] - E_Y)^2*P[i] for i in eachindex(P))
@show cov_XY = sum((X[i] - E_X) * (Y[i] - E_Y) * P[i] for i in eachindex(P))
@show cor_XY = cov_XY/(√var_X * √var_Y);
```

```
E_X = sum((X[i] * P[i] for i = eachindex(P))) = 1.0
E_Y = sum((Y[i] * P[i] for i = eachindex(P))) = -1.0
var_X = sum(((X[i] - E_X) ^ 2 * P[i] for i = eachindex(P))) = 13.8
var_Y = sum(((Y[i] - E_Y) ^ 2 * P[i] for i = eachindex(P))) = 389.0
cov_XY = sum(((X[i] - E_X) * (Y[i] - E_Y) * P[i] for i = eachindex(P))) = -41.0
cor_XY = cov_XY / (√var_X * √var_Y) = -0.5595894432479921
```

2.4 問題: Cauchy-Schwarzの不等式

確率変数達 X, Y について, 以下の不等式を示せ:

$$|\operatorname{cov}(X,Y)| \leq \operatorname{std}(X)\operatorname{std}(Y).$$

簡単のため X, Y が同時確率質量函数を持つと仮定してよい.

注意: Cauchy-Schwarzの不等式より, $\operatorname{std}(X)\operatorname{std}(Y) \neq 0$ のとき, 相関係数

$$\rho_{XY} = \operatorname{cor}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\operatorname{std}(X)\operatorname{std}(Y)}$$

が -1 以上 1 以下になることがわかる. これより, 0 以上 π 以下の θ を

$$\rho_{XY} = \operatorname{cor}(X,Y) = \frac{\operatorname{cov}(X,Y)}{\operatorname{std}(X)\operatorname{std}(Y)} = \cos\theta$$

によって定めることができる. この θ はベクトルのあいだの角度の確率変数の場合の類似物になっている.

解答例1: 右辺の2乗から左辺の2乗を引いた結果が 0 以上になること, すなわち, 次を示せばよい:

$$\operatorname{var}(X)\operatorname{var}(Y) - \operatorname{cov}(X,Y)^2 \geq 0. \tag{$*$}$$

確率変数の分散や共分散は確率変数から定数を引いても変わらないので, X, Y からそれらの期待値を引いたものを考えることによって, X, Y の期待値は 0 であると仮定してよい. このとき,

$$\operatorname{var}(X) = E[X^2], \quad \operatorname{var}(Y) = E[Y^2], \quad \operatorname{cov}(X,Y) = E[XY].$$

X, Y の同時確率質量函数を $P(x,y)$ と書く:

$$E[f(X,Y)] = \sum_{x,y} f(x,y)P(x,y).$$

このとき,

$$E[f(X,Y)]E[g(X,Y)] = \sum_{x,y,x',y'} f(x,y)g(x',y')P(x,y)P(x',y').$$

以上の準備のもとで,

$$\begin{aligned}
2\operatorname{var}(X)\operatorname{var}(Y) &= E[X^2]E[Y^2] + E[Y^2]E[X^2] \\
&= \sum_{x,y,x',y'} (x^2y'^2 + y^2x'^2)P(x,y)P(x',y'), \\
2\operatorname{cov}(X,Y)^2 &= 2E[XY]E[XY] = \sum_{x,y,x',y'} 2xyx'y'P(x,y)P(x',y') \\
&= \sum_{x,y,x',y'} 2xy'yx'P(x,y)P(x',y').
\end{aligned}$$

ゆえに,

$$
\begin{aligned}
2\operatorname{var}(X)\operatorname{var}(Y) - 2\operatorname{cov}(X,Y)^2 &= \sum_{x,y,x',y'} (x^2y'^2 + y^2x'^2 - 2xy'yx')P(x,y)P(x',y') \\
&= \sum_{x,y,x',y'} (xy' - yx')^2 P(x,y)P(x',y') \geq 0.
\end{aligned}
$$

これより$(*)$が成立することがわかる.

解答終

上の解答例でやっていることは本質的に (X,Y) と独立で同分布な確率変数 (X',Y') に対応する離散変数 x', y' を導入したことである. そのことに気付けば, 同時確率質量函数の存在に頼らない一般的な証明法も得られる.

解答例2(より一般的な場合): 右辺の2乗から左辺の2乗を引いた結果が 0 以上になること, すなわち, 次を示せばよい:

$$
\operatorname{var}(X)\operatorname{var}(Y) - \operatorname{cov}(X,Y)^2 \geq 0. \tag{$*$}
$$

確率変数の分散や共分散は確率変数から定数を引いても変わらないので, X, Y からそれらの期待値を引いたものを考えることによって, X, Y の期待値は 0 であると仮定してよい. このとき,

$$
\operatorname{var}(X) = E[X^2], \quad \operatorname{var}(Y) = E[Y^2], \quad \operatorname{cov}(X,Y) = E[XY].
$$

さらに, (X',Y') は (X,Y) と同じ分布を持つ (X,Y) と独立な確率変数であるとする. このとき, 次の公式を使用できる:

$$
E[f(X,Y)] = E[f(X',Y')], \quad E[f(X,Y)g(X',Y')] = E[f(X,Y)]E[g(X',Y')].
$$

以上の準備のもとで,

$$
\begin{aligned}
&2\operatorname{var}(X)\operatorname{var}(Y) = 2E[X^2]E[Y^2] = E[X^2]E[Y'^2] + E[Y^2]E[X'^2] = E[X^2Y'^2 + Y^2X'^2], \\
&2\operatorname{cov}(X,Y)^2 = 2E[XY]^2 = 2E[XY]E[X'Y'] = E[2XY'YX'].
\end{aligned}
$$

ゆえに,

$$
\begin{aligned}
2\operatorname{var}(X)\operatorname{var}(Y) - 2\operatorname{cov}(X,Y)^2 &= E[X^2Y'^2 + Y^2X'^2] - E[2XY'YX'] \\
&= E[X^2Y'^2 + Y^2X'^2 - 2XY'YX'] \\
&= E[(XY' - X'Y)^2] \geq 0.
\end{aligned}
$$

これより$(*)$が成立することがわかる.

解答終

注意: 線形代数の教科書におけるCauchy-Schwarzの不等式の証明と以上で紹介した証明を比較してみよ. もしも比較した線形代数の教科書の証明が以上で紹介した証明と全く違っているならば, その教科書の文脈において上で紹介した証明の方法が使えないかどうかを考えてみよ. 参照にした教科書に上で紹介した証明と全く異なる証明が載っているならば, 上で証明した確率変数に関するCauchy-Schwarzの不等式をその教科書の方法で証明してみよ.

2.5 問題: 等確率有限離散分布の相関係数と $\cos\theta$ の関係

$a_i, b_i \in \mathbb{R}$ であるとし, 確率変数の組 (X,Y) は同じ確率 $1/n$ で $(a_1,b_1),\dots,(a_n,b_n)$ のどれかの値になるものと仮定する. この仮定は

$$
E[f(X,Y)] = \frac{1}{n}\sum_{i=1}^{n} f(a_i,b_i)
$$

の成立と同値である(確率変数の組が定義されていることはその函数の期待値が定義されていることと同じことだと考えてよい). さらに,

$$
\mu_X = E[X] = \frac{1}{n}\sum_{i=1}^{n} a_i, \quad \mu_Y = E[Y] = \frac{1}{n}\sum_{i=1}^{n} b_i
$$

とおき, ベクトル v_X, v_Y を次のように定める:

$$
v_X = \frac{1}{\sqrt{n}}\begin{bmatrix} a_1 - \mu_X \\ \vdots \\ a_n - \mu_X \end{bmatrix}, \quad v_Y = \frac{1}{\sqrt{n}}\begin{bmatrix} b_1 - \mu_Y \\ \vdots \\ b_n - \mu_y \end{bmatrix}
$$

これらの内積(対応する成分の積の和)を (v_X, v_Y) と書き, ノルム(成分の2乗の和の平方根)を $\|v_X\|, \|v_Y\|$ と書き, v_X, v_Y のあいだの角度を θ と書くことにし, $\|v_X\| > 0$, $\|v_Y\| > 0$ であると仮定する. このとき, X, Y の相関係数 $\operatorname{cor}(X,Y)$ について次が成立することを示せ:

$$\mathrm{cor}(X,Y) = \frac{(v_X, v_Y)}{\|v_X\|\|v_Y\|} = \cos\theta.$$

注意: 相関係数とベクトルのあいだの角度の関係についてはこの場合をイメージしておけばよい.

解答例: 内積とノルムの定義より,

$$\begin{aligned}
\|v_X\|^2 &= \frac{1}{n}\sum_{i=1}^{n}(a_i - \mu_X)^2 = E[(X-\mu_X)^2] = \mathrm{var}(X),\\
\|v_Y\|^2 &= \frac{1}{n}\sum_{i=1}^{n}(b_i - \mu_Y)^2 = E[(Y-\mu_Y)^2] = \mathrm{var}(Y),\\
(v_X, v_Y) &= \frac{1}{n}\sum_{i=1}^{n}(a_i - \mu_X)(b_i - \mu_Y) = E[(X-\mu_X)(Y-\mu_Y)] = \mathrm{cov}(X,Y).
\end{aligned}$$

ゆえに,

$$\mathrm{cor}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}} = \frac{(v_X, v_Y)}{\|v_X\|\|v_Y\|} = \cos\theta.$$

解答終

2.6 問題: 相関係数の計算

確率変数達 X, Y は無相関であるとし, それぞれの期待値と分散はどちらも 0 と 1 であるとし, $a, b, c, d \in \mathbb{R}$, $a^2 + b^2 > 0$, $c^2 + d^2 > 0$ であると仮定する. 確率変数 A, B を次のように定める:

$$A = aX + bY, \quad B = cX + dY.$$

確率変数達 A, B の相関係数を求めよ.

解答例: X, Y の期待値はどちらも 0 なのでそれらの一次結合である A, B の期待値も 0 になる. さらに, X, Y の分散はどちらも 1 でそれらは無相関だと仮定したので,

$$\begin{aligned}
&\mathrm{var}(A) = a^2\,\mathrm{var}(X) + b^2\,\mathrm{var}(Y) = a^2 + b^2,\\
&\mathrm{var}(B) = c^2\,\mathrm{var}(X) + d^2\,\mathrm{var}(Y) = c^2 + d^2,\\
&\mathrm{cov}(A,B) = E[AB] = E[acX^2 + (ad+bc)XY + bdY^2]\\
&\quad = ac\,\mathrm{var}(X) + (ad+bc)\,\mathrm{cov}(X,Y) + bd\,\mathrm{var}(Y) = ac + bd.
\end{aligned}$$

ゆえに

$$\mathrm{cor}(A,B) = \frac{\mathrm{cov}(A,B)}{\sqrt{\mathrm{var}(A)}\sqrt{\mathrm{var}(B)}} = \frac{ac+bd}{\sqrt{a^2+b^2}\sqrt{c^2+d^2}}.$$

解答終

注意: ベクトル v_A, v_B を

$$v_A = \begin{bmatrix} a \\ b \end{bmatrix}, \quad v_B = \begin{bmatrix} c \\ d \end{bmatrix}$$

と定め, これらの標準内積(対応する成分の積の和)を (v_A, v_B) と書き, それらのノルム(すべての成分の2乗の和の平方根)を $\|v_A\|, \|v_B\|$ と書き, それらの間の角度を θ と書くとき,

$$\mathrm{cor}(A,B) = \frac{(v_A, v_B)}{\|v_A\|\|v_B\}} = \cos\theta$$

が成立している. 上の問題の状況において, 確率変数 A, B の相関係数が 1 に近いことはベクトル v_A, v_B の向きがほぼ同じであることと同値であり, A, B の相関係数が -1 に近いことは v_A, v_B の向きがほぼ正反対であることに同値であり, A, B が無相関に近いことは v_A, v_B がほぼ直交していることと同値である.

2.7 問題: 共分散が 0 に近くても相関係数が 0 から遠い場合がある

平均が 0 の具体的な確率変数達 X, Y の組で, 共分散 $\sigma_{XY} = \mathrm{cov}(X,Y) = E[XY]$ は 0 に近い値だが, 相関係数 $\rho_{XY} = \mathrm{cov}(X,Y) = \mathrm{cov}(X,Y)/(\mathrm{std}(X)\,\mathrm{std}(Y))$ が 0 から遠い値になっているものを構成せよ.

解答例: $\varepsilon, \delta > 0$ であるとし, 確率変数の組 (X,Y) は同じ確率 1/3 で $(\varepsilon+\delta, \varepsilon)$, $(-\varepsilon, \delta)$, $(-\delta, -\varepsilon-\delta)$ のどれかの値になるものとする. このとき, X, Y の平均はそれぞれ

$$E[X] = \frac{(\varepsilon+\delta)+(-\varepsilon)+(-\delta)}{3} = 0, \quad E[Y] = \frac{\varepsilon+\delta+(-\varepsilon-\delta)}{3} = 0$$

となる. X と Y の分散と共分散と相関係数は以下のようになる:

$$\begin{aligned}
\mathrm{var}(X) &= E[X^2] = \frac{(\varepsilon+\delta)^2+(-\varepsilon)^2+(-\delta)^2}{3} = \frac{2(\varepsilon^2+\varepsilon\delta+\delta^2)}{3}, \\
\mathrm{var}(Y) &= E[Y^2] = \frac{\varepsilon^2+\delta^2+(-\varepsilon-\delta)^2}{3} = \frac{2(\varepsilon^2+\varepsilon\delta+\delta^2)}{3}, \\
\mathrm{cov}(X,Y) &= E[XY] = \frac{(\varepsilon+\delta)\varepsilon+(-\varepsilon)\delta+(-\delta)(-\varepsilon-\delta)}{3} = \frac{(\varepsilon^2+\varepsilon\delta+\delta^2)}{3}, \\
\mathrm{cor}(X,Y) &= \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Y)}} = \frac{1}{2}.
\end{aligned}$$

$\varepsilon, \delta > 0$ の値が小さいとき, X, Y の共分散は小さくなるが, 相関係数は 1/2 で全然 0 に近くない. **解答終**

注意: 3次元ベクトル v_X, v_Y を以下のように定める:

$$v_X = \frac{1}{\sqrt{3}}\begin{bmatrix} \varepsilon+\delta \\ -\varepsilon \\ -\delta \end{bmatrix}, \quad v_Y = \frac{1}{\sqrt{3}}\begin{bmatrix} \varepsilon \\ \delta \\ -\varepsilon-\delta \end{bmatrix}.$$

このとき, これらの標準内積(対応する成分の積の和)を (v_X, v_Y) と書き, それらのノルム(すべての成分の2乗の和の平方根)を $\|v_X\|, \|v_Y\|$ と書き, それらの間の角度を θ と書くとき,

$$\begin{aligned}
&\mathrm{var}(X) = \|v_X\|^2, \quad \mathrm{var}(Y) = \|v_Y\|^2, \quad \mathrm{cov}(X,Y) = (v_X, v_Y), \\
&\mathrm{cor}(X,Y) = \frac{(v_X, v_Y)}{\|v_X\|\|v_Y\|} = \cos\theta
\end{aligned}$$

が成立していることに注意せよ.

2.8 問題: 独立ならば無相関である (実質1行で解ける)

確率変数達 X, Y が独立ならば確率変数達 X, Y は無相関であることを示せ.

解答例: X, Y は独立なので $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$ となる. ゆえに, $\mu_X = E[X]$, $\mu_Y = E[Y]$ とおくと,

$$E[(X-\mu_X)(Y-\mu_Y)] = E[X-\mu_X]E[Y-\mu_Y] = (E[X]-\mu_X)(E[Y]-\mu_Y) = 0 \cdot 0 = 0.$$

解答終

2.9 問題: 無相関でも独立とは限らない

確率変数 X, Y で無相関だが独立でないものを具体的に構成せよ.

解答例1: 確率質量函数 $P(x, y)$ を次の表の通りに定める:

	$y=1$	$y=2$	$y=3$	
$x=1$	$P(1,1)=0$	$P(1,2)=1/4$	$P(1,3)=0$	$P(x=1)=1/4$
$x=2$	$P(2,1)=1/4$	$P(2,2)=0$	$P(2,3)=1/4$	$P(x=2)=2/4$
$x=3$	$P(3,1)=0$	$P(3,2)=1/4$	$P(3,3)=0$	$P(x=3)=1/4$
	$P(y=1)=1/4$	$P(y=2)=2/4$	$P(y=3)=1/4$	

確率変数 X, Y は同時確率質量函数 $P(x, y)$ を持つとする. このとき,

$$E[X] = E[Y] = 1 \cdot (1/4) + 2 \cdot (2/4) + 3 \cdot (1/4) = 2.$$

であり, 確率が 0 でない (x, y) については x または y が X と Y の期待値 2 になるので, 共分散は 0 になる:

$$\begin{aligned}
\mathrm{cov}(X,Y) = E[(X-2)(Y-2)] &= \sum_{x,y}(x-2)(y-2)P(x,y) \\
&= (2-2)(1-2)(1/4) + (1-2)(2-2)(1/4) \\
&\quad + (3-2)(2-2)(1/4) + (2-2)(3-2)(1/4) = 0.
\end{aligned}$$

これで X, Y は無相関であることがわかった.

しかし, たとえば $P(1,1) = 0$ と $P(x=1)P(y=1) = (1/4)(1/4)$ が等しくないので X, Y は独立ではない.

解答終

解答例2: 確率質量函数 $P(x, y)$ を次の表の通りに定めても上と同様に, X, Y は無相関だが独立ではないことを示せる:

	$y = 1$	$y = 2$	$y = 3$	
$x = 1$	$P(1, 1) = 1/8$	$P(1, 2) = 1/8$	$P(1, 3) = 1/8$	$P(x = 1) = 3/8$
$x = 2$	$P(2, 1) = 1/8$	$P(2, 2) = 0$	$P(2, 3) = 1/8$	$P(x = 2) = 2/8$
$x = 3$	$P(3, 1) = 1/8$	$P(3, 2) = 1/8$	$P(3, 3) = 1/8$	$P(x = 3) = 3/8$
	$P(y = 1) = 3/8$	$P(y = 2) = 2/8$	$P(y = 3) = 3/8$	

解答終

注意: 3×3 の表をすべて 1/9 で埋めてしまうと, X, Y は独立になってしまう.

解答例3: $\mathbb{R}^2$ における単位円盤上の一様分布の確率密度函数を次のように定める:

$$p(x, y) = \begin{cases} 1/\pi & (x^2 + y^2 \leq 1) \\ 0 & (x^2 + y^2 > 1) \end{cases}$$

これを同時確率密度函数として持つ確率変数 X, Y を考える:

$$E[f(X, Y)] = \iint_{\mathbb{R}^2} f(x, y)p(x, y)\, dx\, dy.$$

単位円盤の対称性から $E[X] = E[Y] = 0$ となることがわかる. (具体的に積分を計算しても易しい.) それらの共分散は

$$\mathrm{cov}(X, Y) = E[XY] = \frac{1}{\pi} \iint_{x^2+y^2 \leq 1} xy\ dx\, dy$$

と書けるが, $xy \geq 0$ の部分の積分と $xy \leq 0$ の部分の積分が単位円盤の対称性より互いに打ち消しあって $\mathrm{cov}(X, Y) = 0$ となることがわかる. X 単独の密度函数は

$$p(x) = \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = \frac{2}{\pi}\sqrt{1 - x^2}$$

となり, 同様にして $p(y) = (2/\pi)\sqrt{1 - y^2}$ となるが, $-1 < x, y < 1$ かつ $x^2 + y^2 > 1$ のとき, $p(x, y) = 0$ となるが, $p(x)p(y) \neq 0$ となるので, それらは等しくない. ゆえに X, Y は独立ではない.

解答終

この解答例3については, 下の方でプロットした (X, Y) の分布のサンプルの散布図も参照して欲しい. 散布図を見れば直観的にどうなっているかを把握しやすい.

In [31]:

```
n = 10^4
XY = [(r = √rand(); t = 2π*rand(); (r*cos(t), r*sin(t))) for _ in 1:n]
X, Y = first.(XY), last.(XY)
@show cov(X, Y) cor(X, Y)
P1 = scatter(X, Y; msc=:auto, ms=2, alpha=0.7, label="", xlabel="x", ylabel="y")
P2 = histogram(X; norm=true, alpha=0.3, bin=41, label="X")
plot!(x -> 2/π*√(1 - x^2), -1, 1; label="", lw=2)
P3 = histogram(X; norm=true, alpha=0.3, bin=41, label="Y")
plot!(y -> 2/π*√(1 - y^2), -1, 1; label="", lw=2)
plot(P1, P2, P3; size=(800, 400), layout=@layout [a [b; c]])
```

```
cov(X, Y) = -0.0005083072927225841
cor(X, Y) = -0.0020186738360558273
```

Out[31]:

単位円盤上の一様分布は「無相関だが独立ではない場合」の例になっている.

この「無相関だが独立ではない場合」の例のサンプルに含まれる点の個数を減らすと以下のように見える.

In [32]:

```
n = 100
XY = [(r = √rand(); t = 2π*rand(); (r*cos(t), r*sin(t))) for _ in 1:n]
X, Y = first.(XY), last.(XY)
@show cov(X, Y) cor(X, Y)
P1 = scatter(X, Y; msc=:auto, label="", xlabel="x", ylabel="y")
plot!(; xlim=(-1.05, 1.05), ylim=(-1.05, 1.05))
plot!(; size=(400, 400))
```

```
cov(X, Y) = 0.011653127995250636
cor(X, Y) = 0.046497931398409825
```

Out[32]:

上の場合において X, Y はどちらも **半円分布** (semicircle distribution)に従っているが, 独立ではない. X, Y の分布が以上と同じ半円分布のときと X, Y が独立になるならば, (X, Y) の分布のサンプルは以下のようになる.

In [33]:

```
n = 10^4
sc = Semicircle(1)
X, Y = rand(sc, n), rand(sc, n)
@show cov(X, Y) cor(X, Y)
P1 = scatter(X, Y; msc=:auto, ms=2, alpha=0.7, label="", xlabel="x", ylabel="y")
P2 = histogram(X; norm=true, alpha=0.3, bin=41, label="X")
plot!(x -> 2/π*√(1 - x^2), -1, 1; label="", lw=2)
P3 = histogram(X; norm=true, alpha=0.3, bin=41, label="Y")
plot!(y -> 2/π*√(1 - y^2), -1, 1; label="", lw=2)
plot(P1, P2, P3; size=(800, 400), layout=@layout [a [b; c]])
```

```
cov(X, Y) = 0.004473060988369731
cor(X, Y) = 0.0179942903137646
```

Out[33]:

これは X, Y が独立な場合であり, ゆえに無相関にもなっている.

サンプルが含む点の個数を減らすと以下のように見える.

In [34]:

```
n = 100
sc = Semicircle(1)
X, Y = rand(sc, n), rand(sc, n)
@show cov(X, Y)
P1 = scatter(X, Y; msc=:auto, label="", xlabel="x", ylabel="y")
plot!(; xlim=(-1.05, 1.05), ylim=(-1.05, 1.05))
plot!(; size=(400, 400))
```

```
cov(X, Y) = 0.0030840296722668315
```

Out[34]:

2.10 問題: 対ごとに無相関な確率変数達の和の分散はそれぞれの分散の和になる

$X_1, \dots, X_n$ は対ごとにも無相関とは限らない(ゆえに独立とは限らない)確率変数達であるとする. このとき, 期待値を取る操作の線形性より,

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$$

となる. 確率変数達の和の期待値は, 独立性や無相関性が成立していなくても, それぞれの期待値の和になる.

対ごとの無相関性を仮定すると, 分散についても簡明な結果を得ることができる.

$X_1, \dots, X_n$ が **対ごとに無相関な確率変数達** ならば(特に独立な確率変数ならば), 次が成立することを示せ:

$$\mathrm{var}(X_1 + \cdots + X_n) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n).$$

この結果は今後空気のごとく使われる.

ヒント: 互いに直交するベクトル達 $v_1, \dots, v_n$ について, 内積を $(\ ,\ )$ と書くとき, $(v_i, v_j) = \delta_{ij}\|v_i\|^2$ が成立することを使えば, $\|v_1 + \cdots + v_n\|^2 = \|v_1\|^2 + \cdots + \|v_n\|^2$ を示せることと本質的に同じ. この結果はPythagorasの定理(平面の場合は三平方の定理)そのものである.

解答例: $\mu_i = E[X_i]$ とおくと,

$$E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n \mu_i$$

が成立しており, さらに, $\sigma_i^2 = \mathrm{var}(X_i) = E[(X_i - \mu_i)^2]$ とおくと, $X_1, \dots, X_n$ が対ごとに無相関であることより,

$$E[(X_i - \mu_i)(X_j - \mu_j)] = \mathrm{cov}(X_i, X_j) = \sigma_i^2 \delta_{ij}$$

が成立しているので(δ_{ij} は $i = j$ の場合にのみ 1 でそれ以外のとき 0), 一般に

$$\left(\sum_{i=1}^n a_i\right)^2 = \sum_{i=1}^n a_i \sum_{i=j}^n a_j = \sum_{i,j=1}^n a_i a_j$$

と計算できることを使うと,

$$\begin{aligned}
\mathrm{var}\left(\sum_{i=1}^n X_i\right) &= E\left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i\right)^2\right] = E\left[\left(\sum_{i=1}^n (X_i - \mu_i)\right)^2\right] \\
&= E\left[\sum_{i,j=1}^n (X_i - \mu_i)(X_j - \mu_j)\right] = \sum_{i,j=1}^n E\left[(X_i - \mu_i)(X_j - \mu_j)\right] \\
&= \sum_{i,j=1}^n \sigma_i^2 \delta_{ij} = \sum_{i=1}^n \sigma_i^2 = \sum_{i=1}^n \mathrm{var}(X_i).
\end{aligned}$$

解答終

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

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

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

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

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

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

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

になることを示せ. さらに, このとき

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

となることも示せ.

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

解答例1: まず $\hat{\mu}$ の分散の式を整理しよう:

$$\mathrm{var}(\hat{\mu}) = E\left[\left(\sum_{i=1}^n w_i X_i - \mu\right)^2\right] = E\left[\left(\sum_{i=1}^n w_i (X_i - \mu)\right)^2\right] = \sum_{i=1}^n w_i^2 \sigma_i^2.$$

2つめの等号で $\sum_{i=1}^n w_i = 1$ であることを使い, 3つめの等号で $w_i X_i$ 達が対ごとに無相関であることを使った(一般に対ごとに無相関な確率変数の和の分散がそれぞれの分散の和になるのであった). $w_n = 1 - \sum_{i=1}^{n-1} w_i$ とおいて, $w_1, \dots, w_{n-1}$ を独立変数とみなすとき, それらによる $\mathrm{var}(\hat{\mu})$ の偏微分が 0 になるという方程式の解が求める w_i 達になる. $w_n = 1 - \sum_{i=1}^{n-1} w_i$ に注意すれば, $i = 1, \dots, n-1$ について

$$0 = \frac{1}{2}\frac{\partial}{\partial w_i}\mathrm{var}(\hat{\mu}) = w_i \sigma_i^2 - \left(1 - \sum_{i=1}^{n-1} w_i\right)\sigma_i^2 = w_i \sigma_i^2 - w_n \sigma_i^2$$

なので, この方程式の解について, $w_i \sigma_i$ 達は互いにすべて等しくなり, w_i 達の総和は 1 になる. そのとき, 以下のようになることがわかる:

$$w_i = \frac{1/\sigma_i^2}{\sum_{i=1}^n 1/\sigma_i^2}, \quad \mathrm{var}(\hat{\mu}) = \sum_{i=1}^n w_i^2 \sigma_i^2 = \frac{1}{\sum_{i=1}^n 1/\sigma_i^2} < \frac{1}{1/\sigma_i^2} = \sigma_i^2.$$

解答終

答えが分かっているので, 以下のように平方完成の方法によって偏微分を使わずにも証明できる.

解答例2: まず,

$$A = \sum_{i=1}^n 1/\sigma_i^2$$

とおいて, $\sum_{i=1}^n w_i = 1$ の仮定のもとで, $\hat{\mu}$ の分散の式を整理しよう:

$$\begin{aligned}
\mathrm{var}(\hat{\mu}) &= E\left[\left(\sum_{i=1}^n w_i X_i - \mu\right)^2\right] = E\left[\left(\sum_{i=1}^n w_i (X_i - \mu)\right)^2\right] \\
&= E\left[\sum_{i=1}^n w_i^2 (X_i - \mu)^2 + \sum_{i \neq j} w_i w_j (X_i - \mu)(X_j - \mu)\right] \\
&= \sum_{i=1}^n w_i^2 E[(X_i - \mu)^2] + \sum_{i \neq j} w_i w_j E[(X_i - \mu)(X_j - \mu)] \\
&= \sum_{i=1}^n w_i^2 E[(X_i - \mu)^2] = \sum_{i=1}^n w_i^2 \sigma_i^2 = \sum_{i=1}^n \sigma_i^2 \left(w_i - \frac{1/\sigma_i^2}{A} + \frac{1/\sigma_i^2}{A}\right)^2 \\
&= \sum_{i=1}^n \sigma_i^2 \left(\left(w_i - \frac{1/\sigma_i^2}{A}\right)^2 - 2\frac{1/\sigma_i^2}{A}\left(w_i - \frac{1/\sigma_i^2}{A}\right) + \frac{(1/\sigma_i^2)^2}{A^2}\right) \\
&= \sum_{i=1}^n \sigma_i^2 \left(w_i - \frac{1/\sigma_i^2}{A}\right)^2 + \frac{1}{A}.
\end{aligned}$$

2つめの等号で $\sum_{i=1}^n w_i = 1$ であることを使い, 5つめの等号で X_i 達が対ごとに無相関であること($i \neq j$ ならば $E[(X_i - \mu)(X_j - \mu)]$)を使い, 最後の等号で $\sum_{i=1}^n w_i = 1$ と $\sum_{i=1}^n 1/\sigma_i^2 = A$ を使った.

このとき, $\sum_{i=1}^n (1/\sigma_i^2)/A = 1$ に注意すれば, $\mathrm{var}(\hat{\mu})$ は $\sum_{i=1}^n w_i = 1$ という条件のもとで $w_i = (1/\sigma_i^2)/A$ のときに最小になり, 最小値は $1/A$ になることがわかる. そして, さらに,

$$\frac{1}{A} = \frac{1}{\sum_{i=1}^n 1/\sigma_i^2} < \frac{1}{1/\sigma_i^2} = \sigma_i^2.$$

解答終

以下の図も参照:

原点からの距離は誤差の大きさ,
直交は無相関を意味する.

一般に $E[X_1] = \cdots = E[X_n] = \mu$, $E[(X_i-\mu)(X_j-\mu)] = \delta_{ij}\sigma_i^2$ のとき,
$\sum_{i=1}^n w_i = 1$ のときの $\hat{\mu} = \sum_{i=1}^n w_i X_i$ の分散は $w_i = \frac{1/\sigma_i^2}{\sum_{j=1}^n 1/\sigma_j^2}$ のとき
最小になり, その最小値は $\frac{1}{\sum_{j=1}^n 1/\sigma_j^2} < \sigma_i^2$ になる.

2.12 問題: 二項分布と負の二項分布の平均と分散のBernoulli分布と幾何分布の場合への帰着

Bernoulli分布 $\mathrm{Bernoulli}(p)$ の平均と分散がそれぞれ p, $p(1-p)$ であることと, 幾何分布 $\mathrm{Geometric}(p)$ の平均と分散がそれぞれ $(1-p)/p$, $(1-p)/p^2$ であることを認めて, 二項分布 $\mathrm{Binomial}(n, p)$ と負の二項分布 $\mathrm{NegativeBinomial}(k, p)$ の平均と分散を平易な計算で求めてみよ. 以下を示せ:

(1) $K \sim \mathrm{Binomial}(n, p)$ ならば $E[K] = np$, $\mathrm{var}(K) = np(1-p)$.

(2) $M \sim \mathrm{NegativeBinomial}(k, p)$ ならば $E[M] = k(1-p)/p$, $\mathrm{var}(M) = k(1-p)/p^2$.

解答例:

二項分布は試行回数 n の成功確率 p のBernoulli試行で生成された 1 と 0 からなる長さ n の数列中に含まれる 1 の個数の分布であった. Bernoulli試行の確率質量函数は

$$P(x_1, \ldots, x_n) = P(x_1) \cdots P(x_n), \quad P(x_i) = p^{x_i}(1-p)^{1-x_i} \quad (x_i = 1, 0)$$

とBernoulli分布の確率質量函数 $P(x_i)$ の **積で書ける** のであった. この事実はBernoulli分布に従う確率変数達 $X_1, \ldots, X_n$ が独立であることを意味する. そして, Bernoulli試行で生成された 1 と 0 からなる長さ n の数列中に含まれる 1 の個数を意味する確率変数は $K = \sum_{i=1}^n X_i$ と書ける. このことから,

$$E[K] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n p = np,$$
$$\mathrm{var}(K) = \sum_{i=1}^n \mathrm{var}(X_i) = \sum_{i=1}^n p(1-p) = np(1-p).$$

となることがわかる.

幾何分布は成功確率 p のBernoulli試行を 1 が1つ出るまで続けたときに出た 0 の個数の分布であった. $M_1, \dots, M_k$ はそれぞれが成功確率 p の幾何分布に従う独立な確率変数であるとする. このとき, $M = \sum_{i=1}^k M_i$ はBernoulli試行を 1 が k 回出るまで続けたときに 0 が出た個数に等しい. M_i は $i-1$ 番目の 1 から i 番目の 1 が出るまでに出た 0 の個数を意味する確率変数だと解釈される. このことは, M が負の二項分布 $\mathrm{NegativeBinomial}(k, p)$ に従う確率変数になることを意味する. このことから,

$$E[M] = \sum_{i=1}^k E[M_i] = \sum_{i=1}^k \frac{1-p}{p} = \frac{k(1-p)}{p},$$
$$\mathrm{var}(M) = \sum_{i=1}^k \mathrm{var}(M_i) = \sum_{i=1}^k \frac{1-p}{p^2} = \frac{k(1-p)}{p^2}.$$

となることがわかる.

解答終

注意: 計算が大幅に簡単になった!

2.13 問題: 番号が異なる確率変数達が対ごとに無相関なときの確率変数の和の共分散

確率変数達 $X_1, Y_1, \dots, X_n, Y_n$ について次が成立していると仮定する:

$$\mathrm{cov}(X_i, Y_j) = \delta_{ij}\,\mathrm{cov}(X_i, Y_i).$$

このとき, 次が成立することを示せ:

$$\mathrm{cov}(X_1 + \cdots + X_n, Y_1 + \cdots + Y_n) = \mathrm{cov}(X_1, Y_1) + \cdots + \mathrm{cov}(X_n, Y_n).$$

ヒント: ベクトル達 $u_1, v_2, \dots, u_n, v_n$ の中の2つの異なる添え字を持つ u_i と v_j が互いに直交するならば, 内積 (,) について

$$(u_1 + \cdots + u_n, v_1 + \cdots + v_n) = (u_1, v_1) + \cdots + (u_n, v_n)$$

が成立することと本質的に同じことである.

解答例:

記号の簡単のため $A_i = X_i - E[X_i]$, $B_i = Y_i - E[Y_i]$ とおく. このとき, $E[A_i] = E[B_i] = 0$ より,

$$\mathrm{var}(A_i) = E[A_i^2], \quad \mathrm{var}(B_i) = E[B_i^2], \quad \mathrm{cov}(A_i, B_j) = E[A_i B_j].$$

A_i, B_i 達について上の問題を解けばよい. 問題の仮定より, A_i, B_i 達について次が成立している:

$$E[A_i B_j] = \delta_{ij} E[A_i B_i].$$

ゆえに,

$$E\left[\left(\sum_{i=1}^n A_i\right)\left(\sum_{j=1}^n B_j\right)\right] = \sum_{i,j=1}^n E[A_i B_j] = \sum_{i,j=1}^n \delta_{ij} E[A_i B_i] = \sum_{i=1}^n E[A_i B_i].$$

これは

$$\mathrm{cov}(A_1 + \cdots + A_n, B_1 + \cdots + B_n) = \mathrm{cov}(A_1, B_1) + \cdots + \mathrm{cov}(A_n, B_n).$$

の成立を意味する.

解答終

2.14 分散共分散行列とその半正定値性

n 個の確率変数を縦に並べてできるベクトル X と実数を成分に持つベクトル a を考える:

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, \quad a = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^n.$$

このとき, X, a の転置をそれぞれ X, a^T と書くと,

$$a^T X = X^T a = \sum_{i=1}^n a_i X_i$$

なので

$$a^T XX^T a = \left(\sum_{i=1}^{n} a_i X_i \right)^2 \geq 0.$$

ゆえに, 期待値を取る操作の線形性と単調性より,

$$0 \leq E[a^T XX^T a] = a^T E[XX^T]a$$

でかつ, XX^T は $X_i X_j$ を (i, j) 成分とする $n \times n$ 行列になるので, $E[XX^T]$ は $E[X_i X_j]$ を (i, j) 成分とする $n \times n$ の実対称行列になる. 上の計算より, その実対称行列の固有値はすべて 0 以上になることがわかる.

ここで $E[X_i] = 0$ と仮定する. このとき, $E[XX^T]$ は i 番目の対角成分が X_i の分散 $\mathrm{var}(X_i)$ で, $i \neq j$ のときの (i, j) 成分が共分散 $\mathrm{cov}(X_i, X_j)$ であるような $n \times n$ の実対称行列になる.

このとき, 行列 $E[XX^T]$ を **分散共分散行列** と呼ぶ.

分散共分散行列の固有値はすべて 0 以上になる. この結果を分散共分散行列の **半正定値性** と呼ぶことにする. 特に分散共分散行列式は 0 以上になる.

3 標本(サンプル, データ)の平均と分散と共分散と相関係数

3.1 標本平均の定義

n 個の数値 $x_1, \ldots, x_n$ で構成されたデータ(標本, サンプル)に対して,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = \frac{x_1 + \cdots + x_n}{n}$$

をその **標本平均** (sample mean)と呼ぶ.

統計モデルを設定するとは現実世界で得られるデータ(標本, サンプル)の生じ方を確率分布の言葉でモデル化することである. そのとき, 統計モデル内部では, データ(標本, サンプル)を, n 個の数値ではなく, n 個の確率変数達 $X_1, \ldots, X_n$ でモデル化することになる. その場合には, 確率変数達の加法平均

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \frac{X_1 + \cdots + X_n}{n}$$

をその **標本平均** (sample mean)と呼ぶ. X_i 達が確率変数であることより, それらの標本平均 $\bar{X}$ も確率変数になることに注意せよ. この文脈での標本平均 $\bar{X}$ は確率モデル内部における確率変数になる.

統計学入門の文脈では, 標本のモデル化とみなされるn 個の確率変数達 $X_1, \ldots, X_n$ は独立同分布だと仮定することが多い.

以下では, X_i 達は同じ期待値と分散を持ち, 対ごとに無相関であると仮定する(この条件は独立同分布よりも弱い).

注意・警告: 標本のモデル化としての確率変数達の **標本平均と確率変数の平均(期待値)を明瞭に区別しなければいけない**. 各確率変数 X_i の期待値 $E[X_i]$ や確率変数としての標本平均 $\bar{X}$ の期待値

$$E[\bar{X}] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \mu$$

と標本平均 $\bar{X}$ を明瞭に区別することが必要になる.

統計モデル内での標本のモデル化としての確率変数達 $X_1, \ldots, X_n$ の標本平均 $\bar{X}$ は確率変数なので, 数学的には函数とみなされ, 直観的にはランダムに値が決まる変数だとみなされるものになる. それに対して, 標本平均 $\bar{X}$ の期待値 $E[\bar{X}]$ は単なる数値になる. このように, 確率変数としての標本平均 $\bar{X}$ とその期待値 $E[\bar{X}]$ は全く異なる数学的対象になる.

注意: 上で期待値を取る操作 $E[\ ]$ の線形性を用いて, 標本平均の期待値が X_i 達の共通の期待値 μ に一致すること

$$E[\bar{X}] = \mu$$

も示されていることに注意せよ. これは標本平均を取る操作が平均の **不偏推定量** を作る操作になっていることも意味している. この点については後で詳しく説明する.

3.2 問題: 対ごとに無相関な確率変数達の標本平均の分散

同じ期待値 μ と分散 σ^2 を持つ対ごとに無相関な n 個の確率変数達 $X_1, \ldots, X_n$ の標本分散 $\bar{X}$ について次が成立していることを示せ:

$$\mathrm{var}(\bar{X}) = \frac{\sigma^2}{n}$$

注意: この問題で扱っている確率変数としての $\bar{X}$ の分散と次の節で説明する標本のモデル化 $X_1, \ldots, X_n$ の分散は異なる数学的対象であることに注意せよ.

注意: $\mathrm{std}(\bar{X}) = \sigma/\sqrt{n}$ は **標準誤差** (standard error)と呼ばれ, SE と書かれることがある. 未知の真の標準誤差の推定量をも標準誤差と呼ぶことがあるが, このノートではそのスタイルを採用せずに, 未知の真の標準誤差の推定量を「標準誤差の推定量」と丁寧に呼ぶことにする.

解答例: 対ごとに無相関な確率変数達の和の分散はそれぞれの確率変数の分散の和になるのであった. X_i 達が対ごとに無相関であるという仮定より, X_i/n 達も対ごとに無相関になる. そして,

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

であることより,

$$\mathrm{var}(\bar{X}) = \sum_{i=1}^{n} \mathrm{var}\left(\frac{X_i}{n}\right) = \sum_{i=1}^{n} \frac{\mathrm{var}(X_i)}{n^2} = \sum_{i=1}^{n} \frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

解答終

3.3 標本分散と不偏分散の定義

n 個の数値 $x_1, \ldots, x_n$ で構成されたデータ(標本, サンプル)の標本平均を

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + \cdots + x_n}{n}$$

と書くとき,

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

を **標本分散** (sample variance, 標本の無補正分散)と呼ぶ. **慣習的には** 補正無し標本分散が使用されることは少なく, 次のように定義される **不偏分散** (unbiased variance, 標本の不偏補正分散)を使うことが多い:

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

どうして n ではなく $n-1$ で割るかについては後で説明する. このとき,

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

とおくと,

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2 = n\overline{x^2} - 2n\bar{x}^2 + n\bar{x}^2 = n\left(\overline{x^2} - \bar{x}^2\right)$$

なので,

$$s^2 = \frac{n}{n-1}\left(\overline{x^2} - \bar{x}^2\right).$$

$X_1, \ldots, X_n$ は標本のモデル化とみなされる n 個の確率変数達であるとする. X_i 達は同じ期待値と分散を持ち, 対ごとに無相関であると仮定し, それらの標本平均を

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + \cdots + X_n}{n}$$

と書く. このとき, $X_1, \ldots, X_n$ の **不偏分散** S^2 を次のように定義する:

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

このとき,

$$\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

とおくと,

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2 = n\overline{X^2} - 2n\bar{X}^2 + n\bar{X}^2 = n\left(\overline{X^2} - \bar{X}^2\right)$$

なので,

$$S^2 = \frac{n}{n-1}\left(\overline{X^2} - \bar{X}^2\right).$$

この場合には, 不偏分散 S^2 は確率変数になることに注意せよ. この S^2 と単なる数値になる $\bar{X}$ の確率変数としての分散 $\mathrm{var}(\bar{X})$ は異なる数学的対象になるので混乱しないようにして欲しい.

3.4 不偏推定量について: 不偏分散の定義ではどうして n ではなく $n-1$ で割るか

$X_1, \dots, X_n$ に共通の期待値と分散をそれぞれ μ, σ^2 と書くことにする. $X_1, \dots, X_n$ を現実で得られる標本のモデル化とみなす. そのとき, 統計モデル内での μ, σ^2 の値は未知の真の平均と分散の値であって欲しい. この設定では未知の μ と σ^2 を標本から推定する方法が欲しくなる.

未知の μ の推定量として優れているのは標本平均である. 標本のモデル化とみなされる $X_1, \dots, X_n$ の標本平均 $\bar{X}$ についてはその期待値 $E[\bar{X}]$ が未知の真の平均値 μ に等しくなる. この性質を標本平均の **不偏性** という. 未知の値 θ の推定量を $E[T] = \theta$ を満たす確率変数 T とするとき, T は θ の **不偏推定量** (unbiased estimator) と呼ぶ.

次に, 不偏分散の定義でどうして n ではなく $n-1$ で割るかを理解するために, $n-1$ で割る前の量の期待値を計算してみよう.

$X_1, \dots, X_n$ 達の標本平均 $\bar{X}$ の期待値と分散がそれぞれ次のようになることはわかっている:

$$E[\bar{X}] = \mu, \quad \mathrm{var}(\bar{X}) = E[(\bar{X} - \mu)^2] = \frac{\sigma^2}{n}.$$

そして, 以下が成立している:

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \bar{X})^2 &= \sum_{i=1}^{n}((X_i - \mu) - (\bar{X} - \mu))^2 \\
&= \sum_{i=1}^{n}\left((X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2)^2\right) \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - 2\left(\sum_{i=1}^{n}(X_i - \mu)\right)(\bar{X} - \mu) + n(\bar{X} - \mu)^2)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - 2n(\bar{X} - \mu)^2 + n(\bar{X} - \mu)^2)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2.
\end{aligned}$$

4つめの等号で $\sum_{i=1}^{n} X_i = n\bar{X}$ を用いた. ゆえに,

$$E\left[\sum_{i=1}^{n}(X_i - \bar{X})^2\right] = \sum_{i=1}^{n} E[(X_i - \mu)^2] - nE[(\bar{X} - \mu)^2] = n\sigma^2 - n\frac{\sigma^2}{n} = (n-1)\sigma^2.$$

これより, σ^2 の不偏推定量を作るためには $\sum_{i=1}^{n}(X_i - \bar{X})^2$ を $n-1$ で割らなければいけないことがわかる. 実際にそのように定義した σ^2 の不偏推定量が不偏分散 S^2 である:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2, \quad E[S^2] = \sigma^2.$$

このように, $\sum_{i=1}^{n}(X_i - \bar{X})^2$ の確率変数としての分散を計算して $(n-1)\sigma^2$ になることを確認すれば, n ではなく $n-1$ で割る理由は明瞭である.

注意: 不偏分散の平方根 $\sqrt{S^2}$ は標準偏差 σ の不偏推定量になって **いない** ことに注意せよ. 推定量の不偏性は座標に依存する.

標準正規分布の場合に不偏分散と補正されていない標本分散の分布を同時プロットしてみよう.

In [35]:

```
n = 10
L = 10^5
X = rand(Normal(), n, L)
S² = var.(eachcol(X))
BV = (n-1)/n * S²
stephist(S²; norm=true, label="unbiased", c=1)
vline!([mean(S²)]; label="E[unbiased]", c=:blue)
stephist!(BV; norm=true, label="biased", c=2)
vline!([mean(BV)]; label="E[biased]", c=:red)
vline!([1]; label="true variance", c=:black, ls=:dash)
```

Out[35]:

確かに不偏分散(unbiased variance)の分布の平均値は真の値に一致しているが, 補正されていない標本分散(biased)の平均値は真の値よりも少し小さくなっている.

さらにこのグラフから, 不偏分散の場合も非補正標本分散の場合も確率密度函数は真の値よりも小さな値で最大になることにも注意せよ.

注意: 標準正規分布のサイズ n の標本分布において, 不偏分散の $n-1$ 倍は自由度 $n-1$ のχ²分布 $\mathrm{Chisq}(n-1)$ に従うことを示せる. 分布 $\mathrm{Chisq}(n-1)$ の期待値は $n-1$ なので, その $n-1$ 分の 1 の期待値は 1 になる.

注意: 標準誤差の推定量 として

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

を採用しよう. このとき,

$$E\left[\widehat{\mathrm{SE}}^2\right] = \frac{E[S^2]}{n} = \frac{\sigma^2}{n} = \mathrm{var}(\bar{X}) = \mathrm{SE}^2$$

なので, $\widehat{\mathrm{SE}}^2$ は標準誤差の2乗 $\mathrm{SE}^2 = \mathrm{var}(\bar{X}) = \sigma^2/n$ の不偏推定量になっている. ただし, $\widehat{\mathrm{SE}}$ 自身は標準誤差 SE の不偏推定量になって **いない** ことに注意せよ. 推定量の不偏性は座標に依存する.

注意: 不偏推定量の方が不偏でない推定量より常に優れていると考えるのは誤りである. 推定量に不偏性の条件を課すと平均二乗誤差が大きくなってしまうことが多い(不偏分散についても実際にそうなることを後で示す). 不偏推定量を使うか否かはそうすることのメリットとデメリットの両方を考慮に入れて決定するべきである. ただし, 上で紹介した不偏分散については害は相当に小さく, 慣習的に非常によく使われている.

注意: $X_1, \ldots, X_n$ は標本のモデル化とみなされる独立同分布な確率変数達であるとし, X_i 達に共通の期待値と分散をそれぞれ μ, σ^2 と書くことにする. このとき, その標本平均 $\bar{X} = (1/n)\sum_{i=1}^{n} X_i$ の期待値と分散がそれぞれ

$$E[\bar{X}] = \mu, \quad \mathrm{var}(\bar{X}) = \frac{\sigma^2}{n}$$

となり, 不偏分散 $S^2 = (1/(n-1))\sum_{i=1}^{n}(X_i - \bar{X})^2$ の期待値が

$$E[S^2] = \sigma^2$$

となることまではすでにわかっている. この一連のノートではほぼ常に与えられた確率変数の期待値と分散を同時に計算して来た. ここでも, 確率変数としての不偏分散の分散も計算するべきである. そして, さらに標本平均と不偏分散の共分散も計算するべきである. しかし, そのためには X_i 達共通の分布の **歪度** (わいど, skewness)と **尖度** (せんど, kurtosis)の定義を先にしておいた方がよい. 次の節で歪度と尖度を定義する.

3.5 データの共分散の定義

n 個の数値の対 $(x_1, y_1), \ldots, (x_n, y_n)$ で構成されたデータ(標本, サンプル)について, x_i, y_i 達各々の標本平均を

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

と書くとき,

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

を **標本共分散** (sample covariance, 標本の無補正共分散)と呼ぶ. **慣習的には** 補正無しの標本共分散が使用されることは少なく, 次のように定義される **不偏共分散** (unbiased variance, 標本の不偏補正共分散)を使うことが多い:

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$$

n ではなく $n-1$ で割る理由は不偏分散の場合と同様である. 後で詳しく説明する. このとき,

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i$$

とおくと,

$$\begin{aligned}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) &= \sum_{i=1}^{n} x_i y_i - \bar{x}\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i\,\bar{y} + n\bar{x}\bar{y} \\ &= n\overline{xy} - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y} = n\left(\overline{xy} - \bar{x}\bar{y}\right)\end{aligned}$$

なので,

$$s_{xy} = \frac{n}{n-1}\left(\overline{xy} - n\bar{x}\bar{y}\right).$$

$(X_1, Y_1), \ldots, (X_n, Y_n)$ は上の型のデータのモデル化とみなされる n 個の確率変数対達であるとする. X_i 達は対ごとに無相関で共通の同じ期待値 μ_X と分散 σ_X^2 を持ち, Y_i 達も対ごとに無相関で共通の同じ期待値 μ_Y と分散 σ_Y^2 を持ち, X_i と Y_i の共分散は i によらない共通の値 σ_{XY} になっているとし, 異なる番号 i, j について X_i と Y_j は無相関であると仮定する. これらの仮定は次のように式で書ける:

$$\begin{aligned}&E[X_i] = \mu_X, \quad E[Y_i] = \mu_Y \\ &E[(X_i - \mu_X)(X_j - \mu_X)] = \delta_{ij}\sigma_X^2, \\ &E[(Y_i - \mu_Y)(Y_j - \mu_Y)] = \delta_{ij}\sigma_Y^2, \\ &E[(X_i - \mu_X)(Y_j - \mu_Y)] = \delta_{ij}\sigma_{XY}.\end{aligned}$$

それらの X_i 達と Y_i 達それぞれの標本平均を次のように書くことにする:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

このとき, $(X_1, Y_1), \ldots, (X_n, Y_n)$ の **不偏共分散** S_{XY} を次のように定義する:

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

このとき,

$$\overline{XY} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i$$

とおくと,

$$\begin{aligned}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) &= \sum_{i=1}^{n} X_i Y_i - \bar{X}\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} X_i\,\bar{Y} + n\bar{X}\bar{Y} \\ &= n\overline{XY} - n\bar{X}\bar{Y} - n\bar{X}\bar{Y} + n\bar{X}\bar{Y} = n\left(\overline{XY} - \bar{X}\bar{Y}\right)\end{aligned}$$

なので,

$$s_{XY} = \frac{n}{n-1}\left(\overline{XY} - n\bar{X}\bar{Y}\right).$$

この場合には, 不偏分散 S_{XY} は確率変数になることに注意せよ. この S_{XY} と単なる数値になる $\bar{X}, \bar{Y}$ の確率変数としての共分散 $\mathrm{cov}(\bar{X}, \bar{Y})$ は異なる数学的対象になるので混乱しないようにして欲しい.

3.6 問題: 標本平均達の共分散

前節の設定のもとで次が成立することを示せ:

$$\mathrm{cov}(\bar{X}, \bar{Y}) = \frac{\sigma_{XY}}{n}.$$

解答例: 標本平均の分散の計算とほぼ同じになる.

番号が異なる確率変数達が対ごとに無相関なときの確率変数達の和の共分散は各々の対の共分散の和になるのであった. これを確率変数の和達

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}, \quad \bar{Y} = \sum_{i=1}^{n} \frac{Y_i}{n}$$

に適用すると,

$$\mathrm{cov}(\bar{X}, \bar{Y}) = \sum_{i=1}^{n} \mathrm{cov}\left(\frac{X_i}{n}, \frac{Y_i}{n}\right) = \sum_{i=1}^{n} \frac{\mathrm{cov}(X_i, Y_i)}{n^2} = \sum_{i=1}^{n} \frac{\sigma_{XY}}{n^2} = \frac{\sigma_{XY}}{n}.$$

解答終

3.7 問題: 不偏共分散の定義ではどうして n ではなく $n-1$ で割るか

前々の設定のもとで, 不偏共分散 S_{XY} が X_i, Y_i の共分散 σ_{XY} の不偏推定量になっていることを示せ. すなわち次が成立することを示せ:

$$E[S_{XY}] = \sigma_{XY}.$$

注意: これが不偏共分散の定義で n ではなく $n-1$ で割る理由である.

解答例: 不偏分散の場合の議論とほぼ同じ.

$\bar{X}, \bar{Y}$ の期待値と共分散がそれぞれ次のようになることはわかっている:

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

そして, 以下が成立している:

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) &= \sum_{i=1}^{n}((X_i - \mu_X) - (\bar{X} - \mu_X))((Y_i - \mu_Y) - (\bar{Y} - \mu_Y)) \\
&= \sum_{i=1}^{n}\Big((X_i - \mu_X)(Y_i - \mu_Y) - (X_i - \mu_X)(\bar{Y} - \mu_Y) \\
&\qquad - (\bar{X} - \mu_X)(Y_i - \mu_Y) + (\bar{X} - \mu_X)(\bar{Y} - \mu_Y)\Big) \\
&= \sum_{i=1}^{n}(X_i - \mu_X)(Y_i - \mu_Y) - \left(\sum_{i=1}^{n}(X_i - \mu_X)\right)(\bar{Y} - \mu_Y) \\
&- (\bar{X} - \mu_X)\left(\sum_{i=1}^{n}(Y_i - \mu_Y)\right) + n(\bar{X} - \mu_X)(\bar{Y} - \mu_Y) \\
&= \sum_{i=1}^{n}(X_i - \mu_X)(Y_i - \mu_Y) - n(\bar{X} - \mu_X)(\bar{Y} - \mu_Y) \\
&- n(\bar{X} - \mu_X)(\bar{Y} - \mu_Y) + n(\bar{X} - \mu_X)(\bar{Y} - \mu_Y) \\
&= \sum_{i=1}^{n}(X_i - \mu_X)(Y_i - \mu_Y) - n(\bar{X} - \mu_X)(\bar{Y} - \mu_Y).
\end{aligned}$$

4つめの等号で $\sum_{i=1}^{n} X_i = n\bar{X}$, $\sum_{i=1}^{n} Y_i = n\bar{Y}$ を用いた. ゆえに,

$$
\begin{aligned}
E\left[\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})\right] &= \sum_{i=1}^{n} E[(X_i-\mu_X)(Y_i-\mu_Y)] - nE[(\bar{X}-\mu_X)(\bar{Y}-\mu_Y)] \\
&= n\sigma_{XY} - n\frac{\sigma_{XY}}{n} = (n-1)\sigma_{XY}.
\end{aligned}
$$

ゆえに

$$
S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})
$$

より

$$
E[S_{XY}] = \sigma_{XY}.
$$

解答終

3.8 データの相関係数の定義 (以上の定義のまとめにもなっている)

データの **相関係数** (correlation coefficient of sample, Pearsonの相関係数)を定義しよう.

数値データの場合の共分散の定義の設定のもとで, x_i, y_i 達の標本平均, 不偏分散達と不偏共分散を以下のように書くことにする:

$$
\begin{aligned}
&\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \\
&s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2, \quad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2, \\
&s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}).
\end{aligned}
$$

s_x^2, s_y^2 の平方根をそれぞれ s_x, s_y と書く. x_i 達と y_i 達の **相関係数** r_{xy} を次のように定める:

$$
r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}}.
$$

この r_{xy} は数値になる.

データのモデル化である確率変数達の共分散の定義の設定のもとで, X_i, Y_i 達の標本平均, 不偏分散達と不偏共分散を以下のように書くことにする:

$$
\begin{aligned}
&\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i, \\
&S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2, \quad S_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2, \\
&S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y}).
\end{aligned}
$$

S_X^2, S_Y^2 の平方根をそれぞれ S_X, S_Y と書く. X_i 達と Y_i 達の **相関係数** R_{XY} を次のように定める:

$$
R_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}}.
$$

この R_{XY} は確率変数になる. (確率変数とは数学的には「その函数の期待値が定義されている変数」のことであり, 直観的には値がランダムに決まる変数だと思える.)

4 最小二乗法による線形回帰

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

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

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

を **残差** と呼ぶ. このとき, **平均二乗残差**

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

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

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

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

これは一次函数によるデータの近似になっているので, **線形回帰** (ordinary least squares, OLS, linear regression)と呼ばれる. この場合には説明変数が x の1つだけなので **単回帰** (simple regression, simple linear regression)と呼ぶこともある. 説明変数が多変数になった場合には **重回帰** (multiple regression)と呼ぶ.

以下ではさらに, **最小平均二乗残差** (もしくは単に **平均二乗残差**)の値を次のように書くことにする:

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

注意: 最小二乗法は説明変数 x と目的変数 y について対称ではないので, それらの変数の立場を逆転して作った回帰直線は元の回帰直線に一致しない. 下の方で紹介するGalton (1886)の回帰直線の節も参照せよ.

4.1 問題: 最小二乗法による線形回帰の公式の導出

データ x_i の標本平均と不偏分散をそれぞれ $\bar{x}, s_x^2$ と書き, y_i の標本平均と不偏分散をそれぞれ $\bar{y}, s_y^2$ と書き, x_i と y_i の不偏共分散を s_{xy} と書くとき, 以上の $\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2$ は次のように表されることを示せ:

$$\hat{\beta} = \frac{s_{xy}}{s_x^2}, \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \quad \hat{\sigma}^2 = \frac{n-1}{n}\frac{s_x^2 s_y^2 - s_{xy}^2}{s_x^2}.$$

4.1.1 解答例1: 単回帰の公式の素朴な導出

解答例1: 複雑な計算をシンプルに記述するために以下のようにおく:

$$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \quad \overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2, \quad \overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i.$$

このとき,

$$s_x^2 = \frac{n}{n-1}\left(\overline{x^2} - \bar{x}^2\right), \quad s_y^2 = \frac{n}{n-1}\left(\overline{y^2} - \bar{y}^2\right), \quad s_{xy} = \frac{n}{n-1}\left(\overline{xy} - \bar{x}\bar{y}\right).$$

平均二乗残差 $f(\alpha, \beta)$ を最小化する $\alpha = \hat{\alpha}, \beta = \hat{\beta}$ は f の導函数を 0 する. ゆえに, $\hat{\alpha}, \hat{\beta}$ は次の方程式の解である:

$$0 = \frac{1}{2}\frac{\partial f}{\partial \alpha} = \frac{1}{n}\sum_{i=1}^{n}(\alpha + \beta x_i - y_i) = \alpha + \bar{x}\beta - \bar{y},$$
$$0 = \frac{1}{2}\frac{\partial f}{\partial \beta} = \frac{1}{n}\sum_{i=1}^{n} x_i(\alpha + \beta x_i - y_i) = \bar{x}\alpha + \overline{x^2}\beta - \overline{xy}.$$

後者から前者に $\bar{x}$ をかけたものを引くと

$$\left(\overline{x^2} - \bar{x}^2\right)\beta = \overline{xy} - \bar{x}\bar{y}.$$

これの解は

$$\hat{\beta} = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{s_{xy}}{s_x^2}.$$

上の方程式の前者より, 次によって $\hat{\alpha}$ も求めることができる:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

このとき

$$
\begin{aligned}
\hat{\sigma}^2 &= f(\hat{\alpha}, \hat{\beta}) = f(\bar{y} - \hat{\beta}\bar{x}, \hat{\beta}) \\
&= \frac{1}{n}\sum_{i=1}^{n}(y_i - (\bar{y} - \hat{\beta}\bar{x} + \beta x_i))^2 = \frac{1}{n}\sum_{i=1}^{n}((y_i - \bar{y}) - \hat{\beta}(x_i - \bar{x}))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left((y_i - \bar{y})^2 - 2\hat{\beta}(x_i - \bar{x})(y_i - \bar{y}) + \hat{\beta}^2(x_i - \bar{x})^2\right) \\
&= \frac{n-1}{n}\left(s_y^2 - 2\hat{\beta}s_{xy} + \hat{\beta}^2 s_x^2\right) = \frac{n-1}{n}\left(s_y^2 - 2\frac{s_{xy}}{s_x^2}s_{xy} + \frac{s_{xy}^2}{(s_x^2)^2}s_x^2\right) \\
&= \frac{n-1}{n}\frac{s_x^2 s_y^2 - s_{xy}^2}{s_x^2}.
\end{aligned}
$$

解答終

注意: 上の解答例ではわざと素朴な方法で計算してみた.

4.1.2 解答例2: データの5つの要約値だけで平均二乗残差を表示

解答例2: 平均二乗残差 $f(\alpha, \beta)$ の和の二乗の内側に $\bar{y} - (\alpha + \beta\bar{x})$ を引いて足して, 適当に整理して, 展開して, 整理すると以下のようになる:

$$
\begin{aligned}
f(\alpha, \beta) &= \frac{1}{n}\sum_{i=1}^{n}(y_i - (\alpha + \beta x_i) - (\bar{y} - (\alpha + \beta\bar{x})) + (\bar{y} - (\alpha + \beta\bar{x})))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}((y_i - \bar{y}) - \beta(x_i - \bar{x}) + (\bar{y} - (\alpha + \beta\bar{x})))^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\Big((y_i - \bar{y})^2 + \beta^2(x_i - \bar{x})^2 + 2\beta(x_i - \bar{x})(y_i - \bar{y}) + (\bar{y} - (\alpha + \beta\bar{x}))^2 \\
&\qquad + 2\underbrace{(y_i - \bar{y})}_{\text{sum}=0}(\bar{y} - \alpha + \beta\bar{x})) - 2\beta\underbrace{(x_i - \bar{x})}_{\text{sum}=0}(\bar{y} - (\alpha + \beta\bar{x}))\Big) \\
&= \frac{n-1}{n}\left(s_y^2 + \beta^2 s_x^2 - 2\beta s_{xy}\right) + (\bar{y} - (\alpha + \beta\bar{x}))^2 \\
&= \frac{n-1}{n}\left(s_x^2\left(\beta - \frac{s_{xy}}{s_x^2}\right)^2 + \frac{s_x^2 s_y^2 - s_{xy}^2}{s_x^2}\right) + (\bar{y} - (\alpha + \beta\bar{x}))^2.
\end{aligned}
$$

これより, 平均二乗残差 $f(\alpha, \beta)$ は, β, α がそれぞれ

$$
\hat{\beta} = \frac{s_{xy}}{s_x^2}, \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}
$$

に等しいときに, 最小値

$$
\hat{\sigma}^2 = \frac{n-1}{n}\frac{s_x^2 s_y^2 - s_{xy}^2}{s_x^2}
$$

になることがわかる.

解答終

注意: 上の計算では最初に, 回帰係数 α, β の函数 $f(\alpha, \beta)$ (平均二乗残差)が $2n$ 個の数値で構成されているデータ $(x_1, y_1), \ldots, (x_n, y_n)$ を要約している標本平均と分散共分散の5つの数値 $\bar{x}, \bar{y}, s_x^2, s_y^2, s_{xy}^2$ だけで記述されることを示した. 特に **回帰直線と最小平均二乗残差の値はそれらの5つの要約値だけで決まってしまう.** データのそれら5つの値への要約が不適切な場合には最小二乗法を使うべきではない.

注意: 以上とはまったく別の **線形代数** における **直交射影** で最小二乗法を理解することもできる. 最小二乗法は数学的には直交射影の理論そのものだと言ってよい. この点に関する詳しい解説は別の機会に行いたい. 以上では最も素朴な最小二乗法の扱い方をした.

注意: 最小二乗法は **正規分布モデルの最尤法の一種** だとみなされる. これも非常に重要な視点である. この点についての解説も別の機会に行いたい. 最小二乗法を与える正規分布モデルの確率密度函数は

$$
\begin{aligned}
p(y_1, \ldots, y_n | x_1, \ldots, x_n, \alpha, \beta, \sigma) &= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (\alpha + \beta x_i))^2\right) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{n}{2\sigma^2}\left[\frac{n-1}{n}\left(s_x^2\left(\beta - \frac{s_{xy}}{s_x^2}\right)^2 + \frac{s_x^2 s_y^2 - s_{xy}^2}{s_x^2}\right) + (\bar{y} - (\alpha + \beta\bar{x}))^2\right]\right)
\end{aligned}
$$

であり, $\bar{x}, \bar{y}, s_x^2, s_y^2, s_{xy}^2$ はこの統計モデルの **十分統計量** になっている. この統計モデルは以下のようにも記述される:

$$y_i - (\alpha + \beta x_i) \sim \mathrm{Norma}(0, \sigma) \quad (i = 1, \dots, n)$$

4.1.3 コンピュータによる最小二乗法の計算例

WolframAlpha (https://www.wolframalpha.com/)では次のように入力すれば単回帰を実行できる.

[実行 (https://www.wolframalpha.com/input?i=linear+fit+%2811%2C22%29%2C+%2812%2C20%29%2C+%2813%2C25%29%2C+%2814%2C+19%29%2C+%2815%2C+24%2
→ 7.35238 + 1.17143 x

```
linear fit (11,22), (12,20), (13,25), (14, 19), (15, 24), (16, 29)
```

不偏分散と不偏共分散も以下のように入力すれば計算できる.

[実行 (https://www.wolframalpha.com/input?i=variance+%2811%2C+12%2C+13%2C+14%2C+15%2C+16%29)] → 3.5

```
variance (11, 12, 13, 14, 15, 16)
```

[実行 (https://www.wolframalpha.com/input?i=covariance+%2811%2C+12%2C+13%2C+14%2C+15%2C+16%29%2C+%2822%2C+20%2C+25%2C+19%2C+24%2C+29%29)]
→ 4.1

```
covariance (11, 12, 13, 14, 15, 16), (22, 20, 25, 19, 24, 29)
```

他にも便利な道具があるので, 自分で色々試してみて研究しておくとよい.

以下のコードはJulia言語 (https://julialang.org/).

In [36]:
```
[ones(6) [11, 12, 13, 14, 15, 16]] \ [22, 20, 25, 19, 24, 29]
```

Out[36]:
```
2-element Vector{Float64}:
 7.352380952380989
 1.171428571428569
```

In [37]:
```
var([11, 12, 13, 14, 15, 16])
```

Out[37]:
```
3.5
```

In [38]:
```
cov([11, 12, 13, 14, 15, 16], [22, 20, 25, 19, 24, 29])
```

Out[38]:
```
4.1
```

不偏分散と不偏共分散については以下のように素朴に計算してもよい. (素朴な計算も一度はやってみるべきである.)

In [39]:
```
n = 6
```

Out[39]:
```
6
```

In [40]:
```
# xの標本平均
a = (11 + 12 + 13 + 14 + 15 + 16)/n
```

Out[40]:
```
13.5
```

In [41]:
```
# xの不偏分散
((11 - a)^2 + (12 - a)^2 + (13 - a)^2 + (14 - a)^2 + (15 - a)^2 + (16 - a)^2)/(n - 1)
```

Out[41]:
```
3.5
```

次の公式を使った方が入力が楽になる:

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

In [42]:
```
(11^2 + 12^2 + 13^2 + 14^2 + 15^2 + 16^2 - n*a^2)/(n - 1) # こちらの方が入力が楽
```
Out[42]: 3.5

In [43]:
```
# yの標本平均
b = (22 + 20 + 25 + 19 + 24 + 29)/n
```
Out[43]: 23.166666666666668

標本分散についても次の公式を使って計算してみよう:

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_u - \bar{y}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right).$$

In [44]:
```
# xとyの不偏共分散
(11*22 + 12*20 + 13*25 + 14*19 + 15*24 + 16*29 - n*a*b)/(n - 1)
```
Out[44]: 4.1

グラフもプロットしてみよう. さらに色々な計算方法も試してみよう.

In [45]:
```
# プロット用函数
function plot_ols(x, y, α̂, β̂; xlim=nothing,
        title="data and regression line", kwargs...)
    a, b = extrema(x)
    isnothing(xlim) && (xlim = (a - 0.1(b-a), b + 0.1(b-a)))
    scatter(x, y; label="", msw=0)
    plot!(x -> α̂ + β̂*x, xlim...; label="", lw=2)
    plot!(; xlabel="x", ylabel="y", msw=0)
    plot!(; size=(400, 300))
    plot!(; title, kwargs...)
end
```
Out[45]: plot_ols (generic function with 1 method)

In [46]:
```
# 「よろしくみなさん」に擬似乱数のシードを設定(テスト用)
Random.seed!(4649373)
```
Out[46]: TaskLocalRNG()

In [47]:

```
# create test data
n = 100
α₀, β₀, σ₀ = 10, 2, 4
x = rand(Normal(10, 2), n)
noise = rand(Normal(0, σ₀), n)
y = @. α₀ + β₀*x + noise
@show α₀ β₀ σ₀
println()

# ordinary least squares
@show x̄ = sum(x[i] for i in 1:n)/n
@show ȳ = sum(y[i] for i in 1:n)/n
@show sx² = sum((x[i] - x̄)^2 for i in 1:n)/(n - 1)
@show sy² = sum((y[i] - ȳ)^2 for i in 1:n)/(n - 1)
@show sxy = sum((x[i] - x̄)*(y[i] - ȳ) for i in 1:n)/(n - 1)
println()
β̂ = sxy/sx²
α̂ = ȳ - β̂*x̄
σ̂ = √((n - 1)/n*(sx²*sy² - sxy^2)/sx²)
@show α̂ β̂ σ̂

plot_ols(x, y, α̂, β̂)
```

```
α₀ = 10
β₀ = 2
σ₀ = 4

x̄ = sum((x[i] for i = 1:n)) / n = 9.930093374143759
ȳ = sum((y[i] for i = 1:n)) / n = 29.76287823674706
sx² = sum(((x[i] - x̄) ^ 2 for i = 1:n)) / (n - 1) = 3.9902128888981885
sy² = sum(((y[i] - ȳ) ^ 2 for i = 1:n)) / (n - 1) = 37.32614282138912
sxy = sum(((x[i] - x̄) * (y[i] - ȳ) for i = 1:n)) / (n - 1) = 8.920847575776838

α̂ = 7.562346117161635
β̂ = 2.235682111246986
σ̂ = 4.148269982085684
```

Out[47]:

不偏分散と不偏共分散は次のようなコードでも計算できる.

In [48]:

```
@show sx² = var(x)
@show sy² = var(y)
@show sxy = cov(x, y);
```

```
sx² = var(x) = 3.990212888898188
sy² = var(y) = 37.3261428213891
sxy = cov(x, y) = 8.920847575776842
```

以下のように線形代数的な実装にすればシンプルなコードになる.

In [49]:

```
# ordinary least squares
X = [ones(n) x] # design matrix
α̂, β̂ = X \ y
σ̂ = norm(y - (α̂ .+ β̂*x))/√n
@show α̂ β̂ σ̂

plot_ols(x, y, α̂, β̂)
```

```
α̂ = 7.562346117161626
β̂ = 2.235682111246988
σ̂ = 4.148269982085681
```

Out[49]:

正規分布モデルの最尤法でも同じことをできる.

In [50]:

```
# negative log likelihood
function f(x, y, α, β, t)
    -sum(logpdf(Normal(0, exp(t)), y[i] - (α + β*x[i])) for i in eachindex(x))
end

# minimize negative log likelihood = maximize likelihood
o = optimize(((α, β, t),) -> f(x, y, α, β, t), zeros(3), LBFGS())
α̂, β̂, t̂ = o.minimizer
σ̂ = exp(t̂)
@show α̂ β̂ σ̂

plot_ols(x, y, α̂, β̂)
```

```
α̂ = 7.562346115104934
β̂ = 2.2356821114394743
σ̂ = 4.148269982255
```

Out[50]:

4.2 必修問題: 最小二乗法の計算例 (Anscombe's quartet)

以下に示した $(x_1, y_1), \ldots, (x_n, y_n)$ 型の4つのデータのそれぞれについて,

- x_i 達の標本平均 $\bar{x}$
- y_i 達の標本平均 $\bar{y}$
- x_i 達の不偏分散 s_x^2,
- y_i 達の不偏分散 s_y^2,
- x_i 達と y_i 達の不偏共分散 s_{xy},
- 回帰直線 $y = \hat{\alpha} + \hat{\beta}x$ の係数 $\hat{\alpha}, \hat{\beta}$ と最小平均二乗残差の平方根 $\hat{\sigma}$

を求めよ. それぞれについて小数点以下第1桁まで求めよ.

1. (10, 8.04), (8, 6.95), (13, 7.58), (9, 8.81), (11, 8.33), (14, 9.96), (6, 7.24), (4, 4.26), (12, 10.84), (7, 4.82), (5, 5.68)
2. (10, 9.14), (8, 8.14), (13, 8.74), (9, 8.77), (11, 9.26), (14, 8.1), (6, 6.13), (4, 3.1), (12, 9.13), (7, 7.26), (5, 4.74)
3. (10, 7.46), (8, 6.77), (13, 12.74), (9, 7.11), (11, 7.81), (14, 8.84), (6, 6.08), (4, 5.39), (12, 8.15), (7, 6.42), (5, 5.73)
4. (8, 6.58), (8, 5.76), (8, 7.71), (8, 8.84), (8, 8.47), (8, 7.04), (8, 5.25), (19, 12.5), (8, 5.56), (8, 7.91), (8, 6.89)

注意: 実はこれは有名なアンスコムの例 (https://ja.wikipedia.org/wiki/%E3%82%A2%E3%83%B3%E3%82%B9%E3%82%B3%E3%83%A0%E3%81%AE%E4%BE%8B)である. この例の教訓はデータの散布図をプロットすることの重要性である. 回帰直線を求める統計分析は2,3,4のデータについては適切ではない.

コピー&ペースト用: Anscombe's quartet

データ1

- x: 10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5
- y: 8.04, 6.95, 7.58, 8.81, 8.33, 9.96, 7.24, 4.26, 10.84, 4.82, 5.68

データ2

- x: 10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5
- y: 9.14, 8.14, 8.74, 8.77, 9.26, 8.1, 6.13, 3.1, 9.13, 7.26, 4.74

データ3

- x: 10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5
- y: 7.46, 6.77, 12.74, 7.11, 7.81, 8.84, 6.08, 5.39, 8.15, 6.42, 5.73

データ4

- x: 8, 8, 8, 8, 8, 8, 8, 19, 8, 8, 8
- y: 6.58, 5.76, 7.71, 8.84, 8.47, 7.04, 5.25, 12.5, 5.56, 7.91, 6.89

解答例:

データ1:

$$\begin{aligned} &\bar{x} \approx 9.0, \quad \bar{y} \approx 7.5, \\ &s_x^2 \approx 11.0, \quad s_y^2 \approx 4.1, \quad s_{xy} \approx 5.5 \\ &\hat{\alpha} \approx 3.0, \quad \hat{\beta} \approx 0.5, \quad \hat{\sigma} \approx 1.1 \end{aligned}$$

データ2:

$$\begin{aligned} &\bar{x} \approx 9.0, \quad \bar{y} \approx 7.5, \\ &s_x^2 \approx 11.0, \quad s_y^2 \approx 4.1, \quad s_{xy} \approx 5.5 \\ &\hat{\alpha} \approx 3.0, \quad \hat{\beta} \approx 0.5, \quad \hat{\sigma} \approx 1.1 \end{aligned}$$

データ3:

$$\begin{aligned} &\bar{x} \approx 9.0, \quad \bar{y} \approx 7.5, \\ &s_x^2 \approx 11.0, \quad s_y^2 \approx 4.1, \quad s_{xy} \approx 5.5 \\ &\hat{\alpha} \approx 3.0, \quad \hat{\beta} \approx 0.5, \quad \hat{\sigma} \approx 1.1 \end{aligned}$$

データ4:

$$
\begin{aligned}
&\bar{x} \approx 9.0, \quad \bar{y} \approx 7.5, \\
&s_x^2 \approx 11.0, \quad s_y^2 \approx 4.1, \quad s_{xy} \approx 5.5 \\
&\hat{\alpha} \approx 3.0, \quad \hat{\beta} \approx 0.5, \quad \hat{\sigma} \approx 1.1
\end{aligned}
$$

解答終

注意: 小数点以下第1桁までの数値は4つのデータについてすべて一致している! データの散布図を描くと, 回帰直線を描くことが適切なのは最初のデータだけで, 残りの3つのデータについては不適切なことがわかる. グラフは下の方にある.

4.2.1 WaolframAlphaでAnscombeの例1を扱う方法

データ1に関するWolframAlphaの使用例: xの標本平均 (https://www.wolframalpha.com/input?i=mean+10%2C+8%2C+13%2C+9%2C+11%2C+14%2C+6%2C+4%2C+12%2C+7%2C+5&lang=ja), yの標本平均 (https://www.wolframalpha.com/input?i=mean+8.04%2C+6.95%2C+7.58%2C+8.81%2C+8.33%2C+9.96%2C+7.24%2C+4.26%2C+10.84%2C+4.82%2C+5.68), xの不偏分散 (https://www.wolframalpha.com/input?i=var+10%2C+8%2C+13%2C+9%2C+11%2C+14%2C+6%2C+4%2C+12%2C+7%2C+5), yの不偏分散 (https://www.wolframalpha.com/input?i=var+8.04%2C+6.95%2C+7.58%2C+8.81%2C+8.33%2C+9.96%2C+7.24%2C+4.26%2C+10.84%2C+4.82%2C+5.68), x,yの不偏共分散 (https://www.wolframalpha.com/input?i=covar+%2810%2C+8%2C+13%2C+9%2C+11%2C+14%2C+6%2C+4%2C+12%2C+7%2C+5%29%2C+%288.04%2C+6.95%2C+7
回帰直線 (https://www.wolframalpha.com/input?i=linear+fit+%2810%2C+8.04%29%2C+%288%2C+6.95%29%2C+%2813%2C+7.58%29%2C+%289%2C+8.81%29%2C+%2811%2
最小平均二乗残差の平方根 (https://www.wolframalpha.com/input?i=sqrt%28%28n-1%29%2Fn*%28a*c+-+b%5E2%29%2Fa%29+where+n%3D11%2C+a%3D11%2C+b%3D5.5%2C+c%3D4.127)

4.2.2 Julia言語でAnscombeの例を扱う方法

RDatatsets.jl パッケージを入れればその中にAnscombeの例が入っている.

In [51]:
```
using RDatasets
anscombe = dataset("datasets", "anscombe")
```

Out[51]: 11 rows × 8 columns

	X1	X2	X3	X4	Y1	Y2	Y3	Y4
	Int64	Int64	Int64	Int64	Float64	Float64	Float64	Float64
1	10	10	10	8	8.04	9.14	7.46	6.58
2	8	8	8	8	6.95	8.14	6.77	5.76
3	13	13	13	8	7.58	8.74	12.74	7.71
4	9	9	9	8	8.81	8.77	7.11	8.84
5	11	11	11	8	8.33	9.26	7.81	8.47
6	14	14	14	8	9.96	8.1	8.84	7.04
7	6	6	6	8	7.24	6.13	6.08	5.25
8	4	4	4	19	4.26	3.1	5.39	12.5
9	12	12	12	8	10.84	9.13	8.15	5.56
10	7	7	7	8	4.82	7.26	6.42	7.91
11	5	5	5	8	5.68	4.74	5.73	6.89

In [52]:
```
# コピー＆ペーストに利用するための表示
data1 = collect(zip(anscombe.X1, anscombe.Y1))
data2 = collect(zip(anscombe.X2, anscombe.Y2))
data3 = collect(zip(anscombe.X3, anscombe.Y3))
data4 = collect(zip(anscombe.X4, anscombe.Y4))
@show data1 data2 data3 data4;
```

```
data1 = [(10, 8.04), (8, 6.95), (13, 7.58), (9, 8.81), (11, 8.33), (14, 9.96), (6, 7.24), (4, 4.
26), (12, 10.84), (7, 4.82), (5, 5.68)]
data2 = [(10, 9.14), (8, 8.14), (13, 8.74), (9, 8.77), (11, 9.26), (14, 8.1), (6, 6.13), (4, 3.
1), (12, 9.13), (7, 7.26), (5, 4.74)]
data3 = [(10, 7.46), (8, 6.77), (13, 12.74), (9, 7.11), (11, 7.81), (14, 8.84), (6, 6.08), (4,
5.39), (12, 8.15), (7, 6.42), (5, 5.73)]
data4 = [(8, 6.58), (8, 5.76), (8, 7.71), (8, 8.84), (8, 8.47), (8, 7.04), (8, 5.25), (19, 12.
5), (8, 5.56), (8, 7.91), (8, 6.89)]
```

In [53]:
```
# コピー＆ペーストに利用するための表示
@show anscombe.X1 anscombe.Y1
@show anscombe.X2 anscombe.Y2
@show anscombe.X3 anscombe.Y3
@show anscombe.X4 anscombe.Y4;
```

```
anscombe.X1 = [10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5]
anscombe.Y1 = [8.04, 6.95, 7.58, 8.81, 8.33, 9.96, 7.24, 4.26, 10.84, 4.82, 5.68]
anscombe.X2 = [10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5]
anscombe.Y2 = [9.14, 8.14, 8.74, 8.77, 9.26, 8.1, 6.13, 3.1, 9.13, 7.26, 4.74]
anscombe.X3 = [10, 8, 13, 9, 11, 14, 6, 4, 12, 7, 5]
anscombe.Y3 = [7.46, 6.77, 12.74, 7.11, 7.81, 8.84, 6.08, 5.39, 8.15, 6.42, 5.73]
anscombe.X4 = [8, 8, 8, 8, 8, 8, 8, 19, 8, 8, 8]
anscombe.Y4 = [6.58, 5.76, 7.71, 8.84, 8.47, 7.04, 5.25, 12.5, 5.56, 7.91, 6.89]
```

In [54]:
```
function solve_anscombe(x, y;
        title="", xlim=(3, 20), ylim=(2, 14), size=(250, 225))
    @show x̄ = mean(x)
    @show ȳ = mean(y)
    @show sx² = var(x)
    @show sy² = var(y)
    @show sxy = cov(x, y)
    β̂ = sxy/sx²
    α̂ = ȳ - β̂*x̄
    σ̂ = √((n - 1)/n*(sx²*sy² - sxy^2)/sx²)
    @show α̂ β̂ σ̂

    plot_ols(x, y, α̂, β̂; title, xlim, ylim, size)
end
```

Out[54]:
```
solve_anscombe (generic function with 1 method)
```

In [55]:
```
PA1 = solve_anscombe(anscombe.X1, anscombe.Y1, title="Anscombe 1")
```

```
x̄ = mean(x) = 9.0
ȳ = mean(y) = 7.500909090909093
sx² = var(x) = 11.0
sy² = var(y) = 4.127269090909091
sxy = cov(x, y) = 5.501
α̂ = 3.0000909090909103
β̂ = 0.5000909090909091
σ̂ = 1.1672644558967773
```

Out[55]:

In [56]:
```
PA2 = solve_anscombe(anscombe.X2, anscombe.Y2, title="Anscombe 2")
```

```
x̄ = mean(x) = 9.0
ȳ = mean(y) = 7.500909090909091
sx² = var(x) = 11.0
sy² = var(y) = 4.127629090909091
sxy = cov(x, y) = 5.5
α̂ = 3.000909090909091
β̂ = 0.5
σ̂ = 1.167841085079644
```

Out[56]:

In [57]:
```
PA3 = solve_anscombe(anscombe.X3, anscombe.Y3, title="Anscombe 3")
```

```
x̄ = mean(x) = 9.0
ȳ = mean(y) = 7.500000000000001
sx² = var(x) = 11.0
sy² = var(y) = 4.12262
sxy = cov(x, y) = 5.497
α̂ = 3.0024545454545466
β̂ = 0.49972727272727274
σ̂ = 1.1669888559879227
```

Out[57]:

In [58]:
```
PA4 = solve_anscombe(anscombe.X4, anscombe.Y4, title="Anscombe 4")
```

```
x̄ = mean(x) = 9.0
ȳ = mean(y) = 7.50090909090909
sx² = var(x) = 11.0
sy² = var(y) = 4.12324909090909
sxy = cov(x, y) = 5.499
α̂ = 3.0017272727272726
β̂ = 0.49990909090909086
σ̂ = 1.1664075231238866
```

Out[58]:

In [59]:
```
plot(PA1, PA2, PA3, PA4; size=(500, 450), layout=(2,2))
```

Out[59]:

アンスコムの例では1以外のデータについて単純に最小二乗法で回帰直線を引くことは不適切である.

さらに, データを要約せずにその全体の様子を見るためにグラフを描いてみるとことが非常に大事なこともこの例からわかる.

4.3 Galton (1886)が描いた2つの回帰直線

下の方にある図は次の論文の5ページ目からの転載である:

- Galton, Francis. Regression towards mediocrity in hereditary stature. Journal of the Anthropological Institute, Vol. 5 (1886), 246-263. [link (https://galton.org/bib/JournalItem.aspx_action=view_id=157)]

この論文の著者の Francis Galton (https://en.wikipedia.org/wiki/Francis_Galton) (フランシス・ゴルトン (https://ja.wikipedia.org/wiki/%E3%83%95%E3%83%A9%E3%83%B3%E3%82%B7%E3%82%B9%E3%83%BB%E3%82%B4%E3%
は「平均への回帰」の記述や「相関係数」の導入で有名である.

このグラフにおける縦軸は親の身長の「平均値」であり, 横軸は成人した子供の身長である(単位はインチ). 親の身長の「平均値」の正確な意味は論文の2ページ目で以下のように説明されている:

> My data consisted of th e heights of 930 adult children and of their respective parentages, 205 in number. In every case I transmuted the female statures to their corresponding male equivalents and used them in their transmuted form, so that no objection grounded on the sexual difference of stature need be raised when I speak of averages. The factor I used was 1.08, which is equivalent to adding a little less than one-twelfth to each female height.

翻訳: 私のデータは930人の成人した子供とその親205人の身長である. どの場合でも, 女性の身長を対応する男性の身長に変換し, 変換された状態で使用した. それによって, 平均値について語るときに, 身長の性差に基づく異論を唱える必要を無くした. 使用した係数は1.08であり, これは各々の女性の身長にその12分の1弱を足すことに相当する.

上のグラフに描かれた楕円は, データの分布が多変量正規分布に近いことを表していると考えてよい. さらに, 親と子の身長のあいだに相関関係があることがわかる.

直線 ON が説明変数を親の身長, 目的変数を子の身長としたときの回帰直線であり, 直線 OM は2つの変数の立場を逆転させて作った回帰直線である. **2つの回帰直線は互いに一致しないし, 楕円の長軸とも一致しない.**

点 O はデータの平均値(数式で書くと $(\bar{x}, \bar{y})$ もしくは $(\bar{y}, \bar{x})$)であり, 回帰直線 ON が通る点 N は楕円と横軸に平行な直線が接する点になっており, 回帰直線 OM が通る点 M は楕円と縦軸に平行な直線が接する点になっている. (そうなる理由を自分で考えることは数学に関するちょっとした練習になるので, 興味がある人は自分で考えてみるとよい. 少し下の方に解答例がある.)

4.3.1 Galtonのグラフの再現

以下のセルのデータは上の論文の3ページ目にある Table I である. (数値の入力間違いが残っている可能性があるので, 原論文を参照して各自確認すること. 私自身は入力間違いはないと思っているが, それは私がそう思っているだけに過ぎない. このようなことはよくある.)

In [60]:

```
galton = [
    0  0  0  0  0  0  0  0  0  0  0  1  3  0
    0  0  0  0  0  0  0  1  2  1  2  7  2  4
    0  0  0  0  1  3  4  3  5 10  4  9  2  2
    1  0  1  0  1  1  3 12 18 14  7  4  3  3
    0  0  1 16  4 17 27 20 33 25 20 11  4  5
    1  0  7 11 16 25 31 34 48 21 18  4  3  0
    0  3  5 14 15 36 38 28 38 19 11  4  0  0
    0  3  3  5  2 17 17 14 13  4  0  0  0  0
    1  0  9  5  7 11 11  7  7  5  2  1  0  0
    1  1  4  4  1  5  5  0  2  0  0  0  0  0
    1  0  2  4  1  2  2  1  1  0  0  0  0  0
]
# 上の行列の縦軸は親の身長であり，行列の下ほど身長は小さくなる．
# 各種函数の仕様に合わせるために，行列の縦方向を逆転させて扱う．
galton = reverse(galton; dims=1)
# 行列の横軸が子の身長(単位はインチ)
galton_x = 61.2:74.2
# 行列の縦軸が親の身長(単位はインチ)
galton_y = 63.5:73.5

# データのヒートマップによる可視化
heatmap(galton_x, galton_y, galton; size=(320, 220))
```

Out[60]:

以下のセルでは楕円の長軸と短軸と2本の回帰直線をプロットしている.

In [61]:

```
# 適当にランダマイズしてデータを作成
Random.seed!(4649373)
data = hcat(([x-0.5+rand(), y-0.5+rand()]
        for (i, y) in enumerate(galton_y)
            for (j, x) in enumerate(galton_x)
                for _ in 1:galton[i,j])...)

# 多変量正規分布の最尤法
mvnormal = fit_mle(MvNormal, data)

# 楕円の長軸と短軸を求める
μ = mean(mvnormal) # 楕円の中心 [x̄, ȳ]
Σ = cov(mvnormal)  # 分散共分散行列 ((n-2)/n)[sx² sxy; sxy sy²]
d, U = eigen(Σ)    # U[:,j] = Σの固有値 d[j] の固有ベクトル
a = [-U[1,1]/U[2,1], -U[1,2]/U[2,2]] # 長軸・短軸の傾き
@show Σ μ a
g1(x) = a[1]*(x - μ[1]) + μ[2] # 長軸
g2(x) = a[2]*(x - μ[1]) + μ[2] # 短軸

# 最小二乗法による線形回帰
x, y = data[1,:], data[2,:]
X = [x.^0 x]
Y = [y.^0 y]
@show β = X\y # OLS(x, y)の回帰係数
@show γ = Y\x # OLS(y, x)の回帰係数
f1(x) = β[1] + β[2]*x # OLS(x, y)の回帰直線
f2(y) = γ[1] + γ[2]*y # OLS(y, x)の回帰直線

# グラフをプロット
widenlim((a, b)) = (c = 0.05(b-a); (a-c, b+c))
xlim = widenlim(extrema(galton_x))
ylim = widenlim(extrema(galton_y))
xs = range(xlim..., 1000)
ys = range(ylim..., 1000)

plot(legend=:topleft)
scatter!(x, y; label="data", msw=0, alpha=0.7)
contour!(xs, ys, (x,y) → pdf(mvnormal, [x,y]); colorbar=false, ls=:dot)
plot!(xs, g1.(xs); label="major axis", color=:blue)
plot!(xs, g2.(xs); label="minor axis", color=:blue, ls=:dash)
plot!(f2.(ys), ys; label="OLS(y, x)", color=:red)
plot!(xs, f1.(xs); label="OLS(x, y)", color=:red, ls=:dash)
plot!(; xlim, ylim, aspectratio=1, xtick=0:100, ytick=0:100)
plot!(; size=(560, 440))
title!("Ellipse axes and regression lines of Galton (1886)")
```

```
Σ = [6.496979569005935 2.0836915591542615; 2.0836915591542615 3.350202984520065]
μ = [68.09572980291252, 68.30018786422937]
a = [0.4979680162348105, -2.008161101512316]
β = X \ y = [46.460732175978194, 0.320716963478627]
γ = Y \ x = [25.61575900743164, 0.6219597943116172]
```

Out[61]:

注意・警告: 2つの回帰直線は楕円の長軸に一致しない. この点を誤解している人もいるようなので注意すること. 目分量で大雑把に回帰直線を手で引くときには, 楕円の長軸に線を引きたくなるかもしれないが, 誤りになる. 回帰直線は楕円と横軸または縦軸に平行な直線が接する点を通る直線になる.

4.3.2 グラフ中の楕円と2変量正規分布モデルや回帰直線の関係

上の計算での多変量正規分布の最尤法では, 次の確率密度函数で定まる2変量正規分布の標本分布を統計モデルとして採用している:

$$p(x, y|\mu, \Sigma) = \frac{1}{\sqrt{\det(2\pi\Sigma)}}\exp\left(-\frac{1}{2}\left[\,x-\mu_x,\ y-\mu_y\,\right]\Sigma^{-1}\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix}\right).$$

ここで $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} \in \mathbb{R}^2$ で Σ は固有値が正の 2×2 の実対称行列である. 独立なパラメータは5つある. この統計モデルのデータ $(x_1, y_1), \ldots, (x_n, y_n)$ に関する最尤法はその5つのパラメータを次のように決定する:

$$\hat{\mu} = \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix}, \quad \hat{\Sigma} = \frac{n-1}{n}\begin{bmatrix} s_x^2 & s_{xy} \\ s_{xy} & s_y^2 \end{bmatrix}.$$

単回帰の結果を決定したデータの5つの要約値 $\bar{x}, \bar{y}, s_x^2, s_{xy}, s_y^2$ が2変量正規分布モデルの最尤法の解も決定する.

$\hat{\Sigma}$ の式の右辺の $(n-1)/n$ は分散と共分散の推定値として不偏分散と不偏共分散を採用したので必要になった因子である. 分散と共分散の推定値として $n-1$ ではなく n で割った方を採用すればその因子はなくなる. このように正規分布モデルの最尤法の解の公式と不偏分散・不偏共分散は相性は悪い. 慣習的に使用されている不偏分散と不偏共分散にはこのようなトレードオフもある. (どのような場合であっても都合のよい流儀は存在しない.)

以上の統計モデルと楕円の関係について説明しよう.

分散共分散行列 Σ を $\Sigma = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ と書くと,

$$\Sigma^{-1} = \frac{1}{ac-b^2}\begin{bmatrix} c & -b \\ -b & a \end{bmatrix}$$

なので

$$\left[\,x-\mu_x,\ y-\mu_y\,\right]\Sigma^{-1}\begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix} = \frac{c(x-\mu_x)^2 - 2b(x-\mu_x)(y-\mu_y) +}{ac-b^2}.$$

前節で描いたグラフ中の楕円はこれが一定の値になるという条件で定義された図形である.

以下, 計算を簡単に記述するために $\mu_x = \mu_y = 0$ と仮定する. このとき, その楕円の方程式は

$$ay^2 - 2bxy + cx^2 = \text{constant}.$$

になる. この条件を保ったまま (x, y) を微小に動かすと,

$$2ay\,dy - 2bx\,dy - 2by\,dx + 2cx\,dx = 0$$

すなわち

$$(ay - bx)\,dy - (by - cx)\,dx = 0$$

が得られる(これは点 (x, y) を通る楕円の接線の方程式だとみなされる).

点 (x, y) が楕円と y 軸と平行な直線の接点のとき, そのような接点 (x, y) を楕円上に乗せたままで微小に動かすと x 座標は高次の微小量しか変化しないので $dx = 0$ が得られ, $y = (b/a)x$ が得られる. これは, $\Sigma = \hat{\Sigma}$ の場合には $a = ((n-1)/n)s_x^2$, $b = ((n-1)/n)s_{xy}$ なので $y = (s_{xy}/s_x^2)x$ を意味する. その式で定義される直線の傾き s_{xy}/s_x^2 は説明変数が x で目的変数が y のときの単回帰直線の傾き $\bar{\beta}$ に一致している.

これで, 説明変数が x で目的変数が y のときの単回帰直線が楕円と y 軸に平行な直線の接点を通ることも証明された.

注意: このことから, 楕円の内側を通る y 軸に平行な直線 ℓ について, 説明変数が x で目的変数が y のときの単回帰直線と直線 ℓ の交点と楕円と直線 ℓ の2つの交点の距離が等しくなることもわかる. なぜならば, 変換 $(x, y) \mapsto (x, y - (b/a)x)$ によって, 回帰直線 $y = (b/a)x$ が楕円と y 軸に平行な直線の接点を通るという条件も保たれるので($\mu_x = \mu_y = 0$ と仮定している), その変換でうつした先の楕円の軸の片方は x 軸に一致し, うつした先の楕円が x 軸について線対称になることがわかるからである. 下の方で描いたグラフを見よ.

注意: 以上の数学的議論を追い切れない人は前節で描いたグラフやGolton氏自身が描いたグラフをよく見て, 単回帰直線が楕円と横軸(もしくは縦軸)と平行な直線の接点を通っていることを目で確認せよ. 数学的な議論が追い切れない場合であっても, 結論をグラフを見て確認できるならば, 記憶に残って役に立つ知識になることが多い.

4.3.3 楕円と回帰直線の関係の視覚化

In [62]:

```
f(x, y, h) = 6y^2 - 4x*y + 3x^2 - h
g(x, h, s) = (2x + s*√(4x^2 - 6*(3x^2 - h)))/6
h = f(1.5, 0.5, 0)
x = range(-2, 2, 200)
y = range(-1.5, 1.5, 200)
plot(; aspectratio=1, size=(400, 300))
plot!(x, x -> x/3; label="regression line", legend=:topleft)
vline!([-1.5, 1.5]; label="", ls=:dash)
k = -0.4
plot!([k, k], [g(k, h, -1), g(k, h, 1)]; label="", c=3)
plot!([k-0.05, k+0.05], fill((k/3 + g(k, h, -1))/2, 2); label="", c=3)
plot!([k-0.05, k+0.05], fill((k/3 + g(k, h,  1))/2, 2); label="", c=3)
k = 0.8
plot!([k, k], [g(k, h, -1), g(k, h, 1)]; label="", c=4)
plot!([k-0.05, k+0.05], fill((k/3 + g(k, h, -1))/2 + 0.02, 2); label="", c=4)
plot!([k-0.05, k+0.05], fill((k/3 + g(k, h, -1))/2 - 0.02, 2); label="", c=4)
plot!([k-0.05, k+0.05], fill((k/3 + g(k, h,  1))/2 + 0.02, 2); label="", c=4)
plot!([k-0.05, k+0.05], fill((k/3 + g(k, h,  1))/2 - 0.02, 2); label="", c=4)
contour!(x, y, (x, y) -> f(x, y, h); colorbar=false, levels=[0])
plot!(; xlabel="x", ylabel="y")
```

Out[62]:

確かに, 楕円と縦軸に平行な直線の接点を通る直線(=回帰直線)が楕円を上下の長さについて二等分していることがこの図を見ればわかる.

In [63]:

```
f(x, y, h) = 6y^2 - 4x*y + 3x^2 - h
g(x, h, s) = (2x + s*√(4x^2 - 6*(3x^2 - h)))/6
h = f(1.5, 0.5, 0)
x = range(-2, 2, 200)
y = range(-1.5, 1.5, 200)
anim = @animate for a in [reverse(range(-1, 1/3, 40)); range(-1, 1/3, 40)]
    plot(; aspectratio=1, size=(400, 320))
    plot!(x, x -> x/3+a*x; label="", legend=:topleft)
    vline!([-1.5, 1.5]; label="", ls=:dash)
    k = -0.4
    plot!([k, k], [g(k, h, -1)+a*k, g(k, h, 1)+a*k]; label="", c=3)
    plot!([k-0.05, k+0.05], fill((k/3 + g(k, h, -1))/2+a*k, 2); label="", c=3)
    plot!([k-0.05, k+0.05], fill((k/3 + g(k, h,  1))/2+a*k, 2); label="", c=3)
    k = 0.8
    plot!([k, k], [g(k, h, -1)+a*k, g(k, h, 1)+a*k]; label="", c=4)
    plot!([k-0.05, k+0.05], fill((k/3 + g(k, h, -1))/2+a*k + 0.02, 2); label="", c=4)
    plot!([k-0.05, k+0.05], fill((k/3 + g(k, h, -1))/2+a*k - 0.02, 2); label="", c=4)
    plot!([k-0.05, k+0.05], fill((k/3 + g(k, h,  1))/2+a*k + 0.02, 2); label="", c=4)
    plot!([k-0.05, k+0.05], fill((k/3 + g(k, h,  1))/2+a*k - 0.02, 2); label="", c=4)
    contour!(x, y, (x, y) -> f(x, y-a*x, h); colorbar=false, levels=[0])
    plot!(; xlabel="x", ylabel="y")
    plot!(; xtick=-2:0.5:2, ytick=-1.5:0.5:1.5)
    title!("ellipse and regression line")
end
gif(anim, "images/ellipse_and_regressionline.gif")
```

```
┌ Info: Saved animation to
│   fn = D:\OneDrive\work\Statistics\2022\images\ellipse_and_regressionline.gif
└ @ Plots D:\.julia\packages\Plots\1KWPG\src\animation.jl:114
```

Out[63]:

5 モーメントとキュムラントと歪度と尖度

5.1 モーメントとその母函数と特性函数とキュムラント母函数の定義

確率変数 X と $m = 0, 1, 2, \ldots$ 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 **モーメント母函数** (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函数** と呼ぶ. 以下も同様である.

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

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

を **特性函数** (characteristic function)と呼ぶ. 特性函数を扱う場合には $i = \sqrt{-1}$ 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.

モーメント母函数だけではなく, 特性函数もモーメント達の母函数になっている.

モーメント母函数の対数

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

を **キュムラント母函数** (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, 特性函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. モーメント母函数やキュムラント母函数の取り扱いではこの点に注意する必要がある.

注意: モーメント母函数とキュムラント母函数はそれぞれ物理での統計力学における **分配函数** 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, モーメント母函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類似物になっていることになる.

5.2 特性函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 が従う分布が唯一つに決まる** 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.

モーメント母函数 $M_X(t) = E[e^{tX}]$ 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. キュムラント母函数はモーメント母函数の対数である. これらより, **モーメント母函数やキュムラント母函数からも分布が唯一つに決まる** ことがわかる.

5.3 問題: 分布のアフィン変換のキュムラント

確率変数 X と $a, b \in \mathbb{R}$, $a \neq 0$ について, $aX + b$ のモーメント母函数とキュムラント母函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:

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

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. モーメント母函数の対数を取ってキュムラント母函数を定義し, その展開によってキュムラントを定義することにはこのような利点がある. キュムラント母函数を定義することには, 物理の統計力学で分配函数の対数を取って自由エネルギーを定義することと同様の利点がある.

注意: この結果は空気のごとく使われる.

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

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

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函数は $K_X(t) \mapsto K_X(at) + bt$ 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函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

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

になるので,

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

となることがわかる.

解答終

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

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

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

この結果は中心極限定理の証明で使われる.

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

$$tz - \frac{z^2}{2} = -\frac{z^2 - 2tz}{2} = -\frac{(z-t)^2 - t^2}{2} = -\frac{(z-t)^2}{2} + \frac{t^2}{2}$$

なので,

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

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函数は $t^2/2$ というたったの一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函数は 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,

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

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

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

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函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.

5.5 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およびそれらの母函数をそれぞれ **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 $\bar{\mu}_m(X)$, $\bar{\kappa}_m(X)$ と表す. 詳しくは以下の通り:

$$
\begin{aligned}
&\bar{\mu}_m(X) = \mu_m(Z) = E\left[\left(\frac{X-\mu}{\sigma}\right)^m\right], \\
&M_Z(t) = E\left[\exp\left(t\frac{X-\mu}{\sigma}\right)\right] = \sum_{m=0}^{\infty}\bar{\mu}_m(X)\frac{t^m}{m!} = 1 + \frac{t^2}{2} + \bar{\mu}_3(X)\frac{t^3}{3!} + \bar{\mu}_4(X)\frac{t^4}{4!} + \cdots, \\
&K_Z(t) = \log M_Z(t) = \sum_{m=1}^{\infty}\bar{\kappa}_m(X)\frac{t^m}{m!} = \frac{t^2}{2} + \bar{\kappa}_3(X)\frac{t^3}{3!} + \bar{\kappa}_4(X)\frac{t^4}{4!} + \cdots.
\end{aligned}
$$

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

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

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:

$$
\begin{aligned}
\log\left(1 + \frac{t^2}{2} + \bar{\mu}_3(X)\frac{t^3}{3!} + \bar{\mu}_4(X)\frac{t^4}{4!} + O(t^5)\right) &= \frac{t^2}{2} + \bar{\mu}_3(X)\frac{t^3}{3!} + \bar{\mu}_4(X)\frac{t^4}{4!} - \frac{1}{2}\left(\frac{t^2}{2}\right)^2 + O(t^5) \\
&= \frac{t^2}{2} + \bar{\mu}_3(X)\frac{t^3}{3!} + (\bar{\mu}_4(X) - 3)\frac{t^4}{4!} + O(t^5).
\end{aligned}
$$

$\bar{\kappa}_3(X)$ を X もしくは X が従う分布の **歪度** (わいど, skewness) と呼び, $\bar{\kappa}_4(X)$ を **尖度** (せんど, kurtosis)と呼び, 次のようにも書くことにする:

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

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 $\bar{\kappa}_2(X) = 1$, $\bar{\kappa}_m(X) = 0$ $(m \neq 0)$ 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.

このことから, 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. それらのうち最初の2つが上で定義した歪度 $\bar{\kappa}_3(X)$ と尖度 $\bar{\kappa}_4(X)$ になっている.

分布の歪度 $\bar{\kappa}_3(X)$ と尖度 $\bar{\kappa}_4(X)$ 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: $\bar{\kappa}_4(X) = \bar{\mu}_4(X) - 3$ ではなく, 3 を引く前の $\bar{\mu}_4(X)$ 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扱いでキュムラントが非常に便利なことを重視したいので, 3 を引いた側の $\bar{\kappa}_4$ を尖度の定義として採用する. 3 を引いた方の $\bar{\kappa}_4(X) = \bar{\mu}_4(X) - 3$ は **過剰尖度** (かじょうせんど, **excess kurtosis**)と呼ばれることも多い. 正規分布の尖度を($\bar{\kappa}_4$ の方の 0 ではなく) $\bar{\mu}_4$ 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 $\bar{\kappa}(X)$ が表していることを「過剰」と表現している.

In [64]:
```
@vars t μ3 μ4 μ5 κ3 κ4 κ5
Mt = 1 + t^2/2 + μ3*t^3/6 + μ4*t^4/24 + μ5*t^5
expr = series(log(Mt), t)
```

Out[64]:

$$\frac{t^2}{2} + \frac{t^3\mu_3}{6} + t^4\left(\frac{\mu_4}{24} - \frac{1}{8}\right) + t^5\left(-\frac{\mu_3}{12} + \mu_5\right) + O\left(t^6\right)$$

5.6 問題: 正規分布の歪度と尖度

X が正規分布に従う確率変数であるとき, その歪度 $\bar{\kappa}_3(X)$ と尖度 $\bar{\kappa}_4(X)$ が以下のようになることを示せ:

$$\bar{\kappa}_3(X) = \bar{\kappa}_4(X) = 0.$$

解答例: X が正規分布 $\mathrm{Normal}(\mu, \sigma)$ に従う確率変数のとき, その標準化 $Z = (X - \mu)/\sigma$ は標準正規分布に従うので, $Z = (X - \mu)/\sigma$ のキュムラント母函数は $t^2/2$ になり, その 3 次以上の項がすべて 0 になる. ゆえに $\bar{\kappa}_m(X) = 0$ ($m = 3, 4, 5, \ldots$). 特に $\bar{\kappa}_3(X) = \bar{\kappa}_4(X) = 0$.

解答終

5.7 問題: 一様分布のキュムラント母函数と歪度と尖度

一様分布 $\mathrm{Uniform}(0, 1)$ に従う確率変数 X について, そのキュムラント母函数 $K_X(t) = \log E[e^{tX}]$ と歪度 $\bar{\kappa}_3(X)$ と尖度 $\bar{\kappa}_4(X)$ が以下のようになることを示せ:

$$K_X(t) = \log\frac{e^t - 1}{t}, \quad \bar{\kappa}_3(X) = 0, \quad \bar{\kappa}_4(X) = -\frac{6}{5}.$$

一様分布は歪度が 0 でかつ尖度がかなり小さな連続分布の例になっている.

解答例: モーメント母函数とキュムラント母函数は以下のように計算される:

$$\begin{aligned} E[e^{tX}] &= \int_0^1 e^{tx}\,dx = \left[\frac{e^{tx}}{t}\right]_0^1 = \frac{e^t - 1}{t}, \\ K_X(t) &= \log\frac{e^t - 1}{t} = \frac{1}{2}t + \frac{1}{12}\frac{t^2}{2} - \frac{1}{120}\frac{t^4}{4!} + O(t^5). \end{aligned}$$

ゆえに, $\mu = 1/2$, $\sigma = \sqrt{1/12}$, $Z = (X - \mu)/\sigma$ とおくと,

$$K_Z(t) = K_X\left(\frac{t}{\sigma}\right) - \frac{\mu}{\sigma}t = \frac{t^2}{2} - \frac{6}{5}\frac{t^4}{4!} + O(t^5).$$

ゆえに $\bar{\kappa}_3(X) = 0$, $\bar{\kappa}_4 = -6/5$.

解答終

In [65]:
```
@vars t
K_X = log((exp(t) - 1)/t)
κ = [limit(diff(K_X, t, m) , t, 0) for m in 1:4]
```

Out[65]:

$$\begin{bmatrix} \frac{1}{2} \\ \frac{1}{12} \\ 0 \\ -\frac{1}{120} \end{bmatrix}$$

In [66]:
```
μ, σ = κ[1], √κ[2]
κ = [limit(diff(K_X(t⇒t/σ) - μ/σ*t, t, m) , t, 0) for m in 1:4]
```

Out[66]:
$$\begin{bmatrix} 0 \\ 1 \\ 0 \\ -\frac{6}{5} \end{bmatrix}$$

5.8 問題: 独立な確率変数達の和のモーメント母函数と特性函数とキュムラント母函数

独立な確率変数達 $X_1, \dots, X_n$ の和のモーメント母函数と特性函数とキュムラント母函数が次のように表されることを示せ:

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

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,

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

ゆえに

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

解答終

5.9 問題: ガンマ分布のキュムラント母函数と歪度と尖度

ガンマ分布 $\mathrm{Gamma}(\alpha, \theta)$ に従う確率変数 X について, そのキュムラント母函数 $K_X(t) = \log E[e^{tX}]$ と歪度 $\bar{\kappa}_3(X)$ と尖度 $\bar{\kappa}_4(X)$ が以下のようになることを示せ:

$$K_X(t) = -\alpha \log(1 - \theta t), \quad \bar{\kappa}_3(X) = \frac{2}{\sqrt{\alpha}}, \quad \bar{\kappa}_3(X) = \frac{6}{\alpha}.$$

解答例: $t < 1/\theta$ と仮定する. このとき,

$$\mu = \alpha\theta, \quad \sigma = \sqrt{\alpha}\,\theta, \quad Z = \frac{X - \mu}{\sigma}$$

とおくと,

$$\begin{aligned} E[e^{tX}] &= \int_0^\infty e^{tx} \frac{e^{-x/\theta} x^{\alpha-1}}{\theta^\alpha \Gamma(\alpha)}\, dx = \frac{1}{\theta^\alpha \Gamma(\alpha)} \int_0^\infty e^{-((1-\theta t)/\theta)x} x^{\alpha-1}\, dx \\ &= \frac{1}{\theta^\alpha \Gamma(\alpha)} \frac{\theta^\alpha}{(1-\theta t)^\alpha} \Gamma(\alpha) = \frac{1}{(1-\theta t)^\alpha}, \\ K_X(t) &= -\alpha \log(1 - \theta t) \\ &= \alpha\theta t + \alpha\theta^2 \frac{t^2}{2} + \alpha\theta^3 \frac{t^3}{3} + \alpha\theta^4 \frac{t^4}{4} + O(t^5) \\ &= \mu t + \sigma^2 \frac{t^2}{2} + \frac{2\sigma^3}{\sqrt{\alpha}} \frac{t^3}{3!} + \frac{6\sigma^4}{\alpha} \frac{t^4}{4!} + O(t^5), \\ K_Z(t) &= K_X\left(\frac{t}{\sigma}\right) - \frac{\mu}{\sigma} t = \frac{t^2}{2} + \frac{2}{\sqrt{\alpha}} \frac{t^3}{3!} + \frac{6}{\alpha} \frac{t^4}{4!} + O(t^5). \end{aligned}$$

ゆえに,

$$\bar{\kappa}_3(X) = \frac{2}{\sqrt{\alpha}}, \quad \bar{\kappa}_3(X) = \frac{6}{\alpha}.$$

解答終

In [67]:
```
@vars t
@vars α θ positive=true
K_X = -α*log(1 - θ*t)
κ = [diff(K_X, t, m)(t⇒0).simplify().factor() for m in 1:4]
```

Out[67]:

$$\begin{bmatrix} \alpha\theta \\ \alpha\theta^2 \\ 2\alpha\theta^3 \\ 6\alpha\theta^4 \end{bmatrix}$$

In [68]:
```
μ, σ = κ[1], √κ[2]
[diff(K_X(t⇒t/σ) - μ/σ*t, t, m)(t⇒0).simplify().factor() for m in 1:4]
```

Out[68]:

$$\begin{bmatrix} 0 \\ 1 \\ \frac{2}{\sqrt{\alpha}} \\ \frac{6}{\alpha} \end{bmatrix}$$

5.10 問題: 二項分布のキュムラント母函数と歪度と尖度

二項分布 Binomial(n, p) に従う確率変数 X について, そのキュムラント母函数 $K_X(t) = \log E[e^{tX}]$ と歪度 $\bar{\kappa}_3(X)$ と尖度 $\bar{\kappa}_4(X)$ が以下のようになることを示せ:

$$K_X(t) = n\log(1 + p(e^t - 1)), \quad \bar{\kappa}_3(X) = \frac{1-2p}{\sqrt{np(1-p)}}, \quad \bar{\kappa}_4(X) = \frac{1-6p(1-p)}{np(1-p)}.$$

注意: $n = 1$, $p = 1/2$ のとき $\bar{\kappa}_4(X) = -2$ となる. 実はその場合の分布 Bernoulli(1/2) が確率分布全体の中で最小の尖度を持つ分布になっている.

解答例: $X_1, \dots, X_n$ をそれぞれがBernoulli分布 Binomial(p) に従う独立同分布な確率変数達であるとする. このとき, 二項分布に従う確率変数は $X = \sum_{i=1}^n X_i$ と作れる. ゆえに

$$K_X(t) = \sum_{i=1}^n K_{X_i}(t) = nK_{X_1}(t).$$

$X_1 \sim$ Bernoulli(p) について,

$$\begin{aligned} &E[e^{tX_1}] = e^{t1}p + e^{t0}(1-p) = 1 + p(e^t - 1), \\ &K_{X_1}(t) = \log(1 + p(e^t - 1)), \\ &K_X(t) = n\log(1 + p(e^t - 1)) \end{aligned}$$

なので,

$$\mu = np, \quad \sigma = \sqrt{np(1-p)}$$

とおくと,

$$\begin{aligned} &K_X(t) = n\log(1 + p(e^t - 1)) \\ &= npt + np(1-p)\frac{t^2}{2} + np(1-p)(1-2p)\frac{t^3}{3!} + np(1-p)(6p^2 - 6p + 1)\frac{t^4}{4!} + O(t^5) \\ &= \mu t + \sigma^2\frac{t^2}{2} + \sigma^2(1-2p)\frac{t^3}{3!} + \sigma^2(1 - 6p(1-p))\frac{t^4}{4!} + O(t^5), \\ &K_{(X-\mu)/\sigma}(t) = K_X\left(\frac{t}{\sigma}\right) - \frac{\mu}{\sigma}t = \frac{t^2}{2} + \frac{1-2p}{\sigma}\frac{t^3}{3!} + \frac{1-6p(1-p)}{\sigma^2}\frac{t^4}{4!} + O(t^5). \end{aligned}$$

ゆえに

$$\bar{\kappa}_3(X) = \frac{1-2p}{\sigma} = \frac{1-2p}{\sqrt{np(1-p)}}, \quad \bar{\kappa}_4(X) = \frac{1-6p(1-p)}{\sigma^2} = \frac{1-6p(1-p)}{np(1-p)}.$$

解答終

In [69]:
```
kurtosis(Bernoulli(0.5))
```

Out[69]: -2.0

In [70]:
```
@vars n p t
K_X = n * log(1 + p*(exp(t) - 1))
κ = [diff(K_X, t, m)(t⇒0).simplify().factor() for m in 1:4]
```

Out[70]:
$$\begin{bmatrix} np \\ -np(p-1) \\ np(p-1)(2p-1) \\ -np(p-1)\left(6p^2-6p+1\right) \end{bmatrix}$$

In [71]:
```
μ, σ = κ[1], √κ[2]
[diff(K_X(t⇒t/σ) - μ/σ*t, t, m)(t⇒0).simplify().factor() for m in 1:4]
```

Out[71]:
$$\begin{bmatrix} 0 \\ 1 \\ -\dfrac{2p-1}{\sqrt{-np(p-1)}} \\ -\dfrac{6p^2-6p+1}{np(p-1)} \end{bmatrix}$$

6 独立同分布な確率変数達の不偏分散の分散

以下の節においては, $X_1, \ldots, X_n$ は独立同分布な確率変数達であるとし, それらの標本分散と不偏分散を以下のように書くことにする:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

さらに, X_i の期待値と分散をそれぞれ μ, σ^2 と書くことにする.

このとき, 以下が成立することはすでに示した:

$$E[\bar{X}] = \mu, \quad E[S_X^2] = \sigma^2, \quad \mathrm{var}(X) = E[(\bar{X}-\mu)^2] = \frac{\sigma^2}{n}.$$

さらに以下では, X_i の **歪度** (わいど, skewness)と **尖度** (せんど, kurtosis)をそれぞれ $\bar{\kappa}_3 = \bar{\kappa}_3(X_i)$, $\bar{\kappa}_4 = \bar{\kappa}_4(X_i)$ と書くことにする:

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

このとき, 標本平均と不偏分散の共分散と不偏分散の分散は X_i の歪度 $\bar{\kappa}_3$ と尖度 $\bar{\kappa}_4$ を使って次のように書ける:

$$\mathrm{cov}(\bar{X}, S_X^2) = \sigma^3\frac{\bar{\kappa}_3}{n}, \quad \mathrm{var}(S_X^2) = \sigma^4\left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}\right).$$

以下ではこの結果を証明する.

6.1 期待値 0, 分散 1 の場合への帰着

$Z = (X-\mu)/\sigma$ とおくと, Z の期待値と分散はそれぞれ 0 と 1 になる.

$Z_i = (X_i-\mu)/\sigma$ とおくと, $Z_1, \ldots, Z_n$ はそれぞれが Z と同じ分布に従う独立同分布な確率変数達になる. そして, $X_i = \sigma Z_i + \mu$ より,

$$\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i, \quad S_Z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$$

とおくと,

$$\bar{X} = \sigma\bar{Z} + \mu, \quad S_X^2 = \sigma^2 S_Z^2$$

なので,

$$\mathrm{cov}(\bar{X}, S_X^2) = \sigma^3 \,\mathrm{cov}(\bar{Z}, S_Z^2), \quad \mathrm{var}(S_X^2) = \sigma^4 \,\mathrm{var}(S_Z^2).$$

これより, 標本分散と不偏分散の分散と共分散の計算は $\mu = 0, \sigma = 1$ の場合に帰着する.

6.2 標本平均と不偏分散の共分散の計算

標本平均と不偏分散の共分散は X_i の歪度 $\bar{\kappa}_3$ を使って次のように書ける:

$$\mathrm{cov}(\bar{X}, S_X^2) = \frac{\bar{\kappa}_3}{n}\sigma^3.$$

証明: $\mu = 0, \sigma = 1$ と仮定する. このとき, $\mu = 0$ という仮定より,

$$\mathrm{cov}(\bar{X}, S_X^2) = E[\bar{X}(S_X^2 - \sigma^2)] = E[\bar{X}S_X^2] - \sigma^2 \underbrace{E[\bar{X}]}_{=0} = E[\bar{X}S_X^2].$$

そして,

$$\begin{aligned}
n\bar{X}^2 &= \frac{1}{n}\sum_{i=1}^n X_i^2 + \frac{2}{n}\sum_{1\le i<j\le n} X_iX_j, \\
(n-1)S_X^2 &= \sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n (X_i^2 - 2\bar{X}X_i + \bar{X}^2) = \sum_{i=1}^n X_i^2 - n\bar{X}^2 \\
&= \sum_{i=1}^n X_i^2 - \frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{2}{n}\sum_{1\le i<j\le n} X_iX_j = \frac{n-1}{n}\sum_{i=1}^n X_i^2 - \frac{2}{n}\sum_{1\le i<j\le n} X_iX_j, \\
\therefore \quad S_X^2 &= \frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{2}{n(n-1)}\sum_{1\le i<j\le n} X_iX_j.
\end{aligned}$$

ゆえに

$$\bar{X}S_X^2 = \frac{1}{n}\sum_{i=1}^n X_i \left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{2}{n(n-1)}\sum_{1\le i<j\le n} X_iX_j\right).$$

$E[X_i] = \mu = 0$ という仮定と X_i 達の独立性より, i, j, k がすべて等しいとき以外には $E[X_iX_jX_k] = 0$ となり, $\mu = 0, \sigma = 1$ という仮定より $E[X_i^3] = \bar{\kappa}_3$ となるので,

$$\mathrm{cov}(\bar{X}, S_X^2) = E[\bar{X}S_X^2] = \frac{1}{n^2}\sum_{i=1}^n E[X_i^3] = \frac{\bar{\kappa}_3}{n}.$$

$\mu = 0, \sigma = 1$ と仮定しない一般の場合には, これが σ^3 倍されて,

$$\mathrm{cov}(\bar{X}, S_X^2) = \sigma^3\frac{\bar{\kappa}_3}{n}.$$

証明終

6.3 不偏分散の分散の計算

不偏分散の分散は X_i の尖度 $\bar{\kappa}_4$ を使って次のように書ける:

$$\mathrm{var}(S_X^2) = \sigma^4\left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}\right).$$

証明: $\mu = 0, \sigma = 1$ と仮定する. このとき, $E[S_X^2] = \sigma^2 = 1$ より,

$$\mathrm{var}(S_X^2) = E[(S_X^2)^2] - E[S_X^2]^2 = E[(S_X^2)^2] - 1.$$

そして, 前節の計算より

$$S_X^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 - \frac{2}{n(n-1)}\sum_{1\le i<j\le n} X_iX_j$$

なので,

$$
\begin{aligned}
(S_X^2)^2 &= \frac{1}{n^2}\left(\sum_{i=1}^n X_i^4 + 2\sum_{1\le i<j\le n} X_i^2 X_j^2\right) \\
&\quad - \frac{2}{n^2(n-1)}\sum_{i=1}^n \sum_{1\le j<k\le n} X_i^2 X_j X_k \\
&\quad + \frac{4}{n^2(n-1)^2}\sum_{1\le i<j\le n}\sum_{1\le k<l\le n} X_i X_j X_k X_l.
\end{aligned}
$$

$\mu = 0, \sigma = 1$ でかつ $X_1, \dots, X_n$ は独立であるという仮定と, $E[X_i] = 0$, $E[X_i^2] = 1$, $E[X_i^4] = \bar{\kappa}_4 + 3$ であることを使うと,

$$
\begin{aligned}
E\left[\sum_{i=1}^n X_i^4\right] &= n(\bar{\kappa}_4 + 3), \\
E\left[2\sum_{1\le i<j\le n} X_i^2 X_j^2\right] &= 2\frac{n(n-1)}{2} = n(n-1), \\
E\left[\sum_{1\le j<k\le n} X_i^2 X_j X_k\right] &= 0, \\
E\left[\sum_{1\le i<j\le n}\sum_{1\le k<l\le n} X_i X_j X_k X_l\right] &= \frac{n(n-1)}{2}.
\end{aligned}
$$

最後の等式は $i = k, j = l$ の場合のみが和に寄与することを使った. ゆえに

$$
\begin{aligned}
\mathrm{var}(S_X^2) &= \frac{n(\bar{\kappa}_4 + 3) + n(n-1)}{n^2} + \frac{4}{n^2(n-1)^2}\frac{n(n-1)}{2} - 1 \\
&= \frac{\bar{\kappa}_4 + n + 2}{n} + \frac{2}{n(n-1)} - 1 \\
&= \frac{\bar{\kappa}_4 + n + 2}{n} + \frac{2}{n-1} - \frac{2}{n} - \frac{n}{n} \\
&= \frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}.
\end{aligned}
$$

$\mu = 0, \sigma = 1$ と仮定しない一般の場合には, これが σ^4 倍されて,

$$
\mathrm{var}(S_X^2) = \sigma^4\left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}\right).
$$

証明終

6.4 歪度と尖度に関する不等式

以上の結果を用いて, 歪度 $\bar{\kappa}_3$ と尖度 $\bar{\kappa}_4$ のあいだには

$$
\bar{\kappa}_4 \ge \bar{\kappa}_3^2 - 2 \ge -2
$$

が成立することを示そう. 標本平均 $\bar{X}$ と不偏分散 S_X^2 の分散共分散行列

$$
\begin{bmatrix} \mathrm{var}(\bar{X}) & \mathrm{cov}(\bar{X}, S_X^2) \\ \mathrm{cov}(S_X^2, \bar{X}) & \mathrm{var}(S_X^2) \end{bmatrix} = \begin{bmatrix} \sigma^2/n & \sigma^3\bar{\kappa}_3/n \\ \sigma^3\bar{\kappa}_3/n & \sigma^4(\bar{\kappa}_4/n + 2/(n-1)) \end{bmatrix}
$$

は半正定値になる. 特にその行列式は 0 以上になる:

$$
0 \le \frac{\sigma^2}{n}\sigma^4\left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}\right) - \left(\frac{\sigma^3\bar{\kappa}_3}{n}\right)^2 = \frac{\sigma^6}{n^2}\left(\bar{\kappa}_4 + \frac{2}{1-1/n} - \bar{\kappa}_3^2\right).
$$

ゆえに,

$$
\bar{\kappa}_4 \ge \bar{\kappa}_3^2 - \frac{2}{1-1/n}.
$$

$n \to \infty$ とすることによって,

$$
\bar{\kappa}_4 \ge \bar{\kappa}_3^2 - 2.
$$

を得る.

注意: 上の不等式は **Karl Pearson の不等式** と呼ばれることがある.

- Pearson, Karl. IX. Mathematical contributions to the theory of evolution.—XIX. Second supplement to a memoir on skew variation. Philosophical Transactions of the Royal Society of London A, 216 (546): 429–457. [doi (https://doi.org/10.1098/rsta.1916.0009)]

注意: 実は上で紹介した歪度と尖度に関する不等式の証明はひどく遠回りになっている. 分散共分散行列の半正定値性(非負性)に帰着するのであれば, 確率変数 X について $X-\mu$ と $(X-\mu)^2$ の分散共分散行列を求めれみればよい.

$$E[X]=\mu,\quad E[(X-\mu)^2]=\sigma^2,\quad E[(X-\mu)^3]=\sigma^3\bar{\kappa}_3,\quad E[(X-\mu)^4]=\sigma^4(\bar{\kappa}_4+3)$$

のとき,

$$\begin{aligned}
&E[X-\mu]=0,\quad E[(X-\mu)^2]=\sigma^2,\\
&\mathrm{var}(X-\mu)=\sigma^2,\\
&\mathrm{cov}(X-\mu,(X-\mu)^2)=E[(X-\mu)^3]=\sigma^3\bar{\kappa}_3,\\
&\mathrm{var}((X-\mu)^2)=E[(X-\mu)^4]-E[(X-\mu)^2]^2=\sigma^4(\bar{\kappa}_4+3)-\sigma^4=\sigma^4(\bar{\kappa}_4+2).
\end{aligned}$$

これより, $X-\mu$ と $(X-\mu)^2$ の分散共分散行列

$$\begin{bmatrix}\sigma^2 & \sigma^3\bar{\kappa}_3\\ \sigma^3\bar{\kappa}_3 & \sigma^4(\bar{\kappa}_4+2)\end{bmatrix}=\begin{bmatrix}\sigma & 0\\ 0 & \sigma^2\end{bmatrix}\begin{bmatrix}1 & \bar{\kappa}_3\\ \bar{\kappa}_3 & \bar{\kappa}_4+2\end{bmatrix}\begin{bmatrix}\sigma & 0\\ 0 & \sigma^2\end{bmatrix}$$

は半正定値なので, 特にその行列式は 0 以上になり,

$$\bar{\kappa}_4+2\geq\bar{\kappa}_3^3$$

という不等式を得る.

6.5 問題: 不偏分散と不偏補正されていない標本分散の平均二乗誤差の比較

不偏補正されていない($n-1$ ではなく $n-1$ で割ってできる)標本分散

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

と推定先の値 σ^2 の差の二乗の期待値(平均二乗誤差)が次のようになることを示せ:

$$E\left[\left(\frac{n-1}{n}S_X^2-\sigma^2\right)^2\right]=\left(\frac{n-1}{n}\right)^2\mathrm{var}(S_X^2)+\frac{\sigma^4}{n^2}.$$

これより, 次の公式が導かれることを示せ:

$$\mathrm{var}(S_X^2)-E\left[\left(\frac{n-1}{n}S_X^2-\sigma^2\right)^2\right]=\frac{\sigma^4}{n^2}\left(\left(2-\frac{1}{n}\right)\bar{\kappa}_4+3+\frac{2}{n-1}\right).$$

これより, $\bar{\kappa}_4\geq-3/2$ ならば不偏分散の平均二乗誤差の方が不偏補正されていない標本平均の平均二乗誤差よりも大きくなることがわかる.

注意: $\bar{\kappa}_4\geq\bar{\kappa}_3^2-2\geq-2$ なので多くの場合に $\bar{\kappa}_4>-3/2$ となる. 例えば正規分布では $\bar{\kappa}_4=0$ なので,

$$\mathrm{var}(S_X^2)-E\left[\left(\frac{n-1}{n}S_X^2-\sigma^2\right)^2\right]=\frac{\sigma^4}{n^2}\frac{3n-1}{n-1}$$

となり, 不偏分散の平均二乗誤差は補正されていない標本分散より大きくなる. 不偏分散の採用にはこのようなトレードオフも存在するが, この害は小さいと考えられる.

解答例: $(n-1)/n-1=-1/n$ と $E[S_X^2]=\sigma^2$, $E[(S_X^2-\sigma^2)^2]=\mathrm{var}(S_X^2)$ より,

$$\begin{aligned}
E\left[\left(\frac{n-1}{n}S_X^2-\sigma^2\right)^2\right]&=E\left[\left(\frac{n-1}{n}(S_X^2-\sigma^2)-\frac{\sigma^2}{n}\right)^2\right]\\
&=E\left[\left(\frac{n-1}{n}\right)^2(S_X^2-\sigma^2)^2-2\frac{\sigma^2}{n}\frac{n-1}{n}(S_X^2-\sigma^2)+\frac{\sigma^4}{n^2}\right]\\
&=\left(\frac{n-1}{n}\right)^2\mathrm{var}(S_X^2)+\frac{\sigma^4}{n^2}.
\end{aligned}$$

これより,

$$\mathrm{var}(S_X^2) - E\left[\left(\frac{n-1}{n}S_X^2 - \sigma^2\right)^2\right] = \frac{2n-1}{n^2}\mathrm{var}(S_X^2) - \frac{\sigma^4}{n^2}.$$

これに $\mathrm{var}(S_X^2) = \sigma^4\left(\bar{\kappa}_4/n + 2/(n-1)\right)$ を代入すると,

$$\begin{aligned}
\mathrm{var}(S_X^2) - E\left[\left(\frac{n-1}{n}S_X^2 - \sigma^2\right)^2\right] &= \frac{2n-1}{n^2}\sigma^4\left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}\right) - \frac{\sigma^4}{n^2} \\
&= \frac{\sigma^4}{n^2}\left(\frac{2n-1}{n}\bar{\kappa}_4 + \frac{4n-2}{n-1} - 1\right) = \frac{\sigma^4}{n^2}\left(\frac{2n-1}{n}\bar{\kappa}_4 + \frac{3n-1}{n-1}\right) \\
&= \frac{\sigma^4}{n^2}\left(\left(2 - \frac{1}{n}\right)\bar{\kappa}_4 + 3 + \frac{2}{n-1}\right).
\end{aligned}$$

解答終

In [72]:

```julia
function print_varS²_minus_MSEV²(dist, n; L=10^6)
    @show dist
    @show μ = mean(dist)
    @show σ = std(dist)
    @show κ = kurtosis(dist)
    @show n
    @show L = 10^6
    X = rand(dist, n, L)
    S² = var.(eachcol(X))
    V² = var.(eachcol(X); corrected=false)
    var_S² = mean((s² - σ^2)^2 for s² in S²)
    MSE_V² = mean((v² - σ^2)^2 for v² in V²)
    theoretical_val = σ^4/n^2*((2n-1)/n * κ + (3n-1)/(n-1))
    theoretical_value = theoretical_val
    @show theoretical_val
    @show var_S² - MSE_V²
    @show theoretical_value / σ^4
    @show (var_S² - MSE_V²) / σ^4
    nothing
end
```

Out[72]:

```
print_varS²_minus_MSEV² (generic function with 1 method)
```

In [73]:

```julia
print_varS²_minus_MSEV²(Normal(1, 0.5), 5)
```

```
dist = Normal{Float64}(μ=1.0, σ=0.5)
μ = mean(dist) = 1.0
σ = std(dist) = 0.5
κ = kurtosis(dist) = 0.0
n = 5
L = 10 ^ 6 = 1000000
theoretical_val = 0.00875
var_S² - MSE_V² = 0.008711816245110061
theoretical_value / σ ^ 4 = 0.14
(var_S² - MSE_V²) / σ ^ 4 = 0.13938905992176098
```

正規分布の場合に確かに理論通りに不偏分散の平均二乗誤差の方が非補正標本分散の平均二乗誤差よりも大きくなっている.

In [74]:

```julia
print_varS²_minus_MSEV²(Bernoulli(0.5), 5)
```

```
dist = Bernoulli{Float64}(p=0.5)
μ = mean(dist) = 0.5
σ = std(dist) = 0.5
κ = kurtosis(dist) = -2.0
n = 5
L = 10 ^ 6 = 1000000
theoretical_val = -0.0002500000000000002
var_S² - MSE_V² = -0.00025235920018789205
theoretical_value / σ ^ 4 = -0.0040000000000000036
(var_S² - MSE_V²) / σ ^ 4 = -0.004037747203006273
```

これが尖度が最小の場合. 不偏分散の平均二乗誤差の方を確かに小さくできているが, ほんの少しだけである.

7 標本分布における標本平均と不偏分散の同時分布の視覚化

以下の節においては, $X_1, \dots, X_n$ は独立同分布な確率変数達であるとし, それらの標本分散と不偏分散を以下のように書くことにする:

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

さらに, 各 X_i の期待値と分散と歪度と尖度をそれぞれ $\mu, \sigma^2, \bar{\kappa}_3, \bar{\kappa}_4$ と書くことにする. このとき, 標本平均 $\bar{X}$ と不偏分散 S^2 の同時分布の期待値と分散・共分散は以下のようになるのであった:

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

これを, 期待値ベクトルと分散共分散行列の形式で書くと,

$$\begin{aligned}
&\begin{bmatrix} E[\bar{X}] \\ E[S^2] \end{bmatrix} = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}, \\
&\begin{bmatrix} \mathrm{var}(\bar{X}) & \mathrm{cov}(\bar{X}, S^2) \\ \mathrm{cov}(S^2, \bar{X}) & \mathrm{var}(S^2) \end{bmatrix} = \frac{1}{n}\begin{bmatrix} \sigma^2 & \sigma^3\bar{\kappa}_3 \\ \sigma^3\bar{\kappa}_3 & \sigma^4\left(\bar{\kappa}_4 + \frac{2}{1-1/n}\right) \end{bmatrix} \\
&\qquad = \frac{1}{n}\begin{bmatrix} \sigma & 0 \\ 0 & \sigma^2 \end{bmatrix}\begin{bmatrix} 1 & \bar{\kappa}_3 \\ \bar{\kappa}_3 & \bar{\kappa}_4 + \frac{2}{1-1/n} \end{bmatrix}\begin{bmatrix} \sigma & 0 \\ 0 & \sigma^2 \end{bmatrix}.
\end{aligned}$$

以下では X_i の従う分布が色々な場合に $\bar{X}, S^2$ の同時分布を視覚化してみる. その結果を見ると, 一般に以下が成立していそうなことがわかる:

- 標本平均 $\bar{X}$ と不偏分散 S^2 の組 $(\bar{X}, S^2)$ の平面上での分布は, n が十分に大きければ (μ, σ^2) を中心とする楕円状の小さな領域内に集中する分布(実は多変量正規分布で近似される)になり(中心極限定理!), $n \to \infty$ で (μ, σ^2) に集中するようになる(大数の法則!).
- もとの分布が対称ならば $(\bar{X}, S^2)$ の分布も左右対称になり, もとの分布が非対称ならば $(\bar{X}, S^2)$ の分布も左右非対称になり, n が大きなとき斜めに傾いた楕円状の領域内に集中する分布になる.

In [75]:
```
myskewness(dist) = skewness(dist)
myskewness(dist::MixtureModel) = _myskewness(dist)
function _myskewness(dist)
    μ, σ = mean(dist), std(dist)
    quadgk(x -> ((x-μ)/σ)^3*pdf(dist, x), extrema(dist)...)[1]
end

mykurtosis(dist) = kurtosis(dist)
mykurtosis(dist::MixtureModel) = _mykurtosis(dist)
function _mykurtosis(dist)
    μ, σ = mean(dist), std(dist)
    quadgk(x -> ((x-μ)/σ)^4*pdf(dist, x), extrema(dist)...)[1] - 3
end
```

Out[75]:
```
_mykurtosis (generic function with 1 method)
```

In [76]:
```
skewness(Gamma(2, 3)), _myskewness(Gamma(2,3))
```

Out[76]:
```
(1.414213562373095, 1.4142135623731023)
```

In [77]:
```
kurtosis(Gamma(2, 3)), _mykurtosis(Gamma(2,3))
```

Out[77]:
```
(3.0, 2.999999999999999)
```

In [78]:

```julia
function plot_dist(dist; m = 4, xlim = :auto, kwargs...)
    if xlim isa Symbol || length(xlim) != 2
        μ, σ = mean(dist), std(dist)
        a = max(minimum(dist), μ - m*σ)
        b = min(maximum(dist), μ + m*σ)
        xmin, xmax = a - 0.1(b - a), b + 0.1(b - a)
    else
        xmin, xmax = xlim
    end
    if dist isa DiscreteUnivariateDistribution
        x = range(round(xmin) - 1, round(xmax) + 1, 1000)
        plot(x, x -> pdf(dist, round(x)); label="", xlim, kwargs...)
    else
        x = range(xmin, xmax, 1000)
        plot(x, x -> pdf(dist, x); label="", xlim, kwargs...)
    end
end

function plot_X̄_and_SX²(dist; n = 10, L = 10^4,
        sk = myskewness(dist), ku = mykurtosis(dist),
        size = (400, 350), legend = :topleft, alpha = 0.4, kwargs...)
    μ, σ = mean(dist), std(dist)
    Σ = (1/n)*[σ^2 σ^3*sk; σ^3*sk σ^4*(ku+2/(1-1/n))]
    λ, U = eigen(Σ)
    h = quantile(Chisq(2), 0.95)
    f(t) = μ + √(λ[1]*h)*cos(t)*U[1,1] + √(λ[2]*h)*sin(t)*U[1,2]
    g(t) = σ^2 + √(λ[1]*h)*cos(t)*U[2,1] + √(λ[2]*h)*sin(t)*U[2,2]

    X̄ = Vector{Float64}(undef, L)
    S² = similar(X̄)
    tmp = [Vector{eltype(dist)}(undef, n) for _ in 1:Threads.nthreads()]
    Threads.@threads for i in 1:L
        X = rand!(dist, tmp[Threads.threadid()])
        X̄[i] = mean(X)
        S²[i] = var(X)
    end
    scatter(X̄, S²; ms=1.5, msw=0, alpha, label="")
    scatter!([μ], [σ^2]; msc=:auto, label="(μ, σ²)")
    x = range(extrema(X̄)..., 200)
    y = range(extrema(S²)..., 200)
    t = range(0, 2π, 1000)
    plot!(f.(t), g.(t); label="", levels=[h], c=2)
    plot!(; xlabel="x̄", ylabel="s²")
    plot!(; size, legend, kwargs...)
end

function plot_X̄_and_SX²_2x2(dist;
        m = 4, distxlim = :auto, distylim = :auto,
        ns = (10, 40, 160, 640), size=(700, 700), kwargs...)
    μ, σ², sk, ku = mean(dist), var(dist), myskewness(dist), mykurtosis(dist)
    println(dist)
    @show μ σ²
    println("skewness = ", sk)
    println("kurtosis = ", ku)
    flush(stdout)

    D = plot_dist(dist; m, xlim=distxlim, ylim=distylim, size=(300, 200))
    display(D)

    PP = []
    for n in ns
        P = plot_X̄_and_SX²(dist; n, sk, ku, kwargs..., title="n = $n")
        push!(PP, P)
    end
    plot(PP...; size, layout=(2,2),
        titlefontsize=12, guidefontsize=10, tickfontsize=8)
end
```

Out[78]:

```
plot_X̄_and_SX²_2x2 (generic function with 1 method)
```

7.1 正規分布の標本分布における $\bar{X}, S^2$ の同時分布

In [79]:
```
plot_X̄_and_SX²_2x2(Normal(); xlim=(-1.1, 1.1), ylim=(-0.1, 3.0))
```

```
Normal{Float64}(μ=0.0, σ=1.0)
μ = 0.0
σ² = 1.0
skewness = 0.0
kurtosis = 0.0
```

Out[79]:

7.2 一様分布の標本分布における$\bar{X}, S^2$ の同時分布

In [80]:
```
plot_X̄_and_SX²_2x2(Uniform(); xlim=(0.2, 0.8), ylim=(0.01, 0.18))
```

```
Uniform{Float64}(a=0.0, b=1.0)
μ = 0.5
σ² = 0.08333333333333333
skewness = 0.0
kurtosis = -1.2
```

Out[80]:

7.3 単峰型のガンマ分布の標本分布における $\bar{X}, S^2$ の同時分布

In [81]:
```
plot_X̄_and_SX²_2x2(Gamma(3, 4); xlim=(5, 20), ylim=(0, 200))
```

```
Gamma{Float64}(α=3.0, θ=4.0)
μ = 12.0
σ² = 48.0
skewness = 1.1547005383792517
kurtosis = 2.0
```

Out[81]:

7.4 指数分布の標本分布における $\bar{X}, S^2$ の同時分布

In [82]:

```
plot_X̄_and_SX²_2x2(Exponential(); xlim=(0.2, 2.3), ylim=(-0.2, 7))
```

```
Exponential{Float64}(θ=1.0)
μ = 1.0
σ² = 1.0
skewness = 2.0
kurtosis = 6.0
```

Out[82]:

7.5 非対称なベータ分布の標本分布における $\bar{X}, S^2$ の同時分布

In [83]:
```
plot_X̄_and_SX²_2x2(Beta(0.1, 0.9); xlim=(-0.02, 0.4), ylim=(-0.05, 0.2),
    distxlim=(-0.05, 1.05), distylim=(-0.5, 10))
```

```
Beta{Float64}(α=0.1, β=0.9)
μ = 0.1
σ² = 0.045000000000000005
skewness = 2.514157444218836
kurtosis = 5.611111111111112
```

Out[83]:

7.6 非対称な2つ山の混合正規分布の標本分布における $\bar{X}, S^2$ の同時分布

In [84]:

```
plot_X̄_and_SX²_2x2(MixtureModel([Normal(), Normal(20)], [0.95, 0.05]);
    xlim=(-1, 8), ylim=(-2, 120))
```

```
MixtureModel{Normal{Float64}}(K = 2)
components[1] (prior = 0.9500): Normal{Float64}(μ=0.0, σ=1.0)
components[2] (prior = 0.0500): Normal{Float64}(μ=20.0, σ=1.0)

μ = 1.0
σ² = 20.0
skewness = 3.823676241524649
kurtosis = 13.58499999999097
```

Out[84]:

7.7 対称なBernoulli分布の標本分布における $\bar{X}, S^2$ の同時分布

In [85]:
```
plot_X̄_and_SX²_2x2(Bernoulli(0.5); xlim=(-0.02, 1.02), ylim=(-0.01, 0.30))
```

```
Bernoulli{Float64}(p=0.5)
μ = 0.5
σ² = 0.25
skewness = 0.0
kurtosis = -2.0
```

Out[85]:

7.8 非対称なBernoulli分布の標本分布における $\bar{X}, S^2$ の同時分布

In [86]:
```
plot_X̄_and_SX²_2x2(Bernoulli(0.1);
    xlim=(-0.02, 0.6), ylim=(-0.01, 0.3), distxlim=(-1, 2))
```

```
Bernoulli{Float64}(p=0.1)
μ = 0.1
σ² = 0.09000000000000001
skewness = 2.6666666666666665
kurtosis = 5.111111111111109
```

Out[86]:

7.9 Poisson分布の標本分布における $\bar{X}, S^2$ の同時分布

In [87]:
```
plot_X̄_and_SX²_2x2(Poisson(0.5); xlim=(-0.02, 1.2), ylim=(-0.04, 2.0))
```

```
Poisson{Float64}(λ=0.5)
μ = 0.5
σ² = 0.5
skewness = 1.414213562373095
kurtosis = 2.0
```

Out[87]:

7.10 対数正規分布の標本分布における $\bar{X}, S^2$ の同時分布

In [88]:
```
plot_X̄_and_SX²_2x2(LogNormal(); xlim=(-0.2, 5), ylim=(-2, 50))
```

```
LogNormal{Float64}(μ=0.0, σ=1.0)
μ = 1.6487212707001282
σ² = 4.670774270471604
skewness = 6.184877138632554
kurtosis = 110.9363921763115
```

Out[88]:

対数正規分布では $(\bar{X}, S^2)$ の分布の正規分布近似の精度をよくするためには相当に n を大きくする必要がある.

8 正規分布の標本分布の場合

この節では, $X_1, \ldots, X_n$ はそれぞれが正規分布 $\mathrm{Normal}(\mu, \sigma)$ に従う独立な確率変数達であると仮定する:

$$(X_1, \ldots, X_n) \sim \mathrm{Normal}(\mu, \sigma)^n.$$

さらに, それらの標本分散と不偏分散を

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

と書く. このとき, 正規分布の標本分布の特殊性によって, $\bar{X}$ と S^2 は独立になり, $\bar{X}$ は平均 μ, 分散 σ^2/n の正規分布に従い,

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

は自由度 $\nu = n-1$ のχ²分布に従う:

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

以下ではこの結果を証明する.

8.1 準備: $\sqrt{n}\,\bar{x}$ を含む正規直交座標系

$\mathbb{R}^n$ を縦ベクトルの空間とみなし, その標準座標系を $x_1, \dots, x_n$ と書く. このとき, $\mathbb{R}^n$ の任意の正規直交基底 $v_1, \dots, v_n$ について

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} y_i v_i$$

で定義される $\mathbb{R}^n$ の座標系 $y_1, \dots, y_n$ をその直交基底が定める $\mathbb{R}^n$ の **正規直交座標系** と呼ぶ. このとき, ベクトル $a, b \in \mathbb{R}^n$ の標準内積を (a, b) と書く, それに対応するノルムを $\|a\| = \sqrt{(a, a)}$ と書くと,

$$y_i = (v_i, x), \quad \|x\|^2 = \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i^2$$

などが成立している.

以下では $v_n \in \mathbb{R}^n$ は

$$v_n = \frac{1}{\sqrt{n}}\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

と定められていると仮定し, $\bar{x}$ を次のように定める:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

このとき, $\|v_n\|^2 = (1^2 + \cdots + 1^2)/n = 1$ なので v_n は長さ 1 の単位ベクトルになり,

$$(v_n, x) = \frac{x_1 + \cdots + x_n}{\sqrt{n}} = \sqrt{n}\,\bar{x}.$$

さらに, v_n を含む $\mathbb{R}^n$ の任意の正規直交基底 $v_1, \dots, v_{n-1}, v_n$ を取り, それが定める正規直交座標系を $y_1, \dots, y_{n-1}, y_n$ と書くことにする. このとき, $y_n = (v_n, x) = \sqrt{n}\,\bar{x}$ なので, $y_1, \dots, y_{n-1}, y_n$ は $y_n = \sqrt{n}\,\bar{x}$ の正規直交座標系への拡張になっている.

例: v_n を含む正規直交基底が任意に取られているせいで議論が抽象的過ぎるように感じる人達は, 以下の $v_1, \dots, v_{n-1}$ 上の v_n を合わせたものがそのような正規直交基底の具体例になっていることを確認して, それらを使用すればよい:

$$v_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad v_{n-1} = \frac{1}{\sqrt{n(n-1)}}\begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ -(n-1) \end{bmatrix}.$$

これらが内積について $(v_i, v_j) = \delta_{ij}$ を満たしていること(正規直交基底であること)は, この定義式を眺めて暗算すればすぐに確かめられる. **例終**

ベクトル v が

$$v = \begin{bmatrix} x_1 - \mu \\ \vdots \\ x_n - \mu \end{bmatrix} = x - \sqrt{n}\,\mu v_n$$

と定義されているとき,

$$(v_i, v) = (v_i, x) = y_i \quad (i = 1, \dots, n-1), \quad (v_n, v) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (x_i - \mu) = \sqrt{n}(\bar{x} - \mu)$$

なので,

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

さらに不偏分散関係でよく出て来る公式

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

より,

$$\sum_{i=1}^{n} (x_i - \mu)^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2, \quad \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n-1} y_i^2$$

であることもわかる. これらの公式は後で正規分布の確率密度函数達の積に適用される.

8.2 準備 $n-2$ 次元単位球面上の微小面積要素

前節の設定をそのまま引き継ぐ.

$\mathbb{R}^n$ における $\mathbb{R}v_n$ の直交補空間上の正規直交座標系として $y_1, \dots, y_{n-1}$ を採用できる. その直交補空間内での $n-2$ 単位球面 $S^{n-2} : y_1^2 + \cdots + y_{n-1}^2 = 1$ 上の積分で使用可能な座標系 $\theta = (\theta_1, \dots, \theta_{n-2})$ を任意に取って, その座標系での球面上の微小面積要素を $d\theta$ と略して書くことにする. このとき, $n-2$ 次元単位球面 S^{n-2} の面積 A_{n-2} は

$$A_{n-2} = \int_{S^{n-2}} d\theta = \frac{2\pi^{(n-1)/2}}{\Gamma((n-1)/2)}$$

と表される.

例: 具体的に座標系を取っていないせいで議論が抽象的過ぎると感じる人は, 以下の座標系を利用すればよい:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \vdots \\ y_{n-1} \end{bmatrix} = \begin{bmatrix} \cos\theta_1 \cos\theta_2 \cos\theta_3 \cdots \cos\theta_{n-2} \\ \sin\theta_1 \cos\theta_2 \cos\theta_3 \cdots \cos\theta_{n-2} \\ \sin\theta_2 \cos\theta_3 \cdots \cos\theta_{n-2} \\ \sin\theta_3 \cdots \cos\theta_{n-2} \\ \ddots \quad \vdots \\ \sin\theta_{n-2} \end{bmatrix} \qquad \begin{pmatrix} -\pi \le \theta_1 \le \pi, \\ -\frac{\pi}{2} \le \theta_2, \dots, \theta_{n-2} \le \frac{\pi}{2} \end{pmatrix}$$

この座標系は「Gauss積分, ガンマ函数, ベータ函数」のノートの「$n-1$ 次元単位球面上の座標を使った面積の計算」の節で紹介したものと本質的に同じものである. この座標系で微小面積要素は次のように表される:

$$d\theta = \cos^0\theta_1 \cos^1\theta_2 \cos^2\theta_3 \cdots \cos^{n-3}\theta_{n-2}\, d\theta_1\, d\theta_2\, d\theta_3 \cdots d\theta_{n-2}.$$

しかし, この例で紹介した複雑な式は後で使用されない. **例終**

このとき, さらに,

$$r = \sqrt{y_1^2 + \cdots + y_{n-1}^2}$$

とおくと, $\mathbb{R}^n$ における積分のための座標として $(\bar{x}, r, \theta)$ を取れて, その微小体積要素 dV は

$$dV = dy_n\, r^{n-2}\, dr\, d\theta = \sqrt{n}\, r^{n-2}\, d\bar{x}\, dr\, d\theta$$

と書ける. さらに, 不偏分散 S^2 に対応する変数 $u = s^2$ を

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

と定めると, $r = \sqrt{(n-1)u}$, $dr = \frac{\sqrt{n-1}}{2}u^{-1/2}\,du$ なので, これを上の式に代入して,

$$dV = \frac{1}{2}\sqrt{n}\,(n-1)^{(n-1)/2}\,u^{(n-1)/2-1}\,d\bar{x}\,du\,d\theta.$$

これの右辺の定数因子の形は重要ではない.

8.3 正規分布の標本分布における $\bar{X}, S^2$ の同時分布の計算

前節までの設定をそのまま引き継ぐ.

このとき, $(X_1, \dots, X_n) \sim \mathrm{Normal}(\mu, \sigma)^n$ の同時密度函数は以上で容易した座標系で

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

を代入すれば

$$\mathrm{const.}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right) = \mathrm{const.}\exp\left(-\frac{(n-1)u}{2\sigma^2} - \frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right)$$

と表される. 確率変数 S^2 に対応する積分変数を u と書くことにしたのであった. ゆえに,

$$\begin{aligned}
&E[f(\bar{X}, S^2)] \\
&= \mathrm{const.}\int_{S^{n-2}}\left(\int_0^\infty\left(\int_{\mathbb{R}} f(\bar{x}, u)\exp\left(-\frac{(n-1)u}{2\sigma^2} - \frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right)u^{(n-1)/2-1}\,d\bar{x}\right)du\right)d\theta \\
&= \mathrm{const.}\int_0^\infty\left(\int_{\mathbb{R}} f(\bar{x}, u)\exp\left(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right)\exp\left(-\frac{(n-1)u}{2\sigma^2}\right)u^{(n-1)/2-1}\,d\bar{x}\right)du.
\end{aligned}$$

const. の部分は $f = 1$ の場合にこれが 1 になることから決定される. この式から S^2 に対応する通常の変数を $u > 0$ と書くことにすれば, $\bar{X}, S^2$ の同時確率密度函数は次の形になることがわかる:

$$\mathrm{const.}\exp\left(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right)\exp\left(-\frac{(n-1)u}{2\sigma^2}\right)u^{(n-1)/2-1}.$$

これは $\bar{X}$ と S^2 が独立であり,

$$\begin{aligned}
\bar{X} &\sim \mathrm{Normal}\left(\mu, \sqrt{\frac{\sigma^2}{n}}\right), \\
S^2 &\sim \mathrm{Gamma}\left(\frac{n-1}{2}, \frac{2\sigma^2}{n-1}\right) = \frac{\sigma^2\,\mathrm{Chisq}(n-1)}{n-1}
\end{aligned}$$

となっていることを意味している. これらは次と同値である:

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim \mathrm{Normal}(0, 1), \qquad \frac{(n-1)S^2}{\sigma^2} \sim \mathrm{Chisq}(n-1).$$

注意: 以上の議論では定数因子 const. の部分は重要でないので, その具体的な式を書かずに最後まで結論を出した. しかし, その部分について数学的な興味を持っている読者もいると思われるので, 以下では定数因子部分を正確に計算してみよう.

まず, 結論から, $\bar{X}, S^2$ それぞれの確率密度函数は以下の形になることがわかる(S^2 に対応する通常の変数を u と書く):

$$\begin{aligned}
p(\bar{x}|n, \mu, \sigma) &= \frac{1}{\sqrt{2\pi\sigma^2/n}}\exp\left(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}\right), \\
p(u = s^2|n, \mu, \sigma) &= \frac{1}{(2\sigma^2/(n-1))^{(n-1)/2}\Gamma((n-1)/2)}\exp\left(-\frac{(n-1)u}{2\sigma^2}\right)u^{(n-1)/2-1}.
\end{aligned}$$

定数因子部分を略さずに上の計算を書くと以下のようになる:

$$
\begin{aligned}
&E[f(\bar{X}, S^2)] \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \frac{1}{2} \sqrt{n}\,(n-1)^{(n-1)/2} \\
&\times \int_{S^{n-2}} \left(\int_0^\infty \left(\int_{\mathbb{R}} f(\bar{x}, u) \exp\left(-\frac{(n-1)u}{2\sigma^2} - \frac{n(\bar{x}-\mu)^2}{2\sigma^2} \right) u^{(n-1)/2-1}\, d\bar{x} \right) du \right) d\theta \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \frac{1}{2} \sqrt{n}\,(n-1)^{(n-1)/2} \frac{2\pi^{(n-1)/2}}{\Gamma((n-1)/2)} \\
&\times \text{const.} \int_0^\infty \left(\int_{\mathbb{R}} f(\bar{x}, u) \exp\left(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2} \right) \exp\left(-\frac{(n-1)u}{2\sigma^2} \right) u^{(n-1)/2-1}\, d\bar{x} \right) du.
\end{aligned}
$$

そして,

$$
\begin{aligned}
&\frac{1}{(2\pi\sigma^2)^{n/2}} \frac{1}{2} \sqrt{n}\,(n-1)^{(n-1)/2} \frac{2\pi^{(n-1)/2}}{\Gamma((n-1)/2)} \\
&= \frac{1}{\sqrt{2\pi\sigma^2}\,(2\pi\sigma^2)^{(n-1)/2}} \sqrt{n}\,(n-1)^{(n-1)/2} \frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)} \\
&= \frac{1}{\sqrt{2\pi\sigma^2/n}} \times \frac{1}{(2\sigma^2/(n-1))^{(n-1)/2}\Gamma((n-1)/2)}
\end{aligned}
$$

なので, つじつまが合っている.

8.4 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

一般に, Z, Y が独立な確率変数達で, Z は標準正規分布に従い, Y が自由度 ν のχ²分布に従うとき, $T = Z/\sqrt{Y/\nu}$ は自由度 ν の t 分布に従うのであった.

この結果を標本平均 $\bar{X}$ の標準化

$$
Z = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim \text{Normal}(0, 1)
$$

とこれと独立に自由度 $\nu = n - 1$ のχ²分布に従う

$$
Y = \frac{(n-1)S^2}{\sigma^2} \sim \text{Chisq}(n-1)
$$

に適用すると, $Y/(n-1) = S^2/\sigma^2$ なので,

$$
T = \frac{Z}{\sqrt{Y/(n-1)}} = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sqrt{\frac{\sigma^2}{S^2}} = \frac{\bar{X} - \mu}{\sqrt{S^2/n}}
$$

が自由度 $n - 1$ の t 分布に従うことがわかる:

$$
T = \frac{\bar{X} - \mu}{\sqrt{S^2/n}} \sim \text{TDist}(\nu).
$$

注意: この T は Z が含む分散パラメータ σ^2 を標本の不偏分散 S^2 に置き換えた形をしており, Z と違って T は σ^2 を使わずに書けている. T は μ のみを含み σ^2 を含まないので, 分散 σ^2 の設定によらない μ に関する結果を得たいときに, この T は使える. 実際, この T は, 正規分布モデルによって, 平均 μ に関する区間推定や検定を σ^2 の存在に煩わされずに行うために使用される.

8.5 正規分布モデルの使用に関する重要な注意: t 分布の使いどころ

以上では, 正規分布のサイズ n の標本分布においては, 標本平均の標準化の式に含まれる分散のパラメータ σ^2 を不偏分散 S^2 で置き換えると, 自由度 $n - 1$ の t 分布に従う確率変数 T が得られることを長くて精密な計算によって示すことができた.

しかし, 次の疑問は重要である:

> このようにがんばって t 分布に従う統計量 T を正規分布から作っても, 現実世界のデータがぴったり正規分布に従っていることはありそうに思えない. 現実世界におけるデータは近似的にも全然正規分布に従っていない場合が多いだろう. t 分布は現実世界における統計分析に役に立つのだろうか?

この疑問への答えを得るためには中心極限定理について学ぶ必要がある. (グラフのプロットの形ではすでに中心極限定理の例をすでに我々は目で確認しているが, 中心極限定理についは後で詳しく説明する.)

以上で構成した t 分布の自由度は $n-1$ だったので, 標本サイズ n が大きな場合にはその　t 分布は標準正規分布で近似される. 実は正規分布の標本分布に限らず, 適当な緩い条件を満たす任意の確率分布の標本分布について, 中心極限定理より, T と同じ式で定義された統計量が標準正規分布で近似されることを示せる. そのような場合には t 分布を使うことに実践的に大して意味がないが, t 分布を使用しても害はない(正規分布以外の標本分布で使っても害がない).

一方, 左右が対称でおとなしめの確率分布の標本分布については, かなり小さな n であっても中心極限定理がよく効いてくれて, 標本平均 $\bar{X}$ の標準化 Z の分布の標準正規分布による近似の誤差が小さくなる. しかし, 実践的な状況では真の分散の値は不明なので, それを不偏分散で置き換える近似が必要になる. その近似は式の上では以上で構成した T を使うのと同じことになる.

その場合の T はもとの分布が正規分布ではないので t 分布には従わない. しかし, 標準正規分布でよく近似された分布の分散を適当に確率的に揺らがせることによって得られた分布に従うことになる. だから, その分布をどんぶり勘定で t 分布で近似してやれば統計分析の安全性を高めることができるだろうと考えられる. t 分布を使用することにはこういう意味がある.

注意・警告: 以上で説明したように, データが正規分布に従っていなくても, 中心極限定理による標本平均の分布の正規分布近似がうまく行っているという条件が成立していれば, t 分布を使った統計分析は実践的に意味を持ち得る. しかし, 極めて多くの解説で「正規分布の仮定が成立していない場合には t 分布を使った統計分析法を使ってはいけない」と解説されている. それは誤解である. 他の解説を読むときには注意して欲しい.

8.6 必修問題: データの t 値の計算例

次のサイズ $n=10$ のデータを $x_1, \ldots, x_n$ と書く:

- データ: 172.5, 174.0, 181.0, 175.0, 180.7, 169.3, 170.6, 172.9, 169.6, 168.7

このデータと $\mu = 170.0$ について以下を計算せよ:

(1) データの標本平均 $\bar{x}$

(2) データの不偏分散 s^2

(3) データの t 値 $t = (\bar{x}-\mu)/\sqrt{s^2/n}$

(4) 自由度 $n-1$ の t 分布に従う確率変数 T に関する $\text{p-value} = 2(1-P(T \geq |t|))$. (この値をP値と呼ぶ.)

P値が小数点以下第3桁まで合っていれば正解とする.

解答例:

$$\bar{x} \approx 173.43, \quad s^2 \approx 19.5779, \quad t \approx 2.45138, \quad \text{p-value} \approx 0.037.$$

解答終

8.6.1 WolframAlphaによる t 値の計算例

(1) mean 172.5, 174.0, 181.0, 175.0, 180.7, 169.3, 170.6, 172.9, 169.6, 168.7 (https://www.wolframalpha.com/input?i=mean+172.5%2C+174.0%2C+181.0%2C+175.0%2C+180.7%2C+169.3%2C+170.6%2C+172.9%2C+169.6%2C+168.7) → 173.43

(2) var 172.5, 174.0, 181.0, 175.0, 180.7, 169.3, 170.6, 172.9, 169.6, 168.7 (https://www.wolframalpha.com/input?i=var+172.5%2C+174.0%2C+181.0%2C+175.0%2C+180.7%2C+169.3%2C+170.6%2C+172.9%2C+169.6%2C+168.7) → 19.5779

(3) T = (173.43 - 170) / sqrt(19.5779/10) (https://www.wolframalpha.com/input?i=%28173.43+-+170%29+%2F+sqrt%2819.5779%2F10%29&lang=ja) → 2.45138

(4) P値 = 2(1 - cdf(StudentTDistribution(9), 2.45138)) (https://www.wolframalpha.com/input?i=2%281+-+cdf%28StudentTDistribution%289%29%2C+2.45138%29%29)

データの簡易要約: 172.5, 174.0, 181.0, 175.0, 180.7, 169.3, 170.6, 172.9, 169.6, 168.7 (https://www.wolframalpha.com/input?i=172.5%2C+174.0%2C+181.0%2C+175.0%2C+180.7%2C+169.3%2C+170.6%2C+172.9%2C+169.6%2C+168.7)

8.6.2 Julia言語による t 値の計算例

In [89]:

```
x = [172.5, 174.0, 181.0, 175.0, 180.7, 169.3, 170.6, 172.9, 169.6, 168.7]
n = length(x)
μ = 170.0
@show x̄ = mean(x)
@show s² = var(x)
@show t = (x̄ - μ)/√(s²/n)
@show pvalue = 2(1 - cdf(TDist(n-1), abs(t)))

ecdf_x = ecdf(x)
P1 = plot(; legend=:topleft)
scatter!(x, fill(-0.1, n), label="data", c=1)
plot!(x -> ecdf_x(x), 165, 185; label="ecdf of data", c=1)
vline!([x̄]; label="x̄", lw=1.5, c=2)

xlim = (-5, 5)
P2 = plot(; legend=:topleft)
plot!(TDist(n-1), xlim...; label="TDist($(n-1))")
vline!([t]; label="t-value", c=2)
vline!([-t]; label="", ls=:dash, c=2)
plot!(TDist(n-1), first(xlim), -abs(t);
    fill=0, c=1, fc=:red, fa=0.5, label="$(round(100pvalue; digits=2))%")
plot!(TDist(n-1), abs(t), last(xlim);
    fill=0, c=1, fc=:red, fa=0.5, label="")

plot(P1, P2; size=(800, 250))
```

```
x̄ = mean(x) = 173.43
s² = var(x) = 19.57788888888888
t = (x̄ - μ) / √(s² / n) = 2.4513831147919563
pvalue = 2 * (1 - cdf(TDist(n - 1), abs(t))) = 0.03667341466610541
```

Out[89]:

P値が約 3.67% であることの意味は上の右図を見ればわかる. 正規分布のサイズ $n = 10$ の標本分布における T 統計量は自由度 $n - 1 = 9$ の t 分布に従うのであった. 仮に与えられたデータの真の分布が正規分布に近ければ(現実の日本人の成人男性の身長の分布もかなり正規分布に近い形をしている), 計算した t 値がどれだけ外れた値になっているかは, 自由度 $n - 1 = 9$ の t 分布を使ってP値(=データから得られる数値以上に極端な値が統計モデル内で生成される確率)を計算すればわかる. このようなことは後で検定と信頼区間に関する別のノートで詳しく解説する予定である.

8.6.3 問題用の擬似データの生成の仕方

問題用の擬似データは Julia v1.7.2 による以下のコードで生成した:

Input:

```
@show VERSION
using Random
Random.seed!(4649373009)
X = round.(rand(Normal(170, 6), 10); digits=1)
@show X;
```

Output:

```
VERSION = v"1.7.2"
X = [172.5, 174.0, 181.0, 175.0, 180.7, 169.3, 170.6, 172.9, 169.6, 168.7]
```

これは日本人の成人男性の身長の擬似データのつもりである. 擬似乱数のシードを意図的にいじって偏ったデータが生成されるようにしてある. 平均 170, 標準偏差 6 の正規分布からであっても, 3.67% 程度の確率でこれ以上に偏ったデータが生成されることが上の計算で分かったことになる.

9 標本分布における T の分布の視覚化

以下に示すグラフ群を見れば, 正規分布以外の標本分布において T 統計量が従う分布がどれだけ自由度 $\nu = n - 1$ の t 分布からずれるかを確認できる. n を大きくすればずれが小さくなって行くことが確認できる.

$T = (\bar{X} - \mu)/\sqrt{S^2/n}$ だけではなく, $Z = (\bar{X} - \mu)/\sqrt{\sigma^2/n}$ (これの期待値と分散はそれぞれ 0 と 1 になる)と, S^2/σ^2 (これの期待値は 1 になる)の分布も横に並べてプロットした. n を大きくすると, T と Z の分布は正規分布の場合と同じ分布に一致して行くが, S^2/σ^2 の分布の場合には $n \to \infty$ で正規分布の場合もそうでない場合も 1 に台を持つデルタ分布に収束するが, 区間推定に利用できるな近似的な一致は成立していない. このように, 正規分布モデルは平均の区間推定については頑健だが, 分散の区間推定については全然そうではない.

In [90]:

```julia
function plot_X̄_S²_T(dist; n = 10, L = 10^6, kwargs...)
    μ, σ = mean(dist), std(dist)

    Z = Vector{Float64}(undef, L)
    Y = similar(Z)
    T = similar(Z)
    tmp = [Vector{eltype(dist)}(undef, n) for _ in 1:Threads.nthreads()]
    Threads.@threads for i in 1:L
        X = rand!(dist, tmp[Threads.threadid()])
        X̄ = mean(X)
        S² = var(X)
        Z[i] = (X̄ - μ)/(σ/√n)
        Y[i] = S²/σ^2
        T[i] = (X̄ - μ)/√(S²/n)
    end

    normal = Normal()
    P1 = plot()
    if dist isa DiscreteUnivariateDistribution
        ecdf_Z = ecdf(Z)
        plot!(; legend=:topleft)
        plot!(x -> ecdf_Z(x), -4, 5; label="√n(X̄-μ)/σ")
        plot!(x -> cdf(normal, x); label="Normal()", ls=:dash)
        title!("cdf for n = $n")
    else
        stephist!(Z; norm=true, label="√n(X̄-μ)/σ")
        plot!(normal; label="Normal()", ls=:dash)
        plot!(; xlim=(-4, 5))
        title!("pdf for n = $n")
    end

    gamma = Chisq(n - 1)/(n - 1)
    xlim = (-0.2, 3.1)
    P2 = plot()
    if dist isa DiscreteUnivariateDistribution
        ecdf_Y = ecdf(Y)
        plot!(; legend=:bottomright)
        plot!(x -> ecdf_Y(x), xlim...; label="S²/σ²")
        plot!(x -> cdf(gamma, x); label="Chisq(ν)/ν", ls=:dash)
        title!("cdf for ν = $(n-1)")
    else
        stephist!(Y; norm=true, label="S²/σ²")
        plot!(gamma, max(0, 1-8/√(n-1)), 1+8/√(n-1); label="Chisq(ν)/ν", ls=:dash)
        plot!(; xlim)
        title!("pdf for ν = $(n-1)")
    end

    tdist = TDist(n - 1)
    P3 = plot()
    if dist isa DiscreteUnivariateDistribution
        ecdf_T = ecdf(T[.!isnan.(T)])
        plot!(; legend=:topleft)
        plot!(x -> ecdf_T(x), -4, 5; label="T")
        plot!(x -> cdf(tdist, x); label="TDist(ν)", ls=:dash)
        title!("cdf for ν = $(n-1)")
    else
        stephist!(T; norm=true, label="T")
        plot!(tdist; label="TDist(ν)", ls=:dash)
        plot!(; xlim=(-4, 5))
        title!("pdf for ν = $(n-1)")
    end

    plot(P1, P2, P3; size = (800, 180), layout=(1, 3))
    plot!(; legendfontsize=8)
end

function plot_X̄_S²_T_4x3(dist; ns = (10, 40, 160, 640),
        m = 4, distxlim = :auto, distylim = :auto,
        size = (800, 750), kwargs...)
    μ, σ², sk, ku = mean(dist), var(dist), myskewness(dist), mykurtosis(dist)
    println(dist)
    @show μ σ²
    println("skewness = ", sk)
    println("kurtosis = ", ku)
    flush(stdout)

    D = plot_dist(dist; m, xlim=distxlim, ylim=distylim, size=(300, 180))
```

```
    display(D)

    PP = []
    for n in ns
        P = plot_X̄_S²_T(dist; n, kwargs...)
        push!(PP, P)
    end
    plot(PP...; size, layout=(4,1),
        titlefontsize=10, guidefontsize=8, tickfontsize=6)
end
```

Out[90]: plot_X̄_S²_T_4x3 (generic function with 1 method)

9.1 正規分布の標本分布における T の分布

In [91]:
```
plot_X̄_S²_T_4x3(Normal(2, 3))
```

```
Normal{Float64}(μ=2.0, σ=3.0)
μ = 2.0
σ² = 9.0
skewness = 0.0
kurtosis = 0.0
```

Out[91]:

9.2 一様分布の標本分布における T の分布

In [92]:
```
plot_X̄_S²_T_4x3(Uniform())
```

```
Uniform{Float64}(a=0.0, b=1.0)
μ = 0.5
σ² = 0.08333333333333333
skewness = 0.0
kurtosis = -1.2
```

Out[92]:

9.3 単峰型のガンマ分布の標本分布における T の分布

In [93]:
```
plot_X̄_S²_T_4x3(Gamma(3, 4))
```

```
Gamma{Float64}(α=3.0, θ=4.0)
μ = 12.0
σ² = 48.0
skewness = 1.1547005383792517
kurtosis = 2.0
```

Out[93]:

9.4 指数分布の標本分布における R の分布

In [94]:
```
plot_X̄_S²_T_4x3(Exponential())
```

```
Exponential{Float64}(θ=1.0)
μ = 1.0
σ² = 1.0
skewness = 2.0
kurtosis = 6.0
```

Out[94]:

9.5 非対称なベータ分布の標本分布における T の分布

In [95]:
```
plot_X̄_S²_T_4x3(Beta(0.1, 0.9); distxlim=(-0.05, 1.05), distylim=(-0.5, 10))
```

```
Beta{Float64}(α=0.1, β=0.9)
μ = 0.1
σ² = 0.045000000000000005
skewness = 2.514157444218836
kurtosis = 5.611111111111112
```

Out[95]:

9.6 非対称な2つ山の混合正規分布の標本分布における T の分布

In [96]:
```
plot_X̄_S²_T_4x3(MixtureModel([Normal(), Normal(20)], [0.95, 0.05]))
```

```
MixtureModel{Normal{Float64}}(K = 2)
components[1] (prior = 0.9500): Normal{Float64}(μ=0.0, σ=1.0)
components[2] (prior = 0.0500): Normal{Float64}(μ=20.0, σ=1.0)

μ = 1.0
σ² = 20.0
skewness = 3.823676241524649
kurtosis = 13.58499999999097
```

Out[96]:

9.7 対称なBernoulli分布の標本分布における T の分布

In [97]:
```
plot_X̄_S²_T_4x3(Bernoulli(0.5))
```

```
Bernoulli{Float64}(p=0.5)
μ = 0.5
σ² = 0.25
skewness = 0.0
kurtosis = -2.0
```

Out[97]:

9.8 非対称なBernoulli分布の標本分布における T の分布

In [98]:
```
plot_X̄_S²_T_4x3(Bernoulli(0.1))
```

```
Bernoulli{Float64}(p=0.1)
μ = 0.1
σ² = 0.09000000000000001
skewness = 2.6666666666666665
kurtosis = 5.111111111111109
```

Out[98]:

9.9 Poisson分布の標本分布における T の分布

In [99]:
```
plot_X̄_S²_T_4x3(Poisson(0.5))
```

```
Poisson{Float64}(λ=0.5)
μ = 0.5
σ² = 0.5
skewness = 1.414213562373095
kurtosis = 2.0
```

Out[99]:

9.10 対数正規分布の標本分布における T の分布

In [100]:
```
plot_X̄_S²_T_4x3(LogNormal())
```

```
LogNormal{Float64}(μ=0.0, σ=1.0)
μ = 1.6487212707001282
σ² = 4.670774270471604
skewness = 6.184877138632554
kurtosis = 110.9363921763115
```

Out[100]:

In [ ]:

```
1
```

大数の法則と中心極限定理

- 黒木玄

- 2022-04-11〜2022-05-21

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)

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
  - 1.1 Markovの不等式
  - 1.2 Chebyshevの不等式
  - 1.3 大数の弱法則
  - 1.4 不偏分散の $n \to \infty$ での挙動
  - 1.5 大数の強法則
    - 1.5.1 Borel-Cantelliの補題
    - 1.5.2 4次以下のモーメントの存在を仮定した場合の大数の強法則
  - 1.6 大数の強法則に関する文献
  - 1.7 問題: 大数の法則に関する誤解
  - 1.8 大数の法則のデモンストレーション
    - 1.8.1 Bernoulli試行の大数の法則
    - 1.8.2 正規分布のサンプルの大数の法則
    - 1.8.3 ガンマ分布のサンプルの大数の法則
    - 1.8.4 大数の法則が成立しない場合: Cauchy分布
    - 1.8.5 大数の法則が成立しない場合: Pólyaの壺
  - 1.9 $E[X_i] = 0$ のときの和 $X_1 + X_2 + \cdots + X_n$ の分布の様子 (中心極限定理に向けて)
- 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プロット
  - 2.3 必修重要問題: 二項分布の正規分布近似を使った確率の近似計算1
    - 2.3.1 WolframAlphaによる計算1
    - 2.3.2 Julia言語による計算1
    - 2.3.3 Julia言語による視覚化1
  - 2.4 必修重要問題: 二項分布の正規分布近似を使った確率の近似計算2
    - 2.4.1 WolframAlphaによる計算
    - 2.4.2 Julia言語による計算2
    - 2.4.3 Julia言語による視覚化2
  - 2.5 連続性補正とmid-P補正
  - 2.6 問題: 二項分布の中心極限定理を使った極限の計算問題
  - 2.7 二項分布の中心極限定理の証明の方針
  - 2.8 (1) Stirlingの公式を使った二項分布の確率質量函数の近似
  - 2.9 注意: Kullback-Leibler情報量とSanovの定理との関係
  - 2.10 (2) 二項分布の確率質量函数から正規分布の密度函数が出て来ること
- 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函数を使った証明
  - 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函数を使った証明
  - 3.5 中心極限定理の収束の速さと歪度と尖度
  - 3.6 中心極限定理のTaylorの定理のみを使う証明
    - 3.6.1 多重積分の書き方

3.6.2 積分剰余項型のTaylorの定理
3.6.3 積分剰余項の別の表示
3.6.4 微分剰余項型のTaylorの定理
3.6.5 Taylorの定理を使った中心極限定理の証明
3.6.6 中心極限定理の収束の速さと歪度と尖度(再)
3.7 問題: 中心極限定理の収束の様子のグラフ
4 デルタ法
4.1 1次のデルタ法と2次のデルタ法
4.2 おまけ: 2次のデルタ法使用時の近似分散と歪度と尖度に関する不等式
4.3 問題: 2次のデルタ法の不偏分散の平方根の期待値の近似への応用
4.4 問題: 1次のデルタ法の正規分布近似への応用
4.5 問題: 1次のデルタ法による分布の逆数の近似
4.6 例: 自由度が大きなχ²分布の逆数の近似
4.7 問題: デルタ法による分布の対数の近似
4.8 例: 自由度が大きなχ²分布の対数の近似
4.9 問題: ベータ分布の正規分布近似
4.10 問題: ベータ分布の正規分布近似の直接的証明
5 各種の確率分布の正規分布近似
5.1 二項分布の正規分布近似
5.2 負の二項分布の正規分布近似
5.3 Poisson分布の正規分布近似
5.4 ガンマ分布の正規分布近似
5.5 χ²分布の正規分布近似
5.6 逆ガンマ分布の正規分布近似
5.7 ベータ分布の正規分布近似
5.8 t 分布の正規分布近似
5.9 ベータプライム分布の正規分布近似
5.10 F 分布の正規分布近似
5.11 超幾何分布の正規分布近似
5.12 ベータ二項分布の正規分布近似
5.13 ベータ負の二項分布の正規分布近似

大数の法則と中心極限定理のイメージ

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

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

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

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

← $\overline{X}_n \approx \mu$ に集中する。 ← 大数の法則

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

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

平均 0, 標準偏差 σ

← 中心極限定理

$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ と μ の差の大きさはだいたい $\frac{\sigma}{\sqrt{n}}$ の大きさであり、

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

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

平均 0, 標準偏差 1

← 中心極限定理

以上のように、大数の法則によって1点に集中する標本平均の分布を

$\sqrt{n}$ 倍（もしくは $\sigma/\sqrt{n}$ 分の1倍）することによってより精密に見ると中心極限定理が得られる。

In [1]:
```julia
ENV["LINES"], ENV["COLUMNS"] = 100, 100
using BenchmarkTools
using Distributions
using Printf
using QuadGK
using Random
Random.seed!(4649373)
using Roots
using SpecialFunctions
using StaticArrays
using StatsBase
using StatsFuns
using StatsPlots
default(fmt = :png, titlefontsize = 10, size = (400, 250))
using SymPy
```

In [2]:
```julia
# Override the Base.show definition of SymPy.jl:
# https://github.com/JuliaPy/SymPy.jl/blob/29c5bfd1d10ac53014fa7fef468bc8deccadc2fc/src/types.

@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::SymbolicObject)
    print(io, as_markdown("\\displaystyle " * sympy.latex(x, mode="plain", fold_short_frac=fal
end
@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::AbstractArray{Sym})
    function toeqnarray(x::Vector{Sym})
        a = join(["\\displaystyle " * sympy.latex(x[i]) for i in 1:length(x)], "\\\\")
        """\\left[ \\begin{array}{r}$a\\end{array} \\right]"""
    end
    function toeqnarray(x::AbstractArray{Sym,2})
        sz = size(x)
        a = join([join("\\displaystyle " .* map(sympy.latex, x[i,:]), "&") for i in 1:sz[1]],
        "\\left[ \\begin{array}{" * repeat("r",sz[2]) * "}" * a * "\\end{array}\\right]"
    end
    print(io, as_markdown(toeqnarray(x)))
end
```

In [3]:
```julia
x ≲ y = x < y || x ≈ y

mypdf(dist, x) = pdf(dist, x)
mypdf(dist::DiscreteUnivariateDistribution, x) = pdf(dist, round(Int, x))

distname(dist::Distribution) = replace(string(dist), r"{.*}" ⇒ "")
myskewness(dist) = skewness(dist)
mykurtosis(dist) = kurtosis(dist)
function standardized_moment(dist::ContinuousUnivariateDistribution, m)
    μ, σ = mean(dist), std(dist)
    quadgk(x -> (x - μ)^m * pdf(dist, x), extrema(dist)...)[1] / σ^m
end
myskewness(dist::MixtureModel{Univariate, Continuous}) = standardized_moment(dist, 3)
mykurtosis(dist::MixtureModel{Univariate, Continuous}) = standardized_moment(dist, 4) - 3
```

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

1 大数の法則

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

以下では大数の法則について解説する.

1.1 Markovの不等式

確率変数 X と任意の $a > 0$ について

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

証明: 函数 $1_{|x|\geq a}(x)$ を

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

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

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

証明終

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

1.2 Chebyshevの不等式

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

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

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

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

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

証明終

1.3 大数の弱法則

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

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

と書くと, 任意の $\varepsilon > 0$ について,

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

注意: この結論は

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

と同値なので標本平均 $\bar{X}_n$ の分布が $n \to \infty$ で μ の近く(任意に小さな $\varepsilon > 0$ に関する μ から距離 ε 未満の範囲)に集中して行くことを意味している. このとき $\bar{X}_n$ は μ に **確率収束** するというが, 標本平均 $\bar{X}_n$ の分布が $n \to \infty$ で μ の近くに集中して行くことを理解していれば, 確率変数の収束に関する細かいことは応用上あまり神経質になる必要はない. だから, このノートでは「大数の弱法則」と「大数の強法則」の区別に神経質にならずに単に「大数の法則」ということが多い.

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

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

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

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

証明終

1.4 不偏分散の $n \to \infty$ での挙動

$X_1, X_2, X_3, \ldots$ は独立同分布確率変数の列であるとし, その最初の n 個の標本平均と不偏分散をそれぞれ次のように書くことにする:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

以下では $\mu = E[X_i]$ がうまく定義されており, $\sigma^2 = E[(X_i-\mu)^2] < \infty$, $E[|X_i-\mu|^3] < \infty$, $E[(X_i-\mu)^4] < \infty$ となっていると仮定する.

「標本分布について」のノートで以下を示していたのであった. X_i 達が共通に従う分布の歪度(わいど, skewness)と尖度(せんど, kurtosis)をそれぞれ

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

と書くと,

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

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

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

このように, 不偏分散 S_n^2 についても大数の弱法則が成立しており, S_n^2 の分布は $n \to \infty$ で σ^2 の近くに集中して行く.

1.5 大数の強法則

この節の内容は無理して読まなくてもよい. 定義していない用語や記号法が登場する. 大数の強法則の取り扱いには測度論的確率論が必要になるが, この一連のノート群では測度論を避ける方針なので, 大数の強法則について詳しく説明することはできない. 以下で紹介する議論は「よく見る議論」のコピー&ペーストに過ぎない. 必要な用語の定義もされていないし, 内容的にも雑である. しかし, 大数の強法則の証明の雰囲気を伝えたいと思ったので, このような説明をしてみることにした.

確率空間とは σ-algebra (= σ-field) $\mathscr{F}$ と確率測度 $\mathscr{F} \ni A \mapsto P(A) \in \mathbb{R}_{\geq 0}$ が与えられた集合 Ω のことである. 以下では確率空間 Ω 上の点を ω と表す. Ω 上の実数値可測函数を確率変数と呼ぶ.

1.5.1 Borel-Cantelliの補題

確率空間 Ω において次が成立する:

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

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

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

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

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

証明終

1.5.2 4次以下のモーメントの存在を仮定した場合の大数の強法則

$X_1, X_2, X_3, \ldots$ は独立同分布確率変数の列であるとし, その最初の n 個の標本平均を $\bar{X}_n = (1/n)\sum_{i=1}^n X_i$ と書くことにする. さらに, $\mu = E[X_i]$ がうまく定義されており, $\sigma^2 = E[(X_i-\mu)^2] < \infty$, $\mu_4 = E[(X_i-\mu)^4] < \infty$ となっていると仮定する. このとき,

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

注意: これは標本平均が μ に確率 1 で収束することを意味している.

証明: 大数の弱法則の場合と同様に補事象の側の同値な主張を示す. すなわち, 確率変数(可測函数)としての標本平均 $\bar{X}_n$ が $n \to \infty$ で μ に収束しない確率が 0 であることを示そう.

確率変数(可測函数)としての最初の n 個の標本平均 $\bar{X}_n$ の ω における値 $\bar{X}_n(\omega)$ が $n \to \infty$ で μ に収束しないという条件は以下のように書き換えられる. すなわち,

$$A_k(\varepsilon) = \{\, \omega \mid |\bar{X}_k(\omega) - \mu| \geq \varepsilon \,\}$$

とおくと, 以下が成立している:

$\bar{X}_n(\omega)$ は $n \to \infty$ で μ に収束しない.
$\iff$ ある $\varepsilon > 0$ が存在して, 任意の n について, ある $k \geq n$ で $|\bar{X}_k(\omega) - \mu| \geq \varepsilon$ を満たすものが存在する.
$\iff$ ある $\varepsilon > 0$ が存在して, $\omega \in \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k(\varepsilon)$.
$\iff$ $\omega \in \bigcup_{L=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k(1/L)$.

この条件の書き換えによって, 任意の $\varepsilon > 0$ について,

$$\sum_{k=1}^{\infty} P(A_k(\varepsilon)) = \sum_{k=1}^{\infty} P(|\bar{X}_k - \mu| \geq \varepsilon) < \infty$$

を示せれば, Borel-Cantelliの補題より $P\left(\bigcap_{n=1}^{\infty} \bigcup_{k\geq n} A_k(\varepsilon)\right) = 0$ が得られ, それより

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

が導かれ, $\bar{X}_n$ が $n \to \infty$ で μ に収束しない確率が 0 であることが示されることがわかる.

必要ならば X_i を $X_i - \mu$ で置き換えることによって, $\mu = 0$ と仮定しても一般性が失われないので, $\mu = 0$ と仮定する.

$\varepsilon > 0$ を任意に取る. このとき,

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

でかつ

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

なので,

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

これで示すべきことが示された.

証明終

注意: 要するに, Borel-Cantelliの補題から, 任意の $\varepsilon > 0$ について,

$$\sum_{n=1}^{\infty} P(|\bar{X}_n - \mu| \geq \varepsilon) < \infty$$

を示せれば, 大数の強法則が証明される. 大数の弱法則の証明で示したことは

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}\frac{1}{n}$$

に過ぎず, $\sum_{n=1} 1/n = \infty$ と調和級数の和は無限大になるので, 大数の弱法則の証明の議論だけで大数の強法則を示すことはできない.

注意: $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$ はRiemannのゼータ函数と呼ばれ, $s = 1$ では調和級数になって無限大に発散するが, $s > 1$ ならば有限の値に収束する. $\zeta(2) = \sum_{n=1}^{\infty} 1/n^2$ を求めよという問題は **Basel問題** (バーゼル問題)と呼ばれ, 18世紀の超大数学者のEulerが解いた. 答えは $\pi^2/6$ になる. その逆数の $6/\pi^2 = 0.6079271\cdots$ は2つの整数の互いに素になる(最大公約数が 1 になる)「確率」に等しい.

In [4]:
```
6/π^2
```

Out[4]: 0.6079271018540267

In [5]:
```
"""
エラトステネスの篩(sieve of Eratosthenes)によって,
以下の条件を満たす長さ `N` のビットベクトル `IsPrime` を作成する:

>`IsPrime = make_isprime(N)` のとき `IsPrime[p]` は `p` が素数なら真になり, 素数でないなら偽に
"""
function make_isprime(N)
    IsPrime = trues(N)
    # 1は素数ではない
    IsPrime[1] = false
    @inbounds for p in 2:isqrt(N)
        if IsPrime[p]
            for i in 2p:p:N
                # 素数の非自明な倍数は素数ではない
                IsPrime[i] = false
            end
        end
    end
    IsPrime
end

"""
N以下の2つの正の整数が互いに素になる確率を篩法で求める.
"""
function prob_mutually_prime(N)
    IsPrime = make_isprime(N)
    C = trues(N, N)
    @inbounds for p in 2:N
        if IsPrime[p]
            for i in p:p:N
                for j in i:p:N
                    # 共通の素因数 p を持つ i, j を除外する
                    C[i, j] = false
                    C[j, i] = false
                end
            end
        end
    end
    c = count(C)
    c/N^2
end
```

Out[5]: prob_mutually_prime

In [6]:
```
@doc make_isprime
```

Out[6]: エラトステネスの篩(sieve of Eratosthenes)によって, 以下の条件を満たす長さ \texttt{N} のビットベクトル \texttt{IsPrime} を作成する:

```
\begin{quote}
\texttt{IsPrime = make\_isprime(N)} のとき \texttt{IsPrime[p]} は \texttt{p} が素数なら真になり, 素数でないなら偽になる.

\end{quote}
```

In [7]:
```
IsPrime = make_isprime(100)
primes = findall(IsPrime)
@show primes;
```

```
primes = [2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 8
3, 89, 97]
```

In [8]:
```
# 1億以下の素数のリストの作成も一瞬で終わる.
@time make_isprime(10^8);
```

```
  1.038035 seconds (3 allocations: 11.921 MiB)
```

In [9]:
```
@time prob_mutually_prime(10000)
```

```
  0.146887 seconds (5 allocations: 11.922 MiB, 21.27% gc time)
```

Out[9]: 0.60794971

In [10]:
```
@time prob_mutually_prime(100000)
```

```
 50.306610 seconds (5 allocations: 1.164 GiB, 0.05% gc time)
```

Out[10]: 0.6079301507

10万以下の2つの正の整数が互いに素な確率を求めてみたが, 確かに理論値の $1/\zeta(2) = 0.6079271\cdots$ に近い値になっている.

1.6 大数の強法則に関する文献

大数の強法則は前節で採用した条件よりもずっと弱い条件のもとで成立している. そのようなことに興味がある人達は以下の文献を参照せよ.

- N. Etemadi, An elementary proof of the strong law of large numbers, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete volume 55, pages 119–122 (1981) [Springer (https://link.springer.com/article/10.1007/BF01013465#citeas)] [pdf (https://people.math.umass.edu/~yaoli/ptrl.pdf)]
- N. Lizia, A simple proof of the strong law of large numbers with rates, Bulletin of the Australian Mathematical Society, Volume 97, Issue 3, June 2018, pp. 513 - 517. [doi (https://doi.org/10.1017/S0004972718000059)] [pdf (http://www.im.ufrj.br/nuno/SLLN.pdf)]

$X_1, X_2, X_3, \ldots$ が同分布な確率変数の列であるとし, 共通の期待値 $\mu = E[X_i]$ のwell-definedness ($E[|X_i|] < \infty$)を仮定する.

前者の文献では, $X_1, X_2, X_3, \ldots$ に関する(独立性よりもずっと弱い)対ごとの独立性の仮定のもとで大数の強法則を証明している(分散の有限性も仮定していない).

後者の文献では, $X_1, X_2, X_3, \ldots$ の独立性と分散の有限性($\sigma^2 = \mathrm{var}(X_i) < \infty$)のもとで, 収束の速さが分かるような大数の強法則の証明を与えている.

1.7 問題: 大数の法則に関する誤解

当たりとはずれが毎回独立ににどちらも 50% の確率で出るルーレットを回し続ける状況を考える. 当たりが出たら, 胴元から1万円もらい, はずれが出たら1万円支払うというルールになっているとする. この賭け事に参加しているある人がこう言った:

> はずれの方が沢山出続けてたくさん負けてしまった! でも, 確率論の大数の法則によって, これからは当たりの方が出易くなって, 負けは帳消しになるだろう.

この発言が間違っていることについて説明せよ.

解答例: 大数の法則は, はずれが沢山出た後に当たりの方が出易くなったり, 当たりが沢山出たあとにはずれが出易くなることによって成立するのではない.

例えば, いきなり 10 回続けてはずれが出てしまったとする.

その後の90回分では, 当たりとはずれが両方とも45回出たとする. そのとき, トータルでの当たりが出たの割合は $55/100 = 55\%$ で 50% に近付く.

はずれ10回後の990回分でも, 当たりとはずれがどちらも同じ回数の495回出たとする. そのとき, トータルでの当たりが出た割合は $505/1000 = 50.5\%$ でさらに 50% に近付く.

そのとき, 最初に10回外れを引いて10万円負けていたその人は, 大数の法則が予想する結果に近付いているにもかかわらず, 1000回の勝負で最初に負けた10万円を取り戻せていない.

大数の法則は単に公平なギャンブルを続けると勝ちの回数の割合と負けの回数の割合が 50% に近付くことを主張しているだけで，沢山負けた後に勝ち易くなったり，その逆が起こることを何も主張していない．

解答終

他にも沢山の正しい解答例を作れると思います．

1.8 大数の法則のデモンストレーション

In [11]:

```
function plot_law_of_large_numbers(dist, N;
        randfunc = rand, μ = mean(dist), kwargs...)
    X = randfunc(dist, N)
    X̄ = cumsum(X) ./ (1:N)
    plot(1:N, X̄; label="")
    hline!([μ]; label="")
    plot!(; kwargs...)
end

function plot_law_of_large_numbers_6x3(dist, N;
        randfunc = rand, μ = mean(dist), seed = 4649363, title = "", kwargs...)
    Random.seed!(seed)
    PP = []
    for _ in 1:18
        P = plot_law_of_large_numbers(dist, N; randfunc, μ, kwargs...)
        push!(PP, P)
    end
    P0 = plot(; title,
        framestyle=nothing, showaxis=false, ticks=false, margin=0Plots.mm)
    layout = @layout [a{0.005h}; grid(6, 3)]
    plot(P0, PP...; size=(800, 800), layout,
        titlefontsize=12, guidefontsize=8, tickfontsize=6)
end
```

Out[11]: plot_law_of_large_numbers_6x3 (generic function with 1 method)

1.8.1 Bernoulli試行の大数の法則

In [12]:

```
bar(x → pdf(Bernoulli(0.3), x), 0:1; alpha=0.5, label="",
    title="Bernoulli(0.3)", xtick=0:1, ytick=0:0.1:1)
```

Out[12]:

In [13]:

```
plot_law_of_large_numbers_6x3(Bernoulli(0.3), 2^12;
    title="Law of large numbers for Bernoulli(0.3)", ylim=(0.1, 0.5))
```

Out[13]:

1.8.2 正規分布のサンプルの大数の法則

In [14]:
```
plot(Normal(2, 3); label="", title="Normal(2, 3)")
```

Out[14]:

In [15]:
```
plot_law_of_large_numbers_6x3(Normal(2, 3), 2^12;
    title="Law of large numbers for Normal(2, 3)", ylim=(1, 3))
```

Out[15]:

1.8.3 ガンマ分布のサンプルの大数の法則

In [16]:
```
plot(Gamma(2, 3); label="", title="Gamma(2, 3)")
```

Out[16]:

In [17]:
```
plot_law_of_large_numbers_6x3(Gamma(2, 3), 2^12;
    title="Law of large numbers for Gamma(2, 3)", ylim=(5, 7))
```

Out[17]:

1.8.4 大数の法則が成立しない場合: Cauchy分布

確率密度函数が

$$p(x) = \frac{1}{\pi}\frac{1}{1+x^2}$$

のCauchy分布 Cauchy() に従う確率変数 X については

$$E[|X|] = \frac{2}{\pi}\int_0^\infty \frac{x}{1+x^2}\,dx = \infty$$

なので, 期待値 $E[X]$ が定義されない. Cauchy分布の場合には大数の法則は成立しない.

In [18]:
```
plot(Cauchy(), -10, 10; label="", title="Cauchy()")
```

Out[18]:

In [19]:
```
plot_law_of_large_numbers_6x3(Cauchy(), 2^12; μ = 0.0,
    title="Law of large numbers for Cauchy() does not hold", ylim=(-10, 10))
```

Out[19]:

注意: $X_1, X_2, \ldots, X_n$ がCauchy分布のサイズ n の標本であるとき, その標本平均 $\bar{X}$ もCauchy分布に従うことを示せる. そのことはCauchy分布の特性函数が

$$\varphi(t) = E[e^{itX}] = e^{-|t|} \quad (X \sim \mathrm{Cauchy}())$$

となることを認めれば容易に確認できる(この公式のよくある証明では複素解析を使う). 実際, これを認めると,

$$E[e^{it\bar{X}}] = \phi\left(\frac{t}{n}\right)^n = (e^{-|t/n|})^n = e^{-|t|} = \varphi(t) = E[e^{itX}].$$

$\bar{X}$ と $X \sim \mathrm{Cauchy}()$ の特性函数が等しいので $\bar{X} \sim \mathrm{Cauchy}()$.

In [20]:
```
function plot_cauchy(n; dist=Cauchy(), L=10^6, xlim=(-10, 10), kwargs...)
    X̄ = Vector{Float64}(undef, L)
    tmp = [Vector{Float64}(undef, n) for _ in 1:Threads.nthreads()]
    Threads.@threads for i in 1:L
        X = rand!(dist, tmp[Threads.threadid()])
        X̄[i] = mean(X)
    end
    a, b = quantile.(Ref(X̄), (0.001, 0.999))
    stephist(X̄; bin=a:0.1:b, norm=true, label="X̄")
    plot!(Cauchy(), xlim...; label="Cauchy()", ls=:dash)
    title!("Sample means of n = $n size samples of Cauchy()")
    plot!(; xlim, kwargs...)
end
```

Out[20]: plot_cauchy (generic function with 1 method)

In [21]:
```
plot_cauchy(10)
```

Out[21]:

1.8.5 大数の法則が成立しない場合: Pólyaの壺

$\alpha, \beta > 0$ であると仮定する. パラメータ (α, β) から 1 の出る確率が $\alpha/(\alpha+\beta)$ と, 0 が出る確率が $\beta/(\alpha+\beta)$ と決まっているとする. そして, 1 が出たら, パラメータが $(\alpha, \beta) \mapsto (\alpha+1, \beta)$ と更新され, 0 が出たら, パラメータが $(\alpha, \beta) \mapsto (\alpha, \beta+1)$ と更新されると仮定する. このようにして $1, 0$ の列 $X_1, X_2, X_3, \ldots$ がランダムに得られる試行を **Pólyaの壺** の試行と呼ぶことにする.

Pólyaの壺では 1 が出るとその次に 1 が出る確率は上昇し, 0 が出るとその次に 0 が出る確率は上昇する. **Pólyaの壺** の試行では独立性の仮定が満たされていない. 実際, **Pólyaの壺** の試行の n 回分の出目が $(x_1, \ldots, x_n) \in \{1, 0\}^n$ になる確率は, $x_1, \ldots, x_n$ の中に 1 がちょうど k 個含まれているとき,

$$P(x_1, \ldots, x_n | \alpha, \beta) = \binom{n}{k} \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)\ \beta(\beta+1)\cdots(\beta+n-k-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+n-1)}$$

と書ける. (「確率分布達の解釈」のノートに詳しい説明がある.)

しかし, $n-1$ 回目までの出目に条件を付けないとき, n 番目に 1 が出る確率は n によらず $\alpha/(\alpha+\beta)$ のまま一定であることを示せる. すなわち, 各 X_n が従う確率分布は n によらずにどれも成功確率 $\alpha/(\alpha+\beta)$ のBernoulli分布になる. **Pólyaの壺** の試行では同分布の仮定は満たされている.

Pólyaの壺 の試行で得られる $1, 0$ の列 $X_1, X_2, X_3, \ldots$ については大数の法則は成立しない. その n 番目までの標本平均 $\bar{X}_n$ はベータ二項分布 $\mathrm{BetaBinomial}(n, \alpha, \beta)$ に従い, n が大きなときに $\bar{X}_n$ はベータ分布 $\mathrm{Beta}(\alpha, \beta)$ に近似的に従う. そうなることは,

$$P(x_1, \ldots, x_n | \alpha, \beta) = \int_0^1 \prod_{i=1}^n \left(p^{x_i}(1-p)^{1-x_i}\right) \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)}\, dp$$

なので, **Pólyaの壺** の試行が成功確率 p がベータ分布 $\mathrm{Beta}(\alpha, \beta)$ に従ってランダムに決まっているようなBernoulli試行であることから分かる.

In [22]:

```
# pがベータ分布に従ってランダムに決まっているベルヌーイ試行

plot_law_of_large_numbers_6x3(Beta(4, 6), 2^10;
    randfunc = (dist, n) → (p = rand(dist); rand(Bernoulli(p), n)),
    title="Pólya's urn trials for α=4, β=6", ylim=(-0.02, 1.02), ytick=0:0.2:1)
```

Out[22]:

In [23]:

```julia
# 富むものがさらに富み，貧しいものがさらに貧しくなるポリアの壺試行

function rand_polyaurn(beta::Beta, n)
    α, β = Distributions.params(beta)
    X = Vector{Int8}(undef, n)
    for i in 1:n
        x = rand(Bernoulli(α/(α+β)))
        X[i] = x
        if x == 1
            α += 1
        else
            β += 1
        end
    end
    X
end

plot_law_of_large_numbers_6x3(Beta(4, 6), 2^10;
    randfunc = rand_polyaurn, title="Pólya's urn trials for α=4, β=6",
    ylim=(-0.02, 1.02), ytick=0:0.2:1)
```

Out[23]:

In [24]:

```
function plot_polyaurn(n, α, β, bin=:auto; L=10^6, kwargs...)
    X̄ = Vector{Float64}(undef, L)
    tmp = [Vector{Int}(undef, n) for _ in 1:Threads.nthreads()]
    Threads.@threads for i in 1:L
        p = rand(Beta(α, β))
        X = rand!(Bernoulli(p), tmp[Threads.threadid()])
        X̄[i] = mean(X)
    end
    stephist(X̄; norm=true, bin, label="X̄")
    plot!(Beta(α, β), 0, 1; label="Beta(α, β)", ls=:dash)
    title!("Pólya's urn for n=$n, α=$α, β=$β")
    plot!(; xtick=0:0.1:1, kwargs...)
end
```

Out[24]: plot_polyaurn (generic function with 2 methods)

In [25]:

```
plot(plot_polyaurn.((10, 30), 4, 6, (0:0.1:1, 0:0.1:1))...; size=(800, 250))
```

Out[25]:

In [26]:

```
plot(plot_polyaurn.((100, 300), 4, 6, (0:0.01:1, 0:0.01:1))...; size=(800, 250))
```

Out[26]:

1.9 $E[X_i] = 0$ のときの和 $X_1 + X_2 + \cdots + X_n$ の分布の様子 (中心極限定理に向けて)

共通の期待値 0 と分散 σ^2 を持つ独立同分布確率変数列 $X_1, X_2, X_3, \ldots$ を沢山生成して, $X_1 + X_2 + \cdots + X_n$ の動きを同時プロットしてみよう. $X_1 + X_2 + \cdots + X_n$ の動きを **ランダムウォーク** (乱歩)と呼ぶことにする.

ランダムウォークは大雑把に $\pm 2\sigma\sqrt{n}$ の範囲に分布する.

この事実を多くのグラフで視覚的に確認した人達は中心極限定理に限りなく近付いている.

$X_1 + X_2 + \cdots + X_n$ がちらばる速さは $\sqrt{n}$ に比例するので, $(X_1 + X_2 + \cdots + X_n)/\sqrt{n}$ は一定の範囲に分布すると考えられる. その一定の範囲内の分布が正規分布で近似されるというのが中心極限定理の内容である.

In [27]:

```
function plot_randomwalk!(dist, N;
        randfunc = rand, μ = mean(dist), c=1, lw=0.3, alpha=0.3, kwargs...)
    X = randfunc(dist - μ, N)
    X̄ = cumsum(X)
    plot!(1:N, X̄; label="", c, lw, alpha, kwargs...)
end

function plot_randomwalks(dist, N;
        randfunc = rand, μ = mean(dist), σ = std(dist),
        seed = 4649363,
        title = "Random walks for $(distname(dist)) - $μ",
        kwargs...)
    Random.seed!(seed)
    P = plot(; title, )
    for _ in 1:100
        P = plot_randomwalk!(dist, N; randfunc, μ, kwargs...)
    end
    x = range(0, N, 1000)
    y = @. 2σ * √x
    plot!(x,  y; c=2, label="±2σ√n")
    plot!(x, -y; c=2, label="")
    plot!(; xlabel="n", ylabel="X₁+X₂+⋯+Xₙ")
    plot!(; legend=:topleft, size=(500, 320))
end
```

Out[27]: plot_randomwalks (generic function with 1 method)

In [28]:

```
plot_randomwalks(Bernoulli(0.3), 2^12)
```

Out[28]:

In [29]:

```
plot_randomwalks(Normal(2, 3), 2^12)
```

Out[29]:

In [30]:
```
plot_randomwalks(Gamma(2, 3), 2^12)
```

Out[30]:

In [31]:
```
plot_randomwalks(Cauchy(), 2^10; μ=0, σ=10, ylim=(-1000, 1000), alpha=0.7)
```

Out[31]:

In [32]:
```
plot_randomwalks(Beta(4, 6), 2^10;
    μ = 0, σ = 1,
    randfunc = (dist, n) → (p = rand(dist); rand(Bernoulli(p) - mean(dist), n)),
    title="Random walks for Pólya's urn with α=4, β=6", alpha=0.7)
```

Out[32]:

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

二項分布の中心極限定理には以下のような様々な言い換えがある. $0 < p < 1$ と仮定し, n は十分に大きいと仮定する. そのとき,

(1) $K \sim \mathrm{Binomial}(n,p)$ ならば K の分布は平均 np, 分散 $np(1-p)$ の正規分布で近似される.

(2) $K \sim \mathrm{Binomial}(n,p)$ ならば $\dfrac{K}{n}$ の分布は平均 p, 分散 $p(1-p)/n$ の正規分布で近似される.

(3) $K \sim \mathrm{Binomial}(n,p)$ ならば $\dfrac{K-np}{\sqrt{n}}$ の分布は平均 0, 分散 $p(1-p)$ の正規分布で近似される.

(4) $K \sim \mathrm{Binomial}(n,p)$ ならば $\dfrac{K-np}{\sqrt{np(1-p)}}$ の分布は標準正規分布で近似される.

例えば, (1)の結果は, $K \sim \mathrm{Binomial}(n,p)$, $X \sim \mathrm{Normal}\left(np, \sqrt{np(1-p)}\right)$ のとき,

$$P(K \le a) \approx P(X \le a)$$

という近似を使えることを意味する. (2),(3),(4)についても同様のことが言える. これを使うと, 二項分布における確率を対応する正規分布の確率で置き換えて近似計算できる. 二項分布に限らず, 正規分布で近似できる確率分布であれば, どのような分布であっても確率の計算を正規分布という一種類の分布における確率で近似的に計算できる.

例えば, (4)の結果は, $K \sim \mathrm{Binomial}(n,p)$ のとき,

$$E\left[f\left(\frac{K-np}{\sqrt{np(1-p)}}\right)\right] = \sum_{k=0}^{n} f\left(\frac{k-np}{\sqrt{np(1-p)}}\right)\binom{n}{k}p^k(1-p)^{n-k}$$

が, $n \to \infty$ で, $Z \sim \mathrm{Normal}(0,1)$ に関する

$$E[f(Z)] = \int_{-\infty}^{\infty} f(z)\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz.$$

に収束することと同じことだと考えてよい. (3)についても同様のことが言える.

2.2 二項分布の中心極限定理のプロット

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 [33]:

```
function plot_binomial_clt(n, p, s = 1; c = 4.5)
    μ, σ = n*p, √(n*p*(1-p))
    xlim = (μ-c*σ, μ+c*σ)
    plot(x -> mypdf(Binomial(n, p), x), xlim...; label="Binomial(n,p)")
    plot!(Normal(μ, σ), xlim...; label="Normal(μ,σ)", lw=2)
    # plot!(x -> mypdf(Poisson(μ), x), xlim...; label="Poisson(μ)", ls=:dash)
    title!("n = $n, p = $p, μ=np, σ²=np(1-p)")
    plot!(; xtick=0:s:n)
end
```

Out[33]:

```
plot_binomial_clt (generic function with 2 methods)
```

In [34]:
```
plot(plot_binomial_clt(10, 0.3), plot_binomial_clt(10, 0.1); size=(800, 250))
```

Out[34]:

In [35]:
```
plot(plot_binomial_clt(40, 0.3, 2), plot_binomial_clt(40, 0.1); size=(800, 250))
```

Out[35]:

In [36]:
```
plot(plot_binomial_clt(160, 0.3, 5), plot_binomial_clt(160, 0.1, 2); size=(800, 250))
```

Out[36]:

2.3 必修重要問題: 二項分布の正規分布近似を使った確率の近似計算1

確率変数 K は二項分布 Binomial$(100, 1/2)$ に従っていると仮定する.

(1) 確率 $P(K \leq 40)$ を小数点以下第5桁まで求めよ.

(2) 二項分布 Binomial$(100, 1/2)$ と同じ期待値と分散を持つ正規分布に従う確率変数 X について確率 $P(X \leq 40)$ を小数点以下第5桁まで求めよ.

解答例:

(1) $K \sim \mathrm{Binomial}(100, 1/2)$ のとき,

$$P(K \leq 40) \approx 0.02844$$

(2) 一般に $\mathrm{Binomial}(n, p)$ の平均と分散はそれぞれ $\mu = np$, $\sigma^2 = np(1-p)$ なので, $n = 100$, $p = 1/2$ の場合は $\mu = 50$, $\sigma^2 = 25$, $\sigma = 5$ なので, $X \sim \mathrm{Normal}(50, 5)$ で

$$P(X \leq 40) \approx 0.02275$$

解答終

2.3.1 WolframAlphaによる計算1

(1) [cdf(BinomialDistribution(n, p), 40) where n=100, p=0.5 (https://www.wolframalpha.com/input?i=cdf%28BinomialDistribution%28n%2C+p%29%2C+40%29+where+n%3D100%2C+p%3D0.5)](https://www.wolframalpha.com/input?i=cdf%28BinomialDistribution%28n%2C+p%29%2C+40%29+where+n%3D100%2C+p%3D0.5)

(2) [cdf(NormalDistribution(50, 5), 40) (https://www.wolframalpha.com/input?i=cdf%28NormalDistribution%2850%2C+5%29%2C+40%29)](https://www.wolframalpha.com/input?i=cdf%28NormalDistribution%2850%2C+5%29%2C+40%29)

2.3.2 Julia言語による計算1

In [37]:
```
n, p = 100, 1/2
```
Out[37]:
```
(100, 0.5)
```

In [38]:
```
# (1)
p_B = cdf(Binomial(n, p), 40)
```
Out[38]:
```
0.02844396682049039
```

In [39]:
```
# (2)
q_N = cdf(Normal(n*p, √(n*p*(1-p))), 40)
```
Out[39]:
```
0.022750131948179205
```

2.3.3 Julia言語による視覚化1

In [40]:
```
n, p = 100, 1/2
```
Out[40]:
```
(100, 0.5)
```

In [41]:
```
# (1)
plot(x -> mypdf(Binomial(n, p), x), 30, 70; label="", c=1)
vline!([n*p]; label="np", c=1)
plot!(x -> mypdf(Binomial(n, p), x), 30, 40.5;
    label="", fillrange=0, c=1, alpha=0.5)
prob = cdf(Binomial(n, p), 40)
annotate!(35, 0.01, text(@sprintf("%0.5f", prob), 10, :blue))
vline!([40]; label="", c=:gray)
title!("Binomial(n = $n, p = $p)")
```
Out[41]:

In [42]:

```
# (2)
plot(x -> mypdf(Normal(n*p, √(n*p*(1-p))), x), 30, 70; label="", c=:blue, ls=:dash)
vline!([n*p]; label="np", ls=:dashdot, c=:blue)
plot!(x -> mypdf(Normal(n*p, √(n*p*(1-p))), x), 30, 40;
    label="", fillrange=0, c=:blue, alpha=0.5)
prob = cdf(Normal(n*p, √(n*p*(1-p))), 40)
annotate!(35, 0.01, text(@sprintf("%0.5f", prob), 10, :blue))
vline!([40]; label="", c=:gray)
title!("Normal(μ = $n×$p, σ = √$(n*p*(1-p)))")
```

Out[42]:

In [43]:

```
# (1) and (2)
plot(x -> mypdf(Binomial(n, p), x), 30, 70; label="", c=1)
vline!([n*p]; label="np", c=1)

plot!(x -> mypdf(Normal(n*p, √(n*p*(1-p))), x), 30, 70;
    label="", c=:blue, ls=:dash)
#vline!([n*p]; label="np", ls=:dashdot, c=:blue)

vline!([40]; label="", c=:gray)
title!("Binomial(n = $n, p = $p) and normal approx")
```

Out[43]:

2.4 必修重要問題: 二項分布の正規分布近似を使った確率の近似計算2

$0 < p < 1$ であると仮定し, K は二項分布 $\mathrm{Binomial}(100, p)$ に従う確率変数であるとする.

(1) $P(K \leq 40) = 0.025$ を満たす p の値 p_U を小数点以下第4桁まで求めよ.

(2) $P(K \geq 40) = 0.025$ を満たす p の値 p_L を小数点以下第4桁まで求めよ.

$\hat{p} = 40/100 = 0.4$ とおく. X は K と同じ平均 np と推定された分散 $n\hat{p}(1-\hat{p})$ を持つ正規分布に従う確率変数であるとする.

(3) $P(X \leq 40) = 0.025$ を満たす p の値 q_U を小数点以下第4桁まで求めよ.

(4) $P(X \geq 40) = 0.025$ を満たす p の値 q_L を小数点以下第4桁まで求めよ.

注意: これらによって計算される区間 $[p_L, p_U]$ と $[q_L, q_U]$ はデータ「100 回中 40 回当たり」というデータから得られる p の 95% 信頼区間と呼ばれる. 前者は **Clopper-Pearsonの信頼区間** と, 後者は **Waldの信頼区間** と呼ばれる.

解答例: $n = 100$, $k = 40$ とおく.

累積分布函数を cdf と書き, その逆函数の分位点函数を quantile と書くことにする: X が分布 D に従う確率変数であるとき,

累積分布関数を cdf と書き，その逆関数が分位関数を quantile と書くことにする．X が分布 D に従う確率変数であるとき，

$$\begin{aligned}
&\mathrm{cdf}(D, x) = P(X \le x),\\
&\mathrm{quantile}(D, a) = (\text{the value } x \text{ satisfying } P(X \le x) = a).
\end{aligned}$$

以上の記号法のもとで，

$$\begin{aligned}
F_K(p) &= P(K \le k) = \mathrm{cdf}(\mathrm{Binomial}(n, p), k)\\
&= 1 - \mathrm{cdf}(\mathrm{Beta}(k+1, n-k), p),\\
G_K(p) &= P(K \ge k) = 1 - \mathrm{cdf}(\mathrm{Binomial}(n, p), k-1)\\
&= \mathrm{cdf}(\mathrm{Beta}(k, n-k+1), p)
\end{aligned}$$

について，

$$\begin{aligned}
F_K^{-1}(\alpha/2) &= \mathrm{quantile}(\mathrm{Beta}(k+1, n-k), 1-\alpha/2),\\
G_K^{-1}(\alpha/2) &= \mathrm{quantile}(\mathrm{Beta}(k, n-k+1), \alpha/2).
\end{aligned}$$

これらを $n = 100$, $k = 40$, $\alpha = 0.05$ に適用すると，

(1) $p_U \approx 0.5028$

(2) $p_L \approx 0.3033$

$X \sim \mathrm{Normal}(np, \sigma)$ のとき，

$$\begin{aligned}
F_X(p) &= P(X \le k) = \mathrm{cdf}(\mathrm{Normal}(np, \sigma), k)\\
&= \mathrm{cdf}(\mathrm{Normal}(0, 1), (k - np)/\sigma),\\
G_X(p) &= P(X \ge k) = 1 - \mathrm{cdf}(\mathrm{Normal}(np, \sigma), k)\\
&= 1 - \mathrm{cdf}(\mathrm{Normal}(0, 1), (k - np)/\sigma)
\end{aligned}$$

について

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

とおくと，$\mathrm{quantile}(\mathrm{Normal}(0, 1), \alpha/2) = -z_{\alpha/2}$ なので，

$$\begin{aligned}
F_X^{-1}(\alpha/2) &= \frac{k - \sigma\,\mathrm{quantile}(\mathrm{Normal}(0, 1), \alpha/2)}{n} = \frac{k + z_{\alpha/2}\sigma}{n}\\
G_X^{-1}(1-\alpha/2) &= \frac{k - \sigma\,\mathrm{quantile}(\mathrm{Normal}(0, 1), 1-\alpha/2)}{n} = \frac{k - z_{\alpha/2}\sigma}{n}
\end{aligned}$$

$n = 100$, $\hat{p} = 0.4$ のとき，

$$\hat{\sigma}^2 = n\hat{p}(1 - \hat{p}) = 24.$$

であり，上の公式を $n = 100$, $k = 40$, $\sigma = \hat{\sigma}$ の場合に使うと，

(3) $q_U \approx 0.4960$

(4) $q_L \approx 0.3040$

解答終

2.4.1 WolframAlphaによる計算

(1) quantile(BetaDistribution(41, 60), 0.975) (https://www.wolframalpha.com/input?i=quantile%28BetaDistribution%2841%2C+60%29%2C+0.975%29)

(1)検算 cdf(BinomialDistribution(100, 0.502791), 40) (https://www.wolframalpha.com/input?i=cdf%28BinomialDistribution%28100%2C+0.502791%29%2C+40%29)

(2) quantile(BetaDistribution(40, 61), 0.025) (https://www.wolframalpha.com/input?i=quantile%28BetaDistribution%2840%2C+61%29%2C+0.025%29)

(2)検算 1 - cdf(BinomialDistribution(100, 0.303295), 39) (https://www.wolframalpha.com/input?i=1+-+cdf%28BinomialDistribution%28100%2C+0.303295%29%2C+39%29)

quantile(NormalDistribution(0, 1), 0.975) (https://www.wolframalpha.com/input?i=quantile%28NormalDistribution%280%2C+1%29%2C+0.975%29) → 1.95996

(3) (40 + 1.95996 * sqrt(24))/100 (https://www.wolframalpha.com/input?i=%2840+%2B+1.95996+*+sqrt%2824%29%29%2F100)

(3)検算 cdf(NormalDistribution(49.60180, sqrt(24)), 40) (https://www.wolframalpha.com/input?i=cdf%28NormalDistribution%2849.60180%2C+sqrt%2824%29%29%2C+40%29)

(4) (40 - 1.95996 * sqrt(24))/100 (https://www.wolframalpha.com/input?i=%2840+-+1.95996+*+sqrt%2824%29%29%2F100)

(4)検算 1 - cdf(NormalDistribution(30.39820, sqrt(24)), 40) (https://www.wolframalpha.com/input?i=1+-+cdf%28NormalDistribution%2830.39820%2C+sqrt%2824%29%29%2C+40%29)

2.4.2 Julia言語による計算2

記号法: `cdf(d, x)` は分布 `d` の累積分布函数で `ccdf(d, x) = 1 - cdf(d, x)` .

`cdf(Binomial(n, p), k) == ccdf(Beta(k+1, n-k), p)` の確認:

In [44]:
```
n, p = 8, 0.3
[
    [cdf(Binomial(n, p), k) for k in 0:n-1]'
    [ccdf(Beta(k+1, n-k), p) for k in 0:n-1]'
]
```

Out[44]:
```
2×8 Matrix{Float64}:
 0.057648  0.255298  0.551774  0.805896  0.942032  0.988708  0.99871  0.999934
 0.057648  0.255298  0.551774  0.805896  0.942032  0.988708  0.99871  0.999934
```

`ccdf(Binomial(n, p), k-1) == cdf(Beta(k, n-k+1), p)` の確認:

In [45]:
```
n, p = 8, 0.3
[
    [ccdf(Binomial(n, p), k-1) for k in 1:n]'
    [cdf(Beta(k, n-k+1), p) for k in 1:n]'
]
```

Out[45]:
```
2×8 Matrix{Float64}:
 0.942352  0.744702  0.448226  0.194104  0.0579676  0.0112922  0.00129033  6.561e-5
 0.942352  0.744702  0.448226  0.194104  0.0579676  0.0112922  0.00129033  6.561e-5
```

In [46]:
```
n, k = 100, 40
α = 0.05
```

Out[46]: 0.05

In [47]:
```
# (1)
p_U = quantile(Beta(k+1, n-k), 1 - α/2)
```

Out[47]: 0.5027908495776652

In [48]:
```
# (1)検算
cdf(Binomial(n, p_U), k)
```

Out[48]: 0.02499999999999997

In [49]:
```
# (2)
p_L = quantile(Beta(k, n-k+1), α/2)
```

Out[49]: 0.3032947687028774

In [50]:
```
# (2)検算
ccdf(Binomial(n, p_L), k-1)
```

Out[50]: 0.025000000000000005

In [51]:
```
p̂ = k/n
```

Out[51]: 0.4

In [52]:
```
σ̂² = n*p̂*(1-p̂)
```

Out[52]: 24.0

In [53]:
```
σ̂ = √σ̂²
```

Out[53]: 4.898979485566356

In [54]:
```
# z_{α/2}
z = quantile(Normal(0, 1), 1 - α/2)
```
Out[54]: 1.9599639845400576

In [55]:
```
# (3)
q_U = (k + z*σ̂)/n
```
Out[55]: 0.4960182335271064

In [56]:
```
# (3)検算
cdf(Normal(n*q_U, σ̂), k)
```
Out[56]: 0.024999999999999776

In [57]:
```
# (4)
q_L = (k - z*σ̂)/n
```
Out[57]: 0.3039817664728936

In [58]:
```
# (4)検算
1 - cdf(Normal(n*q_L, σ̂), k)
```
Out[58]: 0.0249999999999998

2.4.3 Julia言語による視覚化2

In [59]:
```
n, k, α = 100, 40, 0.05
p̂ = k/n
σ̂² = n*p̂*(1-p̂)
σ̂ = √σ̂²

p_U = quantile(Beta(k+1, n-k), 1 - α/2)
p_L = quantile(Beta(k, n-k+1), α/2)
z = quantile(Normal(0, 1), 1 - α/2)
q_U = (k + z*σ̂)/n
q_L = (k - z*σ̂)/n;
```

In [60]:
```
# (1)
plot(x -> mypdf(Binomial(n, p_U), x), 10, 70; label="", c=1)
vline!([n*p_U]; label="n p_U", c=1)
plot!(x -> mypdf(Binomial(n, p_U), x), 10, 40.5;
    label="", fillrange=0, c=1, alpha=0.5)
annotate!(35, 0.01, text("2.5%", 10, :blue))
vline!([40]; label="", c=:gray)
title!("Binomial(n = $n, p_U = $(@sprintf "%.4f" p_U))")
plot!(; size=(600, 250))
```

Out[60]:

In [61]:

```
# (2)
plot(x -> mypdf(Binomial(n, p_L), x), 10, 70; label="", c=2)
vline!([n*p_L]; label="n p_L", c=2)
plot!(x -> mypdf(Binomial(n, p_L), x), 39.5, 70;
    label="", fillrange=0, c=2, alpha=0.5)
annotate!(45, 0.01, text("2.5%", 10, :red))
vline!([40]; label="", c=:gray)
title!("Binomial(n = $n, p_L = $(@sprintf "%.4f" p_L))")
plot!(; size=(600, 250))
```

Out[61]:

In [62]:

```
# (1) and (2)
plot(x -> mypdf(Binomial(n, p_U), x), 10, 70; label="", c=1)
vline!([n*p_U]; label="n p_U", c=1)
plot!(x -> mypdf(Binomial(n, p_U), x), 10, 40.5;
    label="", fillrange=0, c=1, alpha=0.5)
annotate!(35, 0.01, text("2.5%", 10, :blue))

plot!(x -> mypdf(Binomial(n, p_L), x), 10, 70; label="", c=2)
vline!([n*p_L]; label="n p_L", c=2)
plot!(x -> mypdf(Binomial(n, p_L), x), 39.5, 70;
    label="", fillrange=0, c=2, alpha=0.5)
annotate!(45, 0.01, text("2.5%", 10, :red))

vline!([40]; label="", c=:gray)
title!("[p_L, p_U] = [$(@sprintf "%.4f" p_L), $(@sprintf "%.4f" p_U)]")
plot!(; size=(600, 250))
```

Out[62]:

In [63]:

```
# (3)
plot(x -> mypdf(Normal(n*q_U, p̂), x), 10, 70; label="", c=:blue, ls=:dash)
vline!([n*q_U]; label="n q_U", ls=:dashdot, c=:blue)
plot!(x -> mypdf(Normal(n*q_U, p̂), x), 10, 40;
    label="", fillrange=0, c=:blue, alpha=0.5)
annotate!(35, 0.01, text("2.5%", 10, :blue))
vline!([40]; label="", c=:gray)
title!("Normal(μ = $(n)×$(@sprintf "%.4f" q_L), σ = √($n⋅$p̂(1-$p̂))")
plot!(; size=(600, 250))
```

Out[63]:

In [64]:

```
# (4)
plot(x -> mypdf(Normal(n*q_L, p̂), x), 10, 70; label="", c=:red, ls=:dash)
vline!([n*q_L]; label="n q_L", ls=:dashdot, c=:red)
plot!(x -> mypdf(Normal(n*q_L, p̂), x), 40, 70;
    label="", fillrange=0, c=:red, alpha=0.5)
annotate!(45, 0.01, text("2.5%", 10, :red))
vline!([40]; label="", c=:gray)
title!("Normal(μ = $(n)×$(@sprintf "%.4f" q_L), σ = √($n⋅$p̂(1-$p̂))")
plot!(; size=(600, 250))
```

Out[64]:

In [65]:

```
# (3) and (4)
plot(x -> mypdf(Normal(n*q_U, σ̂), x), 10, 70; label="", c=:blue, ls=:dash)
vline!([n*q_U]; label="n q_U", ls=:dashdot, c=:blue)
plot!(x -> mypdf(Normal(n*q_U, σ̂), x), 10, 40;
    label="", fillrange=0, c=:blue, alpha=0.5)
annotate!(35, 0.01, text("2.5%", 10, :blue))

plot!(x -> mypdf(Normal(n*q_L, σ̂), x), 10, 70; label="", c=:red, ls=:dash)
vline!([n*q_L]; label="n q_L", ls=:dashdot, c=:red)
plot!(x -> mypdf(Normal(n*q_L, σ̂), x), 40, 70;
    label="", fillrange=0, c=:red, alpha=0.5)
annotate!(45, 0.01, text("2.5%", 10, :red))

vline!([40]; label="", c=:gray)
title!("[q_L, q_U] = [$(@sprintf "%.4f" q_L), $(@sprintf "%.4f" q_U)]")
plot!(; size=(600, 250))
```

Out[65]:

In [66]:

```
# (1)(2)(3)(4)
plot(x -> mypdf(Binomial(n, p_U), x), 10, 70; label="", c=1)
vline!([n*p_U]; label="n p_U", c=1)

plot!(x -> mypdf(Binomial(n, p_L), x), 10, 70; label="", c=2)
vline!([n*p_L]; label="n p_L", c=2)

plot!(x -> mypdf(Normal(n*q_U, σ̂), x), 10, 70; label="", c=:blue, ls=:dash)
vline!([n*q_U]; label="n q_U", ls=:dashdot, c=:blue)

plot!(x -> mypdf(Normal(n*q_L, σ̂), x), 10, 70; label="", c=:red, ls=:dash)
vline!([n*q_L]; label="n q_L", ls=:dashdot, c=:red)

vline!([40]; label="", c=:gray)
title!("[p_L, p_U] = [$(@sprintf "%.4f" p_L), $(@sprintf "%.4f" p_U)]"*
    ",  [q_L, q_U] = [$(@sprintf "%.4f" q_L), $(@sprintf "%.4f" q_U)]")
plot!(; size=(600, 250))
```

Out[66]:

次節で使う視覚化のための函数

In [67]:

```julia
function plot_bin_and_normal_cdfs(;
        n=20, p = 0.4, k = 5, k_normal = k, midp = false,
        xlim = (-0.5, 16), annpos = (2, 0.05), midpannpos = (2, 0.07),
        xtick = 0:20, kwargs...
    )
    P1 = plot(x -> mypdf(Binomial(n, p), x), xlim...; label="", c=1)
    if midp
        plot!(x -> mypdf(Binomial(n, p), x), first(xlim), k;
                label="", fillrange=0, c=1, alpha=0.5)
        prob1 = (cdf(Binomial(n, p), k-1) + cdf(Binomial(n, p), k))/2
        annotate!(midpannpos..., text("mid-P value", 9, :blue))
    else
        plot!(x -> mypdf(Binomial(n, p), x), first(xlim), k+0.5;
                label="", fillrange=0, c=1, alpha=0.5)
        prob1 = cdf(Binomial(n, p), k)
    end
    annotate!(annpos..., text(@sprintf("%0.5f", prob1), 10, :blue))
    vline!([k]; label="x = $k", c=:gray)
    title!("Binomial(n = $n, p = $p)")
    plot!(; xtick)

    P2 = plot(x -> mypdf(Normal(n*p, √(n*p*(1-p))), x), xlim...;
        label="", c=:blue, ls=:dash)
    plot!(x -> mypdf(Normal(n*p, √(n*p*(1-p))), x), first(xlim), k_normal;
        label="", fillrange=0, c=:blue, alpha=0.5)
    prob2 = cdf(Normal(n*p, √(n*p*(1-p))), k_normal)
    annotate!(annpos..., text(@sprintf("%0.5f", prob2), 10, :blue))
    vline!([k_normal]; label="x = $k_normal", c=:gray)
    title!("Normal(μ = $(n*p), σ = √$(n*p*(1-p)))")
    plot!(; xtick)

    plot(P1, P2; size=(800, 250))
end
```

Out[67]: plot_bin_and_normal_cdfs (generic function with 1 method)

In [68]:

```julia
pval_exact(k, n, p0) = cdf(Binomial(n, p0), k)
pval_normal(k, n, p0) = cdf(Normal(n*p0, √(n*p0*(1-p0))), k)
pval_conti(k, n, p0) = cdf(Normal(n*p0, √(n*p0*(1-p0))), k+0.5)
pval_midp(k, n, p0) = (cdf(Binomial(n, p0), k-1) + cdf(Binomial(n, p0), k))/2

function prob_alphaerror(pvalfunc, n, p0, α)
    null = Binomial(n, p0)
    sum(pdf(null, k) for k in support(null)
            if pvalfunc(k, n, p0) ≲ α; init = 0.0)
end

function _plot_alphaerrors(n, p0, α, tick)
    p_exact  = prob_alphaerror.(pval_exact, n, p0, α)
    p_normal = prob_alphaerror.(pval_normal, n, p0, α)
    p_conti  = prob_alphaerror.(pval_conti, n, p0, α)
    p_midp   = prob_alphaerror.(pval_midp, n, p0, α)
    plot(; legend=:topleft)
    plot!(α, p_exact; label="exact")
    plot!(α, p_normal; label="normal", ls=:dash)
    plot!(α, p_conti; label="conti. corr.", ls=:dot, lw=1.5)
    plot!(α, p_midp; label="mid-P", ls=:dashdot)
    plot!(α, α; label="", c=:black, ls=:dot)
    xtick = ytick = tick
    plot!(; xtick, ytick, xrotation=90)
    a, b = extrema(α)
    xlim = ylim = (a-0.02(b-a), b+0.02(b-a))
    plot!(; xlim, ylim)
    title!("n = $n, p₀ = $p0")
end

function plot_alphaerrors(n, p0)
    α =
    P1 = _plot_alphaerrors(n, p0, 0:0.001:1, 0:0.1:1)
    P2 = _plot_alphaerrors(n, p0, 0:0.0001:0.1, 0:0.01:1)
    plot(P1, P2; size=(640, 320), layout=(1, 2))
end
```

Out[68]: plot_alphaerrors (generic function with 1 method)

2.5 連続性補正とmid-P補正

まずは次の図を見て欲しい.

In [69]:
```
plot_bin_and_normal_cdfs()
```

Out[69]:

この図の左側と右側は, それぞれ, $n = 20$, $p = 0.4$, $k = 5$ に関する

$$\mathrm{cdf}(\mathrm{Binomial}(n, p), k) \approx 0.12560,$$
$$\mathrm{cdf}(\mathrm{Normal}(np, \sqrt{np(1-p)}), k) \approx 0.08545,$$

の計算を視覚化したものである. ここで cdf は累積分布函数 (cumulative distribution function) を意味する.

後者は前者の正規分布近似による計算であるが, 前者の値 ≈ 0.12560 より小さな値 ≈ 0.08545 になってしまっている. この近似をより正確にするためにはどうすればよいだろうか?

図を見ればわかるように, 正規分布側で 5 以下になる確率ではなく, 5.5 以下になる確率を計算すれば良さそうである. 実際に計算してみると次のようになる.

In [70]:
```
plot_bin_and_normal_cdfs(; k_normal = 5.5)
```

Out[70]:

正規分布近似の側で 5.5 以下になる確率を計算するようにしたら, 二項分布の側で 5 以下である確率の値 ≈ 0.12560 にかなり近い値 ≈ 0.12692 が得られた.

正規分布による近似の側(より一般的には離散分布の連続分布による近似側)をこのように補正することを **連続性補正** (continuity correction)と呼ぶ.

逆に二項分布の側での値を 5 以下になる確率と 4 以下になる確率の平均値で求めるようにすると, 正規分布近似側の値に近付く. 離散分布の側でこのように補正することを **mid-P 補正** (mid-P correction)と呼ぶ. 次の図を見よ.

In [71]:
```
plot_bin_and_normal_cdfs(; midp = true)
```

Out[71]:

mid-P補正は近似である連続分布の側に近似されている離散分布側の値を近付ける補正になっている. 近似される側を近似する側に近付ける補正を使うことは無意味だと感じる人もいるかもしれないが必ずしもそうではないことを説明しよう.

二項分布 $\mathrm{Binomial}(n, p_0)$ において, k 以下になる確率を次のように書き, データ k に関する仮説 $p \geq p_0$ の片側検定の正確な **P値** (P-value) と呼ぶことにする:

$$\mathrm{pval}_{\mathrm{exact}}(k|n, p_0) = \mathrm{cdf}(\mathrm{Binomial}(n, p_0), k).$$

これの正規分布近似版とその連続補正版とmid-P補正版をそれぞれ次のように書くことにする:

$$\begin{aligned}
\mathrm{pval}_{\mathrm{normal}}(k|n, p_0) &= \mathrm{cdf}(\mathrm{Normal}(np_0, \sqrt{np_0(1-p_0)}), k), \\
\mathrm{pval}_{\mathrm{conti}}(k|n, p_0) &= \mathrm{cdf}(\mathrm{Normal}(np_0, \sqrt{np_0(1-p_0)}), k+0.5), \\
\mathrm{pval}_{\mathrm{mid\text{-}P}}(k|n, p_0) &= \frac{1}{2}(\mathrm{cdf}(\mathrm{Binomial}(n, p_0), k-1) + \mathrm{cdf}(\mathrm{Binomial}(n, p_0), k)).
\end{aligned}$$

確率変数 K が $\mathrm{Binomial}(n, p_0)$ に従うとき, データを K としたときのこれらのP値の値が $0 < \alpha < 1$ について α 以下になる確率がどうなるかを計算してみよう. その確率は **第一種の過誤の確率** もしくは **アルファエラー** と呼ばれ, α に近いことが期待される値とみなされる(その理由は後で検定論を学べばわかる).

正確なP値については, 第一種の過誤の確率は確実に α 以下になるが, 大部分の場合に α より真に小さな値になり, 場合によってはかなり小さな値になる(この性質を **保守的** (conservative)と形容することがある). それの良い近似である連続補正された正規分布近似版も保守的である. それらに対して, 正規分布近似によるP値とmid-P補正によるP値の第一種の過誤の確率は α 前後の値になり, 平均的には正確なP値よりも α の値に近くなる.

第一種の過誤の確率を確実に α 以下にしたければ正確なP値が優れているが, 第一種の過誤の確率が α に近い方が良いことにすると正規分布近似によるP値とmid-P補正によるP値の方が優れていることになる.

このようにP値のユーザーが何を欲しているかによってどのP値の採用が適切であるかは変わる.

以下のグラフは横軸を α としたときのP値達の第一種の過誤の確率を同時プロットしたものである.

In [72]:
```
plot_alphaerrors(10, 0.6)
```

Out[72]:

In [73]:
```
plot_alphaerrors(20, 0.6)
```

Out[73]:

In [74]:
```
plot_alphaerrors(40, 0.6)
```

Out[74]:

2.6 問題: 二項分布の中心極限定理を使った極限の計算問題

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

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

解答例: 二項分布の中心極限定理より,

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

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

解答終

上の問題の結果を数値的に確認してみよう.

In [75]:

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

```
p = 0.4

m = 0
n = 10 ^ (m + 1) = 10
prod(1:2:2m - 1) = 1
sum((g(n, p, m, k) * pdf(Binomial(n, p), k) for k = 0:n)) = 1.000000000000001

m = 1
n = 10 ^ (m + 1) = 100
prod(1:2:2m - 1) = 1
sum((g(n, p, m, k) * pdf(Binomial(n, p), k) for k = 0:n)) = 1.0000000000000095

m = 2
n = 10 ^ (m + 1) = 1000
prod(1:2:2m - 1) = 3
sum((g(n, p, m, k) * pdf(Binomial(n, p), k) for k = 0:n)) = 2.998166666666703

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

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

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

2.7 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. $n \to \infty$ 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 $n \to \infty$ で 0 に収束する量を $o(a_n)$ と書く. 例えば $o\left(1/\sqrt{n}\right)$ は $\sqrt{n}$ をかけた後であっても $n \to \infty$ 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 $n \to \infty$ 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函数

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

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と計算の見通しが良くなる. (しかもこの形には重要な意味がある. その点については後述する.)

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

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

を仮定すると, 次が得られる:

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

これより,

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

ならば

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

となることがわかる.

この結果は, n が大きなとき, $K_n \sim \mathrm{Binomial}(n,p)$ に対する $X_n = (K_n - np)/\sqrt{n}$ が従う分布が $\mathrm{Normal}(0, p(1-p))$ で近似されることを意味している($p(1-p)$ は分布 $\mathrm{Bernoulli}(p)$ 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.

2.8 (1) Stirlingの公式を使った二項分布の確率質量函数の近似

n ごとに決まっている k は $n \to \infty$ で

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

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, $n \to \infty$ のとき $k \to \infty$ だけではなく, $n - k \to \infty$ 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

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

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函数

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

に適用すると,

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

2.9 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

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

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)

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

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

- 赤池弘次, エントロピーとモデルの尤度(<講座>物理学周辺の確率統計), 日本物理学会誌, 1980年第35巻7号, pp. 608-614. link (https://www.jstage.jst.go.jp/article/butsuri1946/35/7/35_7_608/_article/-char/ja/)
- 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)

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函数に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.

2.10 (2) 二項分布の確率質量函数から正規分布の密度函数が出て来ること

n ごとに決まっている k は $n \to \infty$ で

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

という条件を満たしているならば

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

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

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

を仮定する. このとき,

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

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

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

ゆえに,

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

したがって,

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

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

これより, $K_n \sim \mathrm{Binomial}(n,p)$, X_\infty \sim \op{Normal}(0, p(1-p)) のとき,

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

とおくと, 有界な連続函数 $f(x)$ について, $n\to\infty$ のとき

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

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

3 中心極限定理

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

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

3.1 中心極限定理のラフな説明

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

中心極限定理は以下のように同値な言い方が色々ある:

(1) それらの和 $X_1 + \cdots + X_n$ 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:

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

(2) それらの加法平均 $\bar{X}_n = (X_1 + \cdots + X_n)/n$ 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:

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

(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:

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

(4) 次の Z_n が従う分布は標準正規分布で近似される:

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

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

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

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

が成立することだと, 大雑把に定義しておく. (函数 $f(x)$ としてどのような函数を許すかについて詳細な説明をするためにはさらなる数学的道具立てが必要でかる, 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, $\sigma^2 = \mathrm{var}(X_k) = E[(X_k - \mu)^2] < \infty$ でかつ, $E[|X_k - \mu|^3] < \infty$ となっていると仮定する. このとき,

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

とおくと, $n \to \infty$ で Z_n の分布は標準正規分布に近付く.

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, $E[Y_k] = 0, \quad E[Y_k^2] = 1$ が成立している. ゆえに Y_k の特性函数 $\varphi(t)$ は k によらず,

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

という形になる. そして,

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

なので, Z_n の特性函数は, $n\to\infty$ で

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

と標準正規分布の特性函数 $e^{-t^2/2}$ 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 $n\to\infty$ で標準正規分布に近付く.

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 の特性函数 $\varphi_X(t) = E[e^{itX}]$ 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 の特性函数と確率変数 Y の特性函数が近ければ, $f(X)$ と $f(Y)$ の期待値も近くなる. このことから, 分布の近似は特性函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, $n\to\infty$ での $\varphi(t/\sqrt{n})^n \to e^{-t^2/2}$ 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 $n\to\infty$ での $\log\varphi\left(t/\sqrt{n}\right)^n \to -t^2/2$ の収束の速さは, $Y_k = (X_k-\mu)/\sigma$ の3次のモーメント $\bar{\mu}_3$ の絶対値の大きさで大体決まっていることがわかる. $\bar{\mu}_3$ の絶対値が小さいほど収束が速く, 大きいほど収束が遅い.

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

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

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

3.4 中心極限定理のキュムラント母函数を使った証明

中心極限定理: $X_1, X_2, X_3, \ldots$ 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, $\sigma^2 = \mathrm{var}(X_k) = E[(X_k - \mu)^2] < \infty$ でかつ, $E[|X_k - \mu|^3] < \infty$ となっており, さらに各 X_k のキュムラント母函数がうまく定義されているとする. このとき,

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

とおくと, $n \to \infty$ で Z_n の分布は標準正規分布に近付く.

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

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

という形になる. そして,

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

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

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

と標準正規分布のキュムラント母函数 $t^2/2$ に収束する.

ゆえに Z_n の分布は $n \to \infty$ で標準正規分布に近付く.

証明終

注意: このようにキュムラント母函数を使うと証明が非常にシンプルになる.

3.5 中心極限定理の収束の速さと歪度と尖度

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

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

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

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

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

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

このとき,

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

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

3.6 中心極限定理のTaylorの定理のみを使う証明

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

その方法は本質的にFourier解析に依存している.

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

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

3.6.1 多重積分の書き方

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

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

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

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

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

3.6.2 積分剰余項型のTaylorの定理

以下では次の結果を示そう:

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

とおくと,

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

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

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

を4回使うと,

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

不定積分を繰り返す過程で,

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

となることから, 次の形の式が得られた:

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

以上とまったく同様にして次を示すことができる:

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

とおくと,

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

この結果を **積分剰余項型のTaylorの定理** と呼ぶことにする.

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

3.6.3 積分剰余項の別の表示

剰余項 $R_n(x,t)$ は次の式で定義されると考えてよい:

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

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

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

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

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

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

Taylorの定理の証明は易しい!

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

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

すなわち,

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

後でTaylorの定理をこの形で使うことになる.

注意: 上で求めた積分剰余項の公式と前節で求めた公式の関係. 前節では

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

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

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

なので, x_n による積分を一番外側に出すと次のように書き直される:

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

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

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

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

3.6.4 微分剰余項型のTaylorの定理

$f(t)$ は C^n 級であると仮定する.

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

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

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

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

これで積分を使わずに微分だけを使った剰余項の表示が得られた.

3.6.5 Taylorの定理を使った中心極限定理の証明

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

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

このとき, Z を標準正規分布に従う確率変数であるとし,

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

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

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

特に次が成立している:

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

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

証明:

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

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

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

このとき,

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

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

まず, $G(a, h)$ を次のように定める:

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

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

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

$f(z)$ にTaylorの定理を適用すると,

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

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

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

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

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

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

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

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

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

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

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

これが示したい不等式であった.

証明終

注意: Z を標準正規分布に従う確率変数とするとき,

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

In [76]:
```
[(2quadgk(z -> z^(2k)*pdf(Normal(), z), 0, Inf)[1],
        2^(k+1/2)*gamma(k+1/2)/√(2π), prod(1:2:2k-1)) for k in 0:4]
```

Out[76]:
```
5-element Vector{Tuple{Float64, Float64, Int64}}:
 (0.9999999999999998, 1.0000000000000002, 1)
 (1.0000000000000009, 1.0000000000000002, 1)
 (3.000000000000005, 3.0000000000000004, 3)
 (15.00000000000029, 15.000000000000004, 15)
 (105.00000000000988, 105.00000000000001, 105)
```

In [77]:
```
[(2√(2π)*quadgk(z -> z^(2k+1)*pdf(Normal(), z), 0, Inf)[1],
        2^(k+1)*gamma(k+1), 2^(k+1)*factorial(k)) for k in 0:4]
```

Out[77]:
```
5-element Vector{Tuple{Float64, Float64, Int64}}:
 (1.9999999999999993, 2.0, 2)
 (4.0, 4.0, 4)
 (16.0000000000001, 16.0, 16)
 (96.00000000000463, 96.0, 96)
 (767.9999999999999, 768.0, 768)
```

3.6.6 中心極限定理の収束の速さと歪度と尖度(再)

前節で示した不等式

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

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

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

3.7 問題: 中心極限定理の収束の様子のグラフ

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 [78]:

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

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

Out[78]:

In [79]:

```
# 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
n = 10
L = 10^6
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[79]:

```
10×1000000 Matrix{Float64}:
 0.672547   0.0848108  0.524284  0.33867     …  0.0986632  0.705264  0.683651  0.764941
 0.337707   0.306371   0.582008  0.324429        0.681252   0.153969  0.680763  0.342673
 0.395861   0.178901   0.979898  0.366622        0.437717   0.640549  0.569048  0.7454
 0.0274189  0.252751   0.49977   0.677686        0.4238     0.836334  0.850832  0.283192
 0.390767   0.0892799  0.33393   0.436294        0.663443   0.274744  0.455814  0.672315
 0.190572   0.150144   0.988904  0.482095     …  0.870442   0.228277  0.556281  0.733509
 0.515534   0.710163   0.722886  0.725216        0.540596   0.111925  0.478272  0.13938
 0.977059   0.106152   0.312564  0.750852        0.232703   0.39544   0.677336  0.210737
 0.989893   0.227767   0.748116  0.177032        0.174567   0.632077  0.585337  0.0599564
 0.306079   0.944226   0.966805  0.00705588      0.940918   0.268038  0.301074  0.0494107
```

In [80]:

```
# Z_n = √n(X̄_n - μ)/σ を大量に計算
μ = mean(dist)
σ = std(dist)
Zns = [√n*(mean(X) - μ)/σ for X in eachcol(Xs)] # Zns はZ_n達という意味
first(Zns, 5)
```

Out[80]:

```
5-element Vector{Float64}:
 -0.2153235809234101
 -2.1354980522614544
  1.817524904240514
 -0.7822010477082603
 -0.6391880610739005
```

In [81]:

```
# 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
histogram(Zns; norm=true, alpha=0.3, bin=100, label="Zₙ")
plot!(Normal(0, 1); label="Normal(0,1)", lw=2)
title!("n = $n")
```

Out[81]:

In [82]:

```
# 以上の手続きを函数化
# さらに
#   * メモリアロケーションの節約と並列化による最適化
#   * 離散分布の場合にも対応
#   * 歪度と尖度を混合モデルの場合にも表示

function plot_central_limit_theorem(dist, n;
        L=10^6,
        μ = mean(dist),
        σ = std(dist),
        a = max(minimum(dist), μ - 5σ),
        b = min(maximum(dist), μ + 5σ),
        disttitle = distname(dist),
        bin = 100,
        kwargs...
    )
    println("skewness(dist) = ", myskewness(dist))
    println("kurtosis(dist) = ", mykurtosis(dist))

    # 分布distをプロット
    if dist isa DiscreteUnivariateDistribution
        P1 = bar(round(Int, a):round(Int, b), x -> pdf(dist, x), ; alpha=0.3, label="")
    else
        P1 = plot(x -> pdf(dist, x), a, b; label="")
    end
    title!(disttitle)

    # 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,
    # Z_n = √n(X̄_n - μ)/σ を大量に計算することを並列化で実行
    # このような計算には並列化が非常に有効である
    Zns = Vector{Float64}(undef, L)
    tmp = [Vector{Float64}(undef, n) for _ in 1:Threads.nthreads()]
    Threads.@threads for i in 1:L # Threads.@threads マクロで並列化
        X = rand!(dist, tmp[Threads.threadid()]) # rand!を使ってアロケーションを節約
        Zns[i] = √n*(mean(X) - μ)/σ
    end

    # Z達のヒストグラムと標準正規分布の密度函数を比較
    if dist isa DiscreteUnivariateDistribution
        summin = max(n*minimum(dist), round(n*μ - 6√n*σ))
        summax = min(n*maximum(dist), round(n*μ + 6√n*σ))
        sumran = summin-0.5:summax+0.5
        bin = @. √n*(sumran/n - μ)/σ
        P2 = histogram(Zns; norm=true, alpha=0.3, bin, label="Zₙ")
    else
        P2 = histogram(Zns; norm=true, alpha=0.3, bin, label="Zₙ")
    end
    plot!(Normal(0,1); label="Normal(0,1)", lw=2)
    title!("n = $n")

    plot(P1, P2; size=(800, 250), kwargs...)
end
```

Out[82]: plot_central_limit_theorem (generic function with 1 method)

In [83]:
```
# 中心極限定理による収束が速い場合1の再現
plot_central_limit_theorem(Uniform(0, 1), 10; a=-0.5, b=1.5)
```

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

Out[83]:

In [84]:
```
# 中心極限定理による収束が速い場合2
plot_central_limit_theorem(Beta(0.5, 0.5), 10; a=0.01, b=0.99)
```

```
skewness(dist) = 0.0
kurtosis(dist) = -1.5
```

Out[84]:

In [85]:
```
# 中心極限定理による収束が速い場合3
plot_central_limit_theorem(Bernoulli(0.5), 10)
```

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

Out[85]:

In [86]:
```
# 中心極限定理による収束が速い場合4
plot_central_limit_theorem(Laplace(), 10)
```

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

Out[86]:

In [87]:
```
# 左右非対称な分布の場合1
plot_central_limit_theorem(Gamma(3, 1), 10)
```

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

Out[87]:

In [88]:
```
# 左右非対称な分布の場合2
plot_central_limit_theorem(Bernoulli(0.25), 10)
```

```
skewness(dist) = 1.1547005383792517
kurtosis(dist) = -0.666666666666667
```

Out[88]:

In [89]:
```
# 左右非対称な分布の場合3
plot_central_limit_theorem(Poisson(1), 10)
```

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

Out[89]:

In [90]:
```
# 中心極限定理による収束が遅い場合1
# 左右の非対称性が大きな分布を試してみる
# 指数分布は左右の非対称性が大きな分布になっている
plot_central_limit_theorem(Exponential(), 10)
```

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

Out[90]:

In [91]:
```
plot_central_limit_theorem(Exponential(), 40)
```

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

Out[91]:

In [92]:
```
plot_central_limit_theorem(Exponential(), 160)
```

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

Out[92]:

In [93]:
```
# 中心極限定理による収束が遅い場合2
# 自由度1のχ²分は左右の非対称性が指数分布よりも大きな分布になっている
plot_central_limit_theorem(Chisq(1), 10; a=0.05, b=6)
```

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

Out[93]:

In [94]:
```
plot_central_limit_theorem(Chisq(1), 40; a=0.05, b=6)
```

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

Out[94]:

In [95]:
```
plot_central_limit_theorem(Chisq(1), 160; a=0.05, b=6)
```

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

Out[95]:

以下で使う, 分布 `MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])` の確率密度函数は次の形になる:

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

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函数が複数の正規分布の確率密度函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 [96]:
```
# 中心極限定理による収束が遅い場合3
# 以下のような分布distも左右の非対称性が大きな分布に分類される
dist = MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])
disttitle = "MixtureModel([Normal(0, 1), Normal(20, 1)], [0.95, 0.05])"
titlefontsize = 9
a, b = -4, 24
plot_central_limit_theorem(dist, 10; a, b, disttitle, titlefontsize)
```

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

Out[96]:

In [97]:
```
plot_central_limit_theorem(dist, 40; a, b, disttitle, titlefontsize)
```

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

Out[97]:

In [98]:
```
plot_central_limit_theorem(dist, 160; a, b, disttitle, titlefontsize)
```

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

Out[98]:

In [99]:
```
# 中心極限定理による収束が遅い場合4
# 対数正規分布は左右の非対称性が非常に大きな分布である
# 右側の裾が太く，外れ値が出やすい
plot_central_limit_theorem(LogNormal(), 10; bin=range(-3, 9, 100))
```

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

Out[99]:

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

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

Out[100]:

In [101]:
```
plot_central_limit_theorem(LogNormal(), 160; bin=range(-3, 9, 100))
```

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

Out[101]:

4 デルタ法

4.1 1次のデルタ法と2次のデルタ法

X は平均 μ, 分散 σ^2 を持つ確率変数であるとし, 滑らかな函数 $f(x)$ にそれを代入して得られる確率変数 $Y = f(X)$ について考える.

x が μ に十分近ければ, $y = f(x)$ はその $x = \mu$ での接線

$$y = f(\mu) + f'(\mu)(x - \mu)$$

で近似される. σ が十分小さければ X の分布の大部分は μ の近くに集まり, 確率変数 $Y = f(X)$ は確率変数

$$Y_1 = f(\mu) + f'(\mu)(X - \mu)$$

で近似される. ゆえに

$$\begin{aligned} &E[Y] \approx E[Y_1] = f(\mu) + f'(\mu)E[X - \mu] = f(\mu), \\ &\mathrm{var}(Y) \approx \mathrm{var}(Y_1) = f'(\mu)^2\sigma^2 \end{aligned}$$

という近似が使える. これを **1次のデルタ法** と呼ぶことにする.

接線よりも精密に2次までの項を加えた

$$y = f(\mu) + f'(\mu)(x - \mu) + \frac{f''(\mu)}{2}(x - \mu)^2$$

は $y = f(x)$ のさらに良い近似になっている. ゆえに, 確率変数 $Y = f(X)$ は, 上の Y_1 よりも,

$$Y_2 = f(\mu) + f'(\mu)(X - \mu) + \frac{f''(\mu)}{2}(X - \mu)^2$$

でより良く近似される. この近似を使えば, $E[Y] \approx E[Y_1] = f(\mu)$ よりも精度が高い

$$E[Y] \approx E[Y_2] = f(\mu) + \frac{f''(\mu)}{2}E[(x - \mu)^2] = f(\mu) + \frac{f''(\mu)\sigma^2}{2}$$

という近似が得られる. これを **2次のデルタ法** と呼ぶことにする.

4.2 おまけ: 2次のデルタ法使用時の近似分散と歪度と尖度に関する不等式

確率変数 X の期待値を μ, 分散を σ^2, 歪度を $\bar{\kappa}_3 = E[((X - \mu)/\sigma)^3]$, 尖度を $\bar{\kappa}_4 = E[((X - \mu)/\sigma)^4] - 3$ と表し, 任意の実数 a, b, c に対して,

$$Y_2 = a + b(X - \mu) + c(X - \mu)^2$$

とおく. このとき, $E[Y_2] = a + c\sigma^2$ となることは易しい. Y_2 の分散は以下のようにして求められる:

$$\begin{aligned} \mathrm{var}(Y_2) &= E[(Y_2 - (a + c\sigma^2))^2] = E[(-c\sigma^2 + b(X - \mu) + c(X - \mu)^2)^2] \\ &= E[c^2\sigma^4 - 2bc\sigma^2(X - \mu) + (b^2 - 2c^2\sigma^2)(X - \mu)^2 + 2bc(X - \mu)^3 + c^2(X - \mu)^4] \\ &= c^2\sigma^4 + (b^2 - 2c^2\sigma^2)\sigma^2 + 2bc\sigma^3\bar{\kappa}_3 + c^2\sigma^4(\bar{\kappa}_4 + 3) \\ &= b^2\sigma^2 + 2bc\sigma^3\bar{\kappa}_3 + c^2\sigma^4(\bar{\kappa}_4 + 2) \\ &= \begin{bmatrix} b\sigma & c\sigma^2 \end{bmatrix} \begin{bmatrix} 1 & \bar{\kappa}_3 \\ \bar{\kappa}_3 & \bar{\kappa}_4 + 2 \end{bmatrix} \begin{bmatrix} b\sigma \\ c\sigma^2 \end{bmatrix}. \end{aligned}$$

これが任意の実数 b, c について 0 以上になることから, この結果に現れた 2×2 行列の行列式は 0 以上になることがわかるので, 次の不等式が得られる:

$$\bar{\kappa}_4 + 2 \geq \bar{\kappa}_3^2.$$

「標本分布について」のノートで標本平均と不偏分散の分散共分散行列の非負性を用いて示した「歪度と尖度に関する不等式」が再び証明された.

注意: 以上の証明は本質的に $X - \mu$ と $(X - \mu)^2$ の分散共分散行列が半正定値になることを使う証明と本質的に同じである.

4.3 問題: 2次のデルタ法の不偏分散の平方根の期待値の近似への応用

$X_1, X_2, \ldots, X_n$ を独立同分布な確率変数達であるとし, それらの標本平均と不偏分散を

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

と書き, X_i 達共通の期待値を μ, 共通の分散を σ^2 と書き, 共通の尖度を $\bar{\kappa}_4 = E[((X_i - \mu)/\sigma)^4] - 3$ と書く. このとき, S^2 の期待値と分散は

$$E[S^2] = \sigma^2, \quad \mathrm{var}(S^2) = \sigma^4\left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}\right)$$

となるのであった. これに2次のデルタ法を適用すると, $S = \sqrt{S^2}$ の期待値の σ からのずれが近似的に

$$E[S] \approx \sigma - \frac{\mathrm{var}(S^2)}{8\sigma^3} = \sigma\left(1 - \frac{1}{8}\left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}\right)\right)$$

のように表されることを確認せよ.

解答例: $f(x) = \sqrt{x}$ とおくと,

$$f'(x) = \frac{1}{2}x^{-1/2}, \quad f''(x) = -\frac{1}{4}x^{-3/2}, \quad f(\sigma^2) = \sigma, \quad f''(\sigma^2) = -\frac{1}{4\sigma^3}$$

なので, $X = S^2$ に2次のデルタ法を適用すると,

$$\begin{aligned} E[S] &\approx f(\sigma^2) + \frac{f''(\sigma^2)\,\mathrm{var}(S^2)}{2} \\ &= \sigma - \frac{\mathrm{var}(S^2)}{8\sigma^3} = \sigma\left(1 - \frac{1}{8}\left(\frac{\bar{\kappa}_4}{n} + \frac{2}{n-1}\right)\right). \end{aligned}$$

解答終

注意: 不偏分散 S^2 は分散 σ^2 の不偏推定量だが, その平方根 $S = \sqrt{S^2}$ は標準偏差 σ の不偏推定量にならないことが上の問題からわかる. $\mathrm{var}(S^2) \geq 0$ なので $E[S]$ の値は σ よりも少し小さくなる.

注意: これを理由に $S = \sqrt{S^2}$ は標準偏差の推定量として不適切であると考えては **いけない**. 推定量が不偏であること自体に特別に優先度が高い価値があるわけではない. 推定法に不偏性の条件を課すと誤差が増えてしまう場合がある.

In [102]:
```
function check_deltamethod_for_S(dist, n; L=10^7)
    σ = std(dist)
    ku = mykurtosis(dist)
    @show dist
    println("kurtosis = ", ku)
    @show n

    S = Vector{Float64}(undef, L)
    tmp = [Vector{Float64}(undef, n) for _ in 1:Threads.nthreads()]
    Threads.@threads for i in 1:L
        X = rand!(dist, tmp[Threads.threadid()])
        S[i] = std(X)
    end
    println("approximation = ", σ*(1 - (1/8)*(ku/n + 2/(n-1))))
    println("mean(S) =       ", mean(S))
    println("σ =             ", σ)
end
```

Out[102]:
```
check_deltamethod_for_S (generic function with 1 method)
```

In [103]:
```
check_deltamethod_for_S(Normal(2, 3), 10)
```

```
dist = Normal{Float64}(μ=2.0, σ=3.0)
kurtosis = 0.0
n = 10
approximation = 2.9166666666666665
mean(S) =       2.9181035910662665
σ =             3.0
```

In [104]:
```
check_deltamethod_for_S(Uniform(), 10)
```

```
dist = Uniform{Float64}(a=0.0, b=1.0)
kurtosis = -1.2
n = 10
approximation = 0.28498650787499025
mean(S) =       0.28475298814557065
σ =             0.28867513459481287
```

In [105]:
```
check_deltamethod_for_S(Gamma(4, 3), 10)
```

```
dist = Gamma{Float64}(α=4.0, θ=3.0)
kurtosis = 1.5
n = 10
approximation = 5.720833333333333
mean(S) =       5.757022288766159
σ =             6.0
```

In [106]:
```
check_deltamethod_for_S(Bernoulli(0.4), 10)
```

```
dist = Bernoulli{Float64}(p=0.4)
kurtosis = -1.833333333333333
n = 10
approximation = 0.4875165001955964
mean(S) =       0.4859214468305372
σ =             0.4898979485566356
```

In [107]:
```
check_deltamethod_for_S(LogNormal(), 10; L=10^6)
```

```
dist = LogNormal{Float64}(μ=0.0, σ=1.0)
kurtosis = 110.9363921763115
n = 10
approximation = -0.895778896909659
mean(S) =       1.7171498580585198
σ =             2.1611974158950877
```

In [108]:
```
check_deltamethod_for_S(LogNormal(), 100; L=10^6)
```

```
dist = LogNormal{Float64}(μ=0.0, σ=1.0)
kurtosis = 110.9363921763115
n = 100
approximation = 1.856045541537819
mean(S) =       2.045538958724601
σ =             2.1611974158950877
```

In [109]:
```
check_deltamethod_for_S(LogNormal(), 1000; L=10^6)
```

```
dist = LogNormal{Float64}(μ=0.0, σ=1.0)
kurtosis = 110.9363921763115
n = 1000
approximation = 2.1306871451883986
mean(S) =       2.140192566082289
σ =             2.1611974158950877
```

4.4 問題: 1次のデルタ法の正規分布近似への応用

確率変数 X が近似的に正規分布 $\mathrm{Normal}(\mu, \sigma)$ に従っているとき, 滑らかな函数 $f(x)$ に対して, σ が十分に小さければ, 確率変数 $Y = f(X)$ は近似的に正規分布 $\mathrm{Normal}(f(\mu), |f'(\mu)|\sigma)$ に従うことを示せ.

解答例: x が μ に十分近ければ $y = f(x)$ は $x = \mu$ における接線 $y = f(\mu) + f'(\mu)(x - \mu)$ で近似される. σ が十分に小さければ確率変数 X の分布の大部分は μ の十分近くに留まる. 以上の状況のもとで, 確率変数 $Y = f(X)$ は確率変数 $Y_1 = f(\mu) + f'(\mu)(X - \mu)$ で近似される. X が正規分布 $\mathrm{Normal}(\mu, \sigma)$ に近似的に従っていると仮定しているので, そのアフィン変換である Y_1 も近似的に正規分布に従い, その期待値と分散は

$$E[Y_1] = f(\mu) + f'(\mu)E[X - \mu] \approx f(\mu), \qquad \mathrm{var}(Y_1) = f'(\mu)^2\,\mathrm{var}(X) \approx f'(\mu)^2\sigma^2$$

となる. 以上をまとめると, Y が近似的に正規分布 $\mathrm{Normal}(f(\mu), |f'(\mu)|\sigma)$ に従うことがわかる.

解答終

注意: デルタ法を使うと, 正規分布で近似される確率変数から正規分布に近似的に従う別の確率変数を作れる. 中心極限定理による正規分布近似はデルタ法を使えば大幅に拡張される. 以上では一変数函数 $f(x)$ に正規分布に近似的に従う確率変数 X を代入する場合のみを扱ったが, 多変数函数に複数の正規分布に近似的に従う確率変数達を代入する場合も同じように一次近似で考えればよい.

4.5 問題: 1次のデルタ法による分布の逆数の近似

正値の確率変数 X は正規分布 $\mathrm{Normal}(\mu, \sigma)$ に近似的に従っていると仮定し, σ/μ は十分に小さいと仮定する. このとき, $Y = 1/X$ が正規分布 $\mathrm{Normal}(1/\mu, \sigma/\mu^2)$ に近似的に従うことを示せ.

解答例: $f(x) = 1/x$ とおくと,

$$f(x) = \frac{1/\mu}{1 + (x/\mu - 1)} = \frac{1}{\mu}\left(1 - \left(\frac{x}{\mu} - 1\right) + \left(\frac{x}{\mu} - 1\right)^2 - \cdots\right)$$

なので, $y = f(x)$ は x/μ が 1 に近ければ, その $x = \mu$ での接線

$$y = \frac{1}{\mu}\left(1 - \left(\frac{x}{\mu} - 1\right)\right) = \frac{1}{\mu} - \frac{x-\mu}{\mu^2}$$

で近似される. σ/μ が十分に小さければ X/μ の分布の大部分は 1 の近くに留まる. ゆえに $Y = 1/X$ の分布は $Y_1 = 1/\mu - (X-\mu)/\mu^2$ の分布で近似される. このことから, $Y = 1/X$ が平均 $1/\mu$, 分散 σ^2/μ^4 の正規分布に近似的に従うことがわかる.

解答終

4.6 例: 自由度が大きなχ^2分布の逆数の近似

例えば, $X \sim \mathrm{Chisq}(\nu)$ のとき, $E[X] = \nu$, $\mathrm{var}(X) = 2\nu$ であり, ν が大きなとき X は近似的に平均 $\mu = \nu$, 分散 $\sigma^2 = 2\nu$ の正規分布に近似的に従い, ν が大きなとき $\sigma/\mu = \sqrt{2/\nu}$ の値は小さくなる. ゆえに上の問題の結果より, $1/X$ は平均 $1/\mu = 1/\nu$, 分散 $\sigma^2/\mu^4 = 2/\nu^3$ の正規分布に近似的に従う.

注意: $X \sim \mathrm{Chisq}(\nu) = \mathrm{Gamma}(\nu/2, 2)$ のとき, $1/X \sim \mathrm{InverseGamma}(\nu/2, 1/2)$ なので,

$$E[1/X] = \frac{1/2}{\nu/2 - 1} = \frac{1}{\nu - 2} \approx \frac{1}{\nu},$$
$$\mathrm{var}(1/X) = \frac{(1/2)^2}{(\nu/2-1)^2(\nu/2-2)} = \frac{2}{(\nu-2)^2(\nu-4)} \approx \frac{2}{\nu^3}.$$

In [110]:
```
InverseChisq(ν) = InverseGamma(ν/2, 1/2)

function plot_invchisq(ν)
    xlim = (max(0, 1/ν - 4√(2/ν^3)), 1/ν + 6√(2/ν^3))
    plot(InverseChisq(ν), xlim...; label="InverseChisq(ν=$ν)")
    plot!(Normal(1/ν, √(2/ν^3)); label="Normal(1/ν, √(2/ν³))", ls=:dash)
    plot!(; legend=:outertop)
end
```

Out[110]:
```
plot_invchisq (generic function with 1 method)
```

In [111]:
```
plot(plot_invchisq.((10, 40))...; size=(800, 250), layout=(1, 2))
```

Out[111]:

In [112]:
```
plot(plot_invchisq.((160, 640))...; size=(800, 250), layout=(1, 2))
```

Out[112]:

4.7 問題: デルタ法による分布の対数の近似

正値の確率変数 X は正規分布 $\mathrm{Normal}(\mu, \sigma)$ に近似的に従っていると仮定し, σ/μ は十分に小さいと仮定する. このとき, $Y = \log X$ が正規分布 $\mathrm{Normal}(\log \mu, \sigma/\mu)$ に近似的に従うことを示せ.

解答例: $f(x) = \log x$ とおくと,

$$f(x) = \log \nu \left(1 + \left(\frac{x}{\mu} - 1\right)\right) = \log \mu + \left(\frac{x}{\mu} - 1\right) - \frac{1}{2}\left(\frac{x}{\mu} - 1\right)^2 + \cdots$$

なので, $y = f(x)$ は x/μ が 1 に近ければ, その $x = \mu$ での接線

$$y = \log \mu + \left(\frac{x}{\mu} - 1\right) = \log \mu + \frac{x - \mu}{\mu}$$

で近似される. σ/μ が十分に小さければ X/μ の分布の大部分は 1 の近くに留まる. ゆえに $Y = \log X$ の分布は $Y_1 = \log \mu + (X - \mu)/\mu$ の分布で近似される. このことから, $Y = \log X$ が平均 $\log \mu$, 分散 σ^2/μ^2 の正規分布に近似的に従うことがわかる.

解答終

注意: 正値の確率変数の分布は対数を取ることによって左右の非対称性が緩和されて, 正規分布近似がより正確になる場合がある (次の節の例を見よ). 分布の非対称性を補正して正規分布近似をより正確にする工夫は色々研究されている.

4.8 例: 自由度が大きなχ²分布の対数の近似

例えば, $X \sim \mathrm{Chisq}(\nu)$ のとき, $E[X] = \nu$, $\mathrm{var}(X) = 2\nu$ であり, ν が大きなとき X は近似的に平均 $\mu = \nu$, 分散 $\sigma^2 = 2\nu$ の正規分布に近似的に従い, ν が大きなとき $\sigma/\mu = \sqrt{2/\nu}$ の値は小さくなる. ゆえに上の問題の結果より, $\log X$ は平均 $\log \mu = \log \nu$, 分散 $\sigma^2/\mu^2 = 2/\nu$ の正規分布に近似的に従う.

この場合には余り大きくない自由度 ν におけるχ²分布の左右の非対称性が対数を取ることによって緩和され, 正規分布による近似の精度が高まる.

In [113]:
```
pdf_logchisq(ν, y) = pdf(Chisq(ν), exp(y))*exp(y)

function plot_logchisq(ν)
    xlim = (max(0, ν - 4√(2ν)), ν + 4√(2ν))
    P0 = plot(Chisq(ν), xlim...; label="Chisq(ν=$ν)")
    plot!(Normal(ν, √(2ν)); label="Normal(ν, √(2ν))", ls=:dash)
    plot!(; legend=:outertop)

    xlim = (log(ν) - 5√(2/ν), log(ν) + 4√(2/ν))
    P1 = plot(y -> pdf(Chisq(ν), exp(y))*exp(y), xlim...; label="LogChisq(ν=$ν)")
    plot!(Normal(log(ν), √(2/ν)); label="Normal(log ν, √(2/ν))", ls=:dash)
    plot!(; legend=:outertop)

    plot(P0, P1; size=(800, 250))
end
```

Out[113]: plot_logchisq (generic function with 1 method)

In [114]:
```
plot_logchisq(10)
```

Out[114]:

In [115]:
```
plot_logchisq(20)
```

Out[115]:

In [116]:
```
plot_logchisq(40)
```

Out[116]:

In [117]:
```
plot_logchisq(80)
```

Out[117]:

4.9 問題: ベータ分布の正規分布近似

確率変数達 X, Y は独立であり, $X \sim \mathrm{Gamma}(\alpha, 1)$, $Y \sim \mathrm{Gamma}(\beta, 1)$ を満たしていると仮定する. このとき $T = X/(X+Y)$ とおくと, $T \sim \mathrm{Beta}(\alpha, \beta)$ となるのであった. α, β が大きなとき, X, Y の分布はそれぞれ $\mathrm{Normal}(\alpha, \sqrt{\alpha})$, $\mathrm{Normal}(\beta, \sqrt{\beta})$ に近似的に従う. 2変数函数 $f(x, y) = x/(x+y)$ に1次のデルタ法を適用して, α, β が大きなとき, $T = X/(X+Y) \sim \mathrm{Beta}(\alpha, \beta)$ が

$$\mathrm{Normal}\left(\frac{\alpha}{\alpha+\beta}, \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^3}}\right)$$

に近似的に従うことを示せ.

注意: $T \sim \mathrm{Beta}(\alpha, \beta)$ の正確な期待値と分散は次のようになる:

$$E[T] = \frac{\alpha}{\alpha+\beta}, \quad \mathrm{var}(T) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

解答例: $f(x,y) = x/(x+y) = 1 - y/(x+y)$ を偏微分すると,

$$\begin{aligned}
f_x(x,y) &= \frac{y}{(x+y)^2}, \quad f_y(x,y) = \frac{-x}{(x+y)^2}, \\
f_{xx}(x,y) &= \frac{-2y}{(x+y)^3}, \quad f_{xy}(x,y) = \frac{x-y}{(x+y)^3}, \quad f_{yy}(x,y) = \frac{2x}{(x+y)^3}
\end{aligned}$$

なので,

$$\begin{aligned}
T &= \frac{X}{X+Y} = f(X,Y) \\
&= f(\alpha,\beta) + f_x(\alpha,\beta)(X-\alpha) + f_y(\alpha,\beta)(Y-\beta) \\
&+ \frac{1}{2}\left(f_{xx}(\alpha,\beta)(X-\alpha)^2 + 2f_{xy}(\alpha,\beta)(X-\alpha)(Y-\beta) + f_{yy}(\alpha,\beta)(Y-\beta)^2\right) + \cdots \\
&= \frac{\alpha}{\alpha+\beta} + \frac{\beta}{(\alpha+\beta)^2}(X-\alpha) + \frac{-\alpha}{(\alpha+\beta)^2}(Y-\beta) \\
&+ \frac{-\beta}{(\alpha+\beta)^3}(X-\alpha)^2 + \frac{\alpha-\beta}{(\alpha+\beta)^3}(X-\alpha)(Y-\beta) + \frac{\alpha}{(\alpha+\beta)^3}(Y-\beta)^2 + \cdots \\
&= \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}\frac{X-\alpha}{\alpha+\beta} + \frac{-\alpha}{\alpha+\beta}\frac{Y-\beta}{\alpha+\beta} \\
&+ \frac{-\beta}{\alpha+\beta}\left(\frac{X-\alpha}{\alpha+\beta}\right)^2 + \frac{\alpha-\beta}{\alpha+\beta}\frac{X-\alpha}{\alpha+\beta}\frac{Y-\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}\left(\frac{Y-\beta}{\alpha+\beta}\right)^2 + \cdots.
\end{aligned}$$

ゆえに, α, β の両方が大きいならば(より正確に言えば $\alpha = N\mu$, $\beta = N(1-\mu, 0 < \mu < 1$ で N が大きいならば),

$$\frac{\mathrm{std}(X)}{\alpha+\beta} = \frac{\sqrt{\alpha}}{\alpha+\beta}, \quad \frac{\mathrm{std}(Y)}{\alpha+\beta} = \frac{\sqrt{\beta}}{\alpha+\beta}$$

がどちらも小さくなり, 1次近似

$$T \approx \frac{\alpha}{\alpha+\beta} + \frac{\beta}{(\alpha+\beta)^2}(X-\alpha) + \frac{-\alpha}{(\alpha+\beta)^2}(Y-\beta)$$

が有効になり,

$$\begin{aligned}
E[T] &\approx \frac{\alpha}{\alpha+\beta}, \\
\mathrm{var}(T) &\approx \frac{\beta^4\,\mathrm{var}(X)}{(\alpha+\beta)^4} + \frac{\alpha^2\,\mathrm{var}(Y)}{(\alpha+\beta)^4} = \frac{\beta^2\alpha + \alpha^2\beta}{(\alpha+\beta)^4} = \frac{\alpha\beta}{(\alpha+\beta)^3}.
\end{aligned}$$

さらにそのとき, X, Y の分布は正規分布で近似されるので, T の分布も正規分布で近似される.

解答終

In [118]:

```
@vars x y
expr = x/(x+y)
[
    expr
    diff(expr, x).simplify()
    diff(expr, y).simplify()
    ;;
    diff(expr, x, x).simplify()
    diff(expr, x, y).simplify()
    diff(expr, y, y).simplify()
]
```

Out[118]:

$$\begin{bmatrix} \frac{x}{x+y} & -\frac{2y}{(x+y)^3} \\ \frac{y}{(x+y)^2} & \frac{x-y}{(x+y)^3} \\ -\frac{x}{(x+y)^2} & \frac{2x}{(x+y)^3} \end{bmatrix}$$

注意: $X_1, X_2, \ldots, X_n$ が独立同分布な確率変数達で $X_i \sim \mathrm{Gamma}(\alpha/n, \theta)$ のとき, それらの和もガンマ分布に $X_1 + X_2 + \cdots + X_n \sim \mathrm{Gamma}(\alpha, \theta)$ と従う. この事実と中心極限定理から, α が大きなときのガンマ分布 $\mathrm{Gamma}(\alpha, \theta)$ が正規

分布 Normal(α, $\sqrt{\alpha\theta^2}$) で近似されることが導かれる.

ゆえに, 上の問題の結果は, パラメータの値が大きなときのベータ分布が正規分布で近似されることが, ガンマ分布に関する中心極限定理の帰結になっていることを示していることになる.

ベータ分布の確率密度函数の形から直接的に正規分布近似を導くことも易しい(次の節の問題を見よ).

注意終

In [119]:
```
function plot_gammanormal(α, θ)
    m, s = α*θ, √(α*θ^2)
    xlim = (m - 4s, m + 4s)
    plot(Gamma(α, θ), xlim...; label="Gamma(α,θ)")
    plot!(Normal(m, s); label="normal approx", ls=:dash)
    #plot!(; legend=:outertop)
    title!("α = $α, θ = $θ")
end
```

Out[119]: plot_gammanormal (generic function with 1 method)

In [120]:
```
plot(plot_gammanormal(5, 20), plot_gammanormal(20, 5); size=(800, 250))
```

Out[120]:

In [121]:
```
plot(plot_gammanormal(80, 1.25), plot_gammanormal(320, 0.3125); size=(800, 250))
```

Out[121]:

In [122]:
```
function plot_betanormal(α, β)
    m, s = α/(α+β), √(α*β/(α+β)^3)
    xlim = (m - 4s, m + 4s)
    plot(Beta(α, β), xlim...; label="Beta(α,β)")
    plot!(Normal(m, s); label="normal approx", ls=:dash)
    #plot!(; legend=:outertop)
    title!("α = $α, β = $β")
end
```

Out[122]: plot_betanormal (generic function with 1 method)

In [123]:
```
[3, 7] .* (4 .^ (0:3))'
```

Out[123]:
```
2×4 Matrix{Int64}:
 3  12   48  192
 7  28  112  448
```

In [124]:
```
plot(plot_betanormal(3, 7), plot_betanormal(12, 28); size=(800, 250))
```

Out[124]:

In [125]:
```
plot(plot_betanormal(48, 112), plot_betanormal(192, 448); size=(800, 250))
```

Out[125]:

4.10 問題: ベータ分布の正規分布近似の直接的証明

α, β が大きなときにベータ分布 $\mathrm{Beta}(\alpha, \beta)$ が正規分布で近似されることを, その確率密度函数の具体形を使って直接証明せよ.

解答例: $0 < \mu < 1$ と仮定し,

$$\alpha = N\mu, \quad \beta = N(1-\mu), \quad t = \mu + \frac{x}{\sqrt{N}}$$

とおいて, N を大きくすると, Stirlingの公式より,

$$\begin{aligned}
\frac{1}{B(\alpha,\beta)} &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\Gamma(N)}{\Gamma(N\mu)\Gamma(N(1-\mu))} \\
&\sim \frac{N\mu\ N(1-\mu)}{N} \frac{N^N e^{-N}\sqrt{2\pi N}}{(N\mu)^{N\mu}e^{-N\mu}\sqrt{2\pi N\mu}\,(N(1-\mu))^{N(1-\mu)}e^{-N(1-\mu)}\sqrt{2\pi N(1-\mu)}} \\
&= \frac{N\mu(1-\mu)}{\mu^{N\mu}\sqrt{2\pi N\mu}\,(1-\mu)^{N(1-\mu)}\sqrt{1-\mu}} \\
&= \frac{1}{\sqrt{2\pi\mu(1-\mu)}} \frac{\sqrt{N}}{\mu^{N\mu-1}(1-\mu)^{N(1-\mu)-1}}.
\end{aligned}$$

そして

$$
\begin{aligned}
&N\mu \log\left(1+\frac{x}{\sqrt{N}\,\mu}\right)\\
&= N\mu\left(\frac{x}{\sqrt{N}\,\mu}-\frac{x^2}{2N\mu^2}+O(N^{-3/2})\right)\\
&= \sqrt{N}\,x-\frac{x^2}{2\mu}+O(N^{-1/2}),\\
&N(1-\mu)\log\left(1-\frac{x}{\sqrt{N}(1-\mu)}\right)\\
&= N(1-\mu)\left(-\frac{x}{\sqrt{N}(1-\mu)}-\frac{x^2}{2N(1-\mu)^2}+O(N^{-3/2})\right)\\
&= -\sqrt{N}\,x-\frac{x^2}{2(1-\mu)}+O(N^{-1/2}),\\
&\frac{x^2}{2\mu}+\frac{x^2}{2(1-\mu)}=\frac{x^2}{2\mu(1-\mu)}
\end{aligned}
$$

より, 以下が成立することがわかる:

$$
\begin{aligned}
&t^{\alpha-1}(1-t)^{\beta-1}\,dt\\
&=\left(\mu+\frac{x}{\sqrt{N}}\right)^{N\mu-1}\left(1-\mu-\frac{x}{\sqrt{N}}\right)^{N(1-\mu)-1}\frac{dx}{\sqrt{N}}\\
&=\mu^{N\mu-1}(1-\mu)^{N(1-\mu)-1}\left(1+\frac{x}{\sqrt{N}\,\mu}\right)^{N\mu-1}\left(1-\frac{x}{\sqrt{N}(1-\mu)}\right)^{N(1-\mu)-1}\frac{dx}{\sqrt{N}}\\
&=\frac{\mu^{N\mu-1}(1-\mu)^{N(1-\mu)-1}}{\sqrt{N}}\exp\left(-\frac{x^2}{2\mu(1-\mu)}\right)(1+O(n^{-1/2}))\,dx.
\end{aligned}
$$

以上の結果より,

$$
\frac{t^{\alpha-1}(1-t)^{\beta-1}\,dt}{B(\alpha,\beta)}\,dt=\frac{1}{\sqrt{2\pi\mu(1-\mu)}}\exp\left(-\frac{x^2}{2\mu(1-\mu)}\right)(1+O(n^{-1/2}))\,dx.
$$

この公式はベータ分布が上で導入した x 座標では平均 0, 分散 $\mu(1-\mu)$ の正規分布で近似されることを意味している.

解答終

注意: $\alpha=N\mu$, $\beta=N(1-\mu)$, $T\sim\mathrm{Beta}(\alpha,\beta)$ のとき,

$$
E[T]=\frac{\alpha}{\alpha+\beta}=\mu,\qquad \mathrm{var}(T)=\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}=\frac{\mu(1-\mu)}{N+1}
$$

なので, $X=\sqrt{N}(T-\mu)$ とおくと, $N\to\infty$ のとき,

$$
E[X]=0,\qquad \mathrm{var}(X)=N\,\mathrm{var}(T)=\frac{N}{N+1}\mu(1-\mu)\to\mu(1-\mu).
$$

この $\mu(1-\mu)$ が上の解答例の中で登場した正規分布の分散である.

5 各種の確率分布の正規分布近似

以下に挙げるもの以外にも正規分布で近似される確率分布はたくさんある. 各自, 研究してノートにまとめておくとよいだろう. 具体的な数値例をコンピュータで作っておくと, どのようなときにどれだけ近似がうまく行くかについての「土地勘」が得られる.

In [126]:

```
function plot_normalapprox(dist; title=distname(dist), kwargs...)
    μ, σ = mean(dist), std(dist)
    normal = Normal(μ, σ)
    xlim = (μ - 4.5σ, μ + 4.5σ)
    plot(x -> mypdf(dist, x), xlim...; label="")
    plot!(normal, xlim...; label="normal approx", ls=:dash)
    plot!(; title, kwargs...)
end
```

Out[126]:

```
plot_normalapprox (generic function with 1 method)
```

In [127]:

```julia
# 2022-05-13: Distributions.jl で BetaNegativeBinomial(k, α, β)
# がまだ実装されていないので自前で必要最小限の実装を行う.

module O

using Distributions
using SpecialFunctions

struct BetaNegativeBinomial{T} <: DiscreteUnivariateDistribution
    k::T
    α::T
    β::T
end

function logpdf_betanegbin(x, k, α, β)
    m = round(x)
    m < 0 && return -Inf
    (loggamma(k+m) - loggamma(m+1) - loggamma(k) + logbeta(α+k, β+m) - logbeta(α, β))
end

function Distributions.logpdf(dist::BetaNegativeBinomial, x)
    (; k, α, β) = dist
    logpdf_betanegbin(x, k, α, β)
end

function Distributions.mean(dist::BetaNegativeBinomial)
    (; k, α, β) = dist
    k*β/(α - 1)
end

function Distributions.var(dist::BetaNegativeBinomial)
    (; k, α, β) = dist
    k*(α+k-1)*β*(α+β-1)/((α-1)^2*(α-2))
end

end
```

Out[127]: Main.O

5.1 二項分布の正規分布近似

$0 < p < 1$ でありかつ,

$$(X_1, \ldots, X_n) \sim \mathrm{Bernoulli}(p)^n = \mathrm{Binomial}(1, p)^n$$

であるとき,

$$K_n = X_1 + \cdots + X_n \sim \mathrm{Binomial}(n, p)$$

であり, 中心極限定理によって独立同分布な確率変数達の和の分布が正規分布で近似されることより, n が大きなとき, K_n の分布は正規分布で近似される.

In [128]:

```julia
plot_normalapprox(Binomial(50, 0.2))
```

Out[128]:

5.2 負の二項分布の正規分布近似

$0 < p < 1$ でありかつ,

$$(Y_1, \dots, Y_k) \sim \mathrm{Geometric}(p) = \mathrm{NegativeBinomial}(1, p)^k$$

であるとき,

$$N_k = Y_1 + \cdots + Y_k \sim \mathrm{NegativeBinomial}(k, p)$$

であり, 中心極限定理によって独立同分布な確率変数達の和の分布が正規分布で近似されることより, k が大きなとき, N_k の分布は正規分布で近似される.

In [129]:
```
plot_normalapprox(NegativeBinomial(50, 0.7))
```

Out[129]:

5.3 Poisson分布の正規分布近似

$\lambda > 0$ でありかつ,

$$(X_1, \dots, X_n) \sim \mathrm{Poisson}(\lambda/n)^n$$

であるとき,

$$K = X_1 + \cdots + X_n \sim \mathrm{Poisson}(\lambda)$$

であり, 中心極限定理によって独立同分布な確率変数達の和の分布が正規分布で近似されることより, λ が大きなとき, K の分布は正規分布で近似される.

In [130]:
```
plot_normalapprox(Poisson(30))
```

Out[130]:

5.4 ガンマ分布の正規分布近似

$\alpha > 0, \theta > 0$ でありかつ,

$$(Y_1, \dots, Y_n) \sim \mathrm{Gamma}(\alpha/n, \theta)$$

であるとき,

$$T = Y_1 + \cdots + Y_n \sim \mathrm{Gamma}(\alpha, \theta)$$

であり, 中心極限定理によって独立同分布な確率変数達の和の分布が正規分布で近似されることより, α が大きなとき, T の分布は正規分布で近似される.

In [131]:
```
plot_normalapprox(Gamma(50, 0.1))
```

Out[131]:

5.5 χ^2分布の正規分布近似

$\nu > 0$ でありかつ,

$$(Y_1, \ldots, Y_n) \sim \mathrm{Chisq}(\nu/n) = \mathrm{Gamma}(\nu/(2n), 2)$$

であるとき,

$$Y = Y_1 + \cdots + Y_n \sim \mathrm{Chisq}(\nu) = \mathrm{Gamma}(\nu/2, 2)$$

であり, 中心極限定理によって独立同分布な確率変数達の和の分布が正規分布で近似されることより, ν が大きなとき, Y の分布は正規分布で近似される.

In [132]:
```
plot_normalapprox(Chisq(100))
```

Out[132]:

5.6 逆ガンマ分布の正規分布近似

$\alpha > 0$ が大きなとき, 逆ガンマ分布 $\mathrm{InverseGamma}(\alpha, \theta)$ は正規分布で近似される.

In [133]:
```
plot_normalapprox(InverseGamma(100, 100); title="InverseGamma(100, 100)")
```

Out[133]:

5.7 ベータ分布の正規分布近似

すでに上で示したように, $\alpha, \beta > 0$ が大きなとき, ベータ分布 $\mathrm{Beta}(\alpha, \beta)$ は正規分布で近似される.

In [134]:
```
plot_normalapprox(Beta(15, 35))
```

Out[134]:

5.8 t 分布の正規分布近似

「Gauss積分, ガンマ函数, ベータ函数」のノートですでに示したように, $\nu > 0$ が大きなとき, t 分布 $\mathrm{TDist}(\nu)$ は標準正規分布で近似される.

In [135]:
```
plot_normalapprox(TDist(20))
```

Out[135]:

5.9 ベータプライム分布の正規分布近似

$\alpha, \beta > 0$ が大きなとき, ベータプライム分布 $\mathrm{BetaPrime}(\alpha, \beta)$ は正規分布で近似される. (証明に興味がある人は自分で証明してみよ. ベータプライム分布は F 分布のスケール変換に等しい.)

In [136]:
```
plot_normalapprox(BetaPrime(100, 200))
```

Out[136]:

5.10 F 分布の正規分布近似

$\nu_1, \nu_2 > 0$ が大きなとき, F 分布 $\mathrm{FDist}(\nu_1, \nu_2) = (\nu_2/\nu_1)\,\mathrm{BetaPrime}(\nu_1/2, \nu_2/2)$ は正規分布で近似される.

In [137]:
```
plot_normalapprox(FDist(200, 400))
```

Out[137]:

5.11 超幾何分布の正規分布近似

s, f, n が大きなとき, 超幾何分布 Hypergeometric(s, f, n) は正規分布で近似される.

In [138]:
```
plot_normalapprox(Hypergeometric(40, 60, 80))
```

Out[138]:

5.12 ベータ二項分布の正規分布近似

n, α, β が大きなとき, ベータ二項分布 BetaBinomial(n, α, β) は正規分布で近似される.

In [139]:
```
plot_normalapprox(BetaBinomial(30, 30, 70))
```

Out[139]:

5.13 ベータ負の二項分布の正規分布近似

k, α, β が大きなとき, ベータ二項分布 BetaNegativeBinomial(k, α, β) は正規分布で近似される.

In [140]:

```
plot_normalapprox(O.BetaNegativeBinomial(100, 300, 100);
    title="BetaNegativeBinomial(100, 200, 400)")
```

Out[140]:

In [ ]:

```
```

条件付き確率分布, 尤度, 推定, 記述統計

- 黒木玄

- 2022-05-22

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)

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 条件付き確率分布
  - 1.1 離散確率分布の条件付き確率分布
  - 1.2 条件付き期待値の基本性質(離散分布版)
  - 1.3 離散分布の条件付き確率分布の簡単な例
  - 1.4 問題: 離散分布の条件付き確率分布として二項分布が得られること
  - 1.5 離散分布の場合のBayesの定理
  - 1.6 2×2の分割表での条件付き確率分布(偽陽性率, 偽陰性率)
  - 1.7 必修の易しい計算問題: 有病率によって偽陽性率と偽陰性率がどのように変化するか
  - 1.8 有病率によって偽陽性率と偽陰性率がどのように変化するかの視覚化
  - 1.9 おまけの問題: モンティ・ホール問題 (Monty Hall problem)
  - 1.10 連続確率分布の条件付き確率分布
  - 1.11 条件付き期待値の基本性質(連続分布版)
  - 1.12 連続分布版のBayesの定理
  - 1.13 問題: 連続分布の条件付き確率分布の例 (分散が固定された正規分布モデルのBayes統計)
  - 1.14 分散が固定された正規分布モデルのBayes統計の事後分布の視覚化
  - 1.15 問題: 2変量の正規分布とその条件付き確率分布の例
  - 1.16 2変量の正規分布とその条件付き確率分布の例の視覚化
- 2 2×2の分割表の分布
  - 2.1 4つのPoisson分布の積
  - 2.2 四項分布
  - 2.3 2つの二項分布の積
  - 2.4 Fisherの非心超幾何分布
  - 2.5 Fisherの非心超幾何分布の正規分布近似の視覚化
  - 2.6 四項分布は4つのPoisson分布の積の条件付き確率分布
  - 2.7 2つの二項分布の積は四項分布は4つの条件付き確率分布
  - 2.8 Fisherの非心超幾何分布は2つの二項分布の積の条件付き確率分布
  - 2.9 独立性条件
  - 2.10 Pearsonのχ^2統計量
- 3 尤度 (ゆうど)と推定
  - 3.1 尤度に関するまとめ
  - 3.2 尤度函数と尤度の定義と諸注意
  - 3.3 例: Bernoulli試行の尤度函数と最尤法
  - 3.4 Bernoulli試行モデルの尤度函数の視覚化
  - 3.5 問題: 尤度が高くても全然もっともらしくない例(オーバーフィッティングの例)
  - 3.6 問題: 正規分布の標本分布モデルの尤度函数と最尤法
  - 3.7 正規分布の標本分布モデルの尤度函数の視覚化
  - 3.8 問題: 最小二乗法がある種の正規分布モデルの最尤法に一致すること
  - 3.9 最尤法の観点から見た十分統計量
    - 3.9.1 Bernoulli試行モデルの十分統計量
    - 3.9.2 正規分布の標本分布モデルの十分統計量
    - 3.9.3 最小二乗法を与える正規分布モデルの尤度函数を記述する統計量
- 4 記述統計 (要約統計)

4.1 n 個の 1 と 0 からなるデータ $x_1, \ldots, x_n$ の要約
4.2 より一般の離散的なデータの要約
4.3 n 個の実数からなるデータ $x_1, \ldots, x_n$ の要約
4.4 n 個の実数の対からなるデータ $(x_1, y_1), \ldots, (x_n, y_n)$ の要約
4.5 Anscombeの例(アンスコムの例, 再)
4.6 問題: DataSaurusの例
4.6.1 Julia言語を使う解答例
4.6.2 Julia言語の使い方
4.6.3 インターネット上の各種計算サイトを使う方法

In [1]:

```
ENV["LINES"], ENV["COLUMNS"] = 100, 100
using BenchmarkTools
using Distributions
using LinearAlgebra
using Printf
using QuadGK
using Random
Random.seed!(4649373)
using Roots
using SpecialFunctions
using StaticArrays
using StatsBase
using StatsFuns
using StatsPlots
default(fmt = :png, titlefontsize = 10, size = (400, 250))
using SymPy
```

In [2]:

```
# Override the Base.show definition of SymPy.jl:
# https://github.com/JuliaPy/SymPy.jl/blob/29c5bfd1d10ac53014fa7fef468bc8deccadc2fc/src/types.

@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::SymbolicObject)
    print(io, as_markdown("\\displaystyle " * sympy.latex(x, mode="plain", fold_short_frac=fal
end
@eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::AbstractArray{Sym})
    function toeqnarray(x::Vector{Sym})
        a = join(["\\displaystyle " * sympy.latex(x[i]) for i in 1:length(x)], "\\\\")
        """\\left[ \\begin{array}{r}$a\\end{array} \\right]"""
    end
    function toeqnarray(x::AbstractArray{Sym,2})
        sz = size(x)
        a = join([join("\\displaystyle " .* map(sympy.latex, x[i,:]), "&") for i in 1:sz[1]],
        "\\left[ \\begin{array}{" * repeat("r",sz[2]) * "}" * a * "\\end{array}\\right]"
    end
    print(io, as_markdown(toeqnarray(x)))
end
```

In [3]:

```
safemul(x, y) = x == 0 ? x : x*y
safediv(x, y) = x == 0 ? x : x/y

x ≲ y = x < y || x ≈ y

mypdf(dist, x) = pdf(dist, x)
mypdf(dist::DiscreteUnivariateDistribution, x) = pdf(dist, round(Int, x))

distname(dist::Distribution) = replace(string(dist), r"{.*}" => "")
myskewness(dist) = skewness(dist)
mykurtosis(dist) = kurtosis(dist)
function standardized_moment(dist::ContinuousUnivariateDistribution, m)
    μ, σ = mean(dist), std(dist)
    quadgk(x -> (x - μ)^m * pdf(dist, x), extrema(dist)...)[1] / σ^m
end
myskewness(dist::MixtureModel{Univariate, Continuous}) = standardized_moment(dist, 3)
mykurtosis(dist::MixtureModel{Univariate, Continuous}) = standardized_moment(dist, 4) - 3
```

Out[3]:

```
mykurtosis (generic function with 2 methods)
```

1 条件付き確率分布

我々は複雑な現実世界に立ち向かうために, 各種の確率分布を数学的モデル(統計モデル)として採用して, モデルとデータを比較することによって, 現実世界の様子を推測しようとする.

そのときに, 統計モデル内部に現実世界で得たデータの数値を持ち込んで利用する方法として,

- 条件付き確率分布を考えること. 統計モデル(数学的フィクション)内部で現実世界から得たデータと同じ数値が生成された場合に制限した条件付き確率分布を考える.
- 尤度を考えること. 統計モデル(数学的フィクション)内部で現実世界から得たデータと同じ数値が生成される確率またはその密度を考える.

の2つが特に基本的である.

以下では前者の条件付き確率分布の概念について一般的に説明する.

1.1 離散確率分布の条件付き確率分布

離散集合 $\mathscr{X}$ (大文字のグザイ)の中を動く離散変数 x とその確率質量函数 $P(x)$ によって, 離散確率分布 D が与えられているとする. (注意: x は $(x_1, \ldots, x_n)$ のように多変数を1つにまとめて書いたものかもしれない. その場合には Ξ は $\mathbb{R}^n$ の離散部分集合であると考える.)

集合 Ξ 上の函数 $y = f(x)$ が与えられたとき, 確率変数 $Y = f(X)$ が得られる. この Y の函数の期待値は

$$E[g(Y)] = \sum_{x \in \mathscr{X}} g(f(x))P(x)$$

と定義される. (確率変数とはその函数の期待値が定義されているような変数であると言ってよい.)

数値 y について, $P(Y = y) > 0$ であると仮定する.

条件 $Y = y$ (もしくは条件 $f(x) = y$)が定める **条件付き確率分布** (conditional probability distribution) $D|_{Y=y}$ を以下のように定める:

(1) 条件付き確率分布 $D|_{Y=y}$ は離散集合 $\mathscr{X}_y = \{\, x \in \mathscr{X} \mid f(x) = y \,\}$ 上の確率分布である.

(2) 条件付き確率分布 $D|_{Y=y}$ は確率質量函数

$$P(x|y) = \frac{P(x)}{P(Y = y)}, \quad P(Y = y) = \sum_{f(x)=y} P(x) = \sum_{x \in \mathscr{X}_y} P(x)$$

によって定義される. このとき,

$$\sum_{x \in \mathscr{X}_y} P(x|y) = \frac{\sum_{x \in \mathscr{X}_y} P(x)}{P(Y = y)} = 1$$

なので確かにこの函数は「確率の総和が1になる」という条件を満たしており, 確率質量函数になっている.

条件 $Y = y$ という条件のもとでの変数 x に対応する確率変数 X と y の函数 $f(X, y)$ の **条件付き期待値** を次のように定める:

$$E[f(X, Y)|Y = y] = E[f(X, y)|y] = \sum_x f(x, y)p(x|y).$$

これと同じ記号法を一般のパラメータ θ 付きの確率質量函数 $P(x|\theta)$ の場合にも使うことにする:

$$E[f(X, \theta)|\theta] = \sum_x f(x, \theta)P(x|\theta).$$

1.2 条件付き期待値の基本性質(離散分布版)

条件付き期待値 $g(y) = E[f(X, y)|y]$ は y の函数とみなせるので, 確率変数 $g(Y) = E[f(X, Y)|Y]$ が得られる. このとき次が成立することを示せ:

$$E[g(Y)] = E[E[f(X, Y)|Y]] = E[f(X, Y)].$$

解答例: 定義通りに計算しよう:

$$
\begin{aligned}
E[f(X,Y)|Y] = E[g(Y)] &= \sum_y g(y)P(y) \\
&= \sum_y E[f(X,y)|y]P(y) \\
&= \sum_y \left(\sum_x f(x,y)P(x|y) \right) P(y) \\
&= \sum_{x,y} f(x,y)P(x|y)P(y) \\
&= \sum_{x,y} f(x,y)P(x,y) = E[f(X,Y)].
\end{aligned}
$$

要するに x, y の同時確率質量函数が $P(x,y) = P(x|y)P(y)$ と書けることから, この問題の結論が出て来る.

解答終

注意: 測度論的確率論ではこれを一般化して条件付き期待値を定義する(Radon–Nikodýmの定理の話になる).

注意: 個人的な意見では以上のようなシンプルで初等的な例さえ知らずに, いきなり測度論的確率論を勉強しようとするのは不健全な勉強の仕方で好ましくないと思う. 測度論が無用な事柄について多くの意味のある具体例を知った後でそれらを測度論を使って整理し直すようにした方が測度論的確率論についてもスムーズに理解できように思われる.

1.3 離散分布の条件付き確率分布の簡単な例

正二十面体のサイコロの20個の面のうち 6 つの面には1と書いてあり, 5つの面には2と書いてあり, 4つの面には3と書いてあり, 3つの面には4と書いてあり, 残りの2つの面には5と書いてあるとする. そのサイコロをふったときにすべての面は等確率で出るとする. このサイコロの出目の確率分布は

$$
P(1) = \frac{6}{20}, \quad P(2) = \frac{5}{20}, \quad P(3) = \frac{4}{20}, \quad P(4) = \frac{3}{20}, \quad P(5) = \frac{2}{20}
$$

によって定まる集合 $\mathscr{X} = \{1,2,3,4,5\}$ 上の有限離散分布によってモデル化される.

$\mathscr{X}$ 上の函数 $f(x)$ を x が偶数ならば $f(x) = 0$, x が奇数ならば $f(x) = 1$ と定める.

このとき, 確率変数 $Y = f(X)$ に関する条件 $Y = y$ が定める条件付き確率分布を求めてみよう.

この場合の $\mathscr{X}_y = \{\, x \in \mathscr{X} \mid f(x) = y \,\}$ は次のようになる:

$$
\mathscr{X}_0 = \{2,4\}, \quad \mathscr{X}_1 = \{1,2,3\}.
$$

さらに,

$$
P(Y=0) = P(2) + P(4) = \frac{8}{20}, \quad P(Y=1) = P(1) + P(3) + P(5) = \frac{12}{20}
$$

となる. ゆえに, $P(x|0)$ $(x \in \mathscr{X}_0)$ は

$$
\begin{aligned}
P(2|0) &= \frac{P(2)}{P(Y=0)} = \frac{5/20}{8/20} = \frac{5}{8}, \\
P(4|0) &= \frac{P(4)}{P(Y=0)} = \frac{3/20}{8/20} = \frac{3}{8}
\end{aligned}
$$

となり, $P(x|1)$ $(x \in \mathscr{X}_1)$ は

$$
\begin{aligned}
P(1|0) &= \frac{P(1)}{P(Y=1)} = \frac{6/20}{12/20} = \frac{6}{12}, \\
P(2|0) &= \frac{P(3)}{P(Y=1)} = \frac{4/20}{12/20} = \frac{4}{12}, \\
P(3|0) &= \frac{P(5)}{P(Y=1)} = \frac{2/20}{12/20} = \frac{2}{12}.
\end{aligned}
$$

となる. 偶数の目は2,4の2通りで20面のうち8面に2,4と書かれており, そのうちの5面が2で残りの3面が4なので, 偶数の目に制限した場合に2の目が出る条件付き確率は $P(2|0) = 5/8$ になり, 4の目が出る条件付き確率は $P(4|0) = 3/8$ になる. 奇数の目の場合も同様である.

1.4 問題: 離散分布の条件付き確率分布として二項分布が得られること

2つの0以上の整数の組 (x, y) 全体の集合 $\mathscr{X} = \mathbb{Z}_{\geq 0}^2 = \{ (x, y) \mid x, y = 0, 1, 2, \dots \}$ 上の確率質量函数を次のようにPoisson分布の確率質量函数の積で定める($\lambda, \mu > 0$ と仮定する):

$$P(x, y) = e^{-(\lambda+\mu)} \frac{\lambda^x}{x!} \frac{\mu^y}{y!} \quad (x, y = 0, 1, 2, \dots)$$

集合 $\mathscr{X}$ 上の函数 $f(x, y)$ を $f(x, y) = x + y$ と定める. $N = f(X, Y) = X + Y$ によって確率変数 N を定めることができる(ここで X, Y はいつものように変数 x, y に対応する確率変数である). 確率変数 N の函数の期待値は次のように表される:

$$E[g(N)] = \sum_{(x,y)\in\mathscr{X}} g(x + y)P(x, y).$$

n が 0 以上の整数であるとする. 以下を示せ:

(1) N は平均 $\lambda + \mu$ のPoisson分布に従う.

(2) この場合には, 条件 $N = n$ が定める条件付き確率分布は本質的に二項分布になる.

解答例: $\mathscr{X}_n = \{ (x, y) \in \mathscr{X} \mid f(x, y) = x + y = n \}$, $P(N = n)$, $P(x, y|n) = P(x, y)/P(N = n)$ $((x, y \in \mathscr{X}_n)$ を順番に計算して行こう. まず,

$$\mathscr{X}_n = \{ (x, n - x) \mid x = 0, 1, 2, \dots, n\}$$

であることはすぐにわかる. 次に,

$$\begin{aligned} P(N = n) &= \sum_{(x,y)\in\mathscr{X}_n} P(x, y) = \sum_{x=0}^{n} e^{-(\lambda+\mu)} \frac{\lambda^x}{x!} \frac{\mu^{n-x}}{(n - x)!} \\ &= e^{-(\lambda+\mu)} \frac{1}{n!} \sum_{x=0}^{n} \frac{n!}{x!(n - x)!} \lambda^x \mu^{n-x} = e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^n}{n!}. \end{aligned}$$

これは確率変数 N が平均 $\lambda + \mu$ のPoisson分布 $\mathrm{Poisson}(\lambda + \mu)$ に従うことを意味し, (1)が示された. したがって, $(x, n - x) \in \mathscr{X}_n$ のとき,

$$\begin{aligned} P(x, n - x|n) &= \frac{P(x, n - x)}{P(N = n)} = \frac{e^{-(\lambda+\mu)}(\lambda^x/x!)(\mu^{n-x}/(n - x)!)}{e^{-(\lambda+\mu)}(\lambda + \mu)^n/n!} \\ &= \frac{n!}{x!(n - x)!} \left(\frac{\lambda}{\lambda + \mu} \right)^x \left(\frac{\mu}{\lambda + \mu} \right)^{n-x} \\ &= \frac{n!}{x!(n - x)!} \left(\frac{\lambda}{\lambda + \mu} \right)^x \left(1 - \frac{\lambda}{\lambda + \mu} \right)^{n-x}. \end{aligned}$$

これは条件 $N = n$ が定める条件付き確率分布が本質的に二項分布 $\mathrm{Binomial}(n, \lambda/(\lambda + \mu))$ であることを意味している. これで(2)も示された.

解答終

1.5 離散分布の場合のBayesの定理

条件付き確率分布の概念を理解していれば, Bayesの定理(ベイズの定理)については忘れてもよい. しかし, Bayesの定理という用語が使われる頻度は高いので念のために説明しておくことにする.

離散集合 $\mathscr{X} = \{(x, y)\}$ 上の離散確率分布が確率質量函数 $P(x, y)$ によって定義されているとする. 以下では $P(X = x)$ や $P(Y = y)$ を以下のように書くことにする:

$$P(x) = P(X = x) = \sum_y P(x, y), \quad P(y) = P(Y = y) = \sum_x P(x, y).$$

ここで, 前者の式の $\sum_y$ の y は各 x ごとに $P(x, y)$ が 0 にならないような y 全体を動き, 後者の式の $\sum_x$ の x は各 y ごとに $P(x, y)$ が 0 にならないような x 全体を動く.

記号法上の注意: 慣習に従って $P(x)$ と $P(y)$ の違いを独立変数名 x, y の違いで区別するという記号法を採用している. これは簡潔に書くための手抜きであり, $P_X(x)$ や $P_Y(y)$ のように書けば独立変数名の違いで区別する必要はなくなる. 以下では簡潔さを重視するスタイルを採用したので混乱しないように注意して欲しい. $P(x|y)$ と $P(y|x)$ の違いも独立変数名 x, y の違いで区別することになる.

このとき, y に対応する確率変数 Y によって $Y = y$ という条件で定義される条件付き確率分布の確率質量函数

$$P(x|y) = \frac{P(x, y)}{P(y)}, \quad P(y) = \sum_x P(x, y) \tag{1}$$

だけではなく, x に対応する確率変数 X によって $X = x$ という条件で定義される条件付き確率分布の確率質量函数

$$P(y|x) = \frac{P(x,y)}{P(x)}, \qquad P(x) = \sum_y P(x,y) \tag{2}$$

が定義される. このとき

$$\begin{aligned} P(x,y) &= P(x|y)P(y) = P(y|x)P(x), \\ P(x) &= \sum_y P(x,y) = \sum_y P(x|y)P(y), \\ P(y) &= \sum_x P(x,y) = \sum_x P(y|x)P(x) \end{aligned} \tag{3}$$

なので,

$$\begin{aligned} P(y|x) &= \frac{P(x|y)P(y)}{\sum_y P(x|y)P(y)}, \\ P(x|y) &= \frac{P(y|x)P(x)}{\sum_x P(y|x)P(x)}. \end{aligned} \tag{B}$$

この結果(B)を **Bayesの定理** (ベイズの定理)と呼ぶ.

しかし, 以上の議論を見ればわかるように, Bayesの定理は条件付き確率分布の定義を書き直したものに過ぎず, 価値ある新しい考え方が得られるような結果ではない. しかし, 具体的もしくは理論的な計算で条件付き確率の概念が(B)の形式で使われることがあることは知っておいて損がないと思われるので, このように紹介することにした.

注意: 多くの初学者はBayesの定理を「与えられた公式」として直接使う計算を直観的に理解できないことをやってしまっているという理由でよく間違ってしまうようだ. そのようになってしまいそうな人は無理して(B)の公式を直接使おうとせずに, (1),(2)(および(3))まで戻って地道に計算するようにした方がよいだろう.

注意: 以上のようにBayesの定理は条件付き確率の定義から自明に導かれる無理して使う必要がない公式に過ぎない. そのような自明でつまらない結果について, 「逆確率」「主観確率」のような用語を用いて特別な価値があるかのように説明するスタイルが伝統的になってしまっているので惑わされないように注意が必要である. Bayesの定理は, 条件付き確率の定義を書き直しただけの, 「逆確率」「主観確率」のような用語と無関係に一般的かつ普遍的に成立している自明な数学的定理に過ぎない.

注意: 所謂Bayes統計についても条件付き確率分布の概念まで戻って考えればBayesの定理を使わずに理解可能である.

1.6 2×2の分割表での条件付き確率分布(偽陽性率, 偽陰性率)

病気Dに罹っているいるかどうかに関するある検査法を使うと,

- 病気Dに罹っている人は 75% の確率で陽性だと判定でき,
- 病気Dに罹っていない人は 95% の確率で陰性と判定できるもの

と仮定する. このとき, この検査法の **感度** は 75% であり, **特異度** は 95% であるという.

この検査を病気Dに罹っている確率が p の人(**有病率** p の人)に適用する状況について考える. このとき, その人が病気Dに罹っているか否かと検査結果が陽性か陰性かで分類することによって, 次のような 2×2 の確率の表を作れる:

	病気有	病気無	
陽性	$0.75p$	$0.05(1-p)$	$0.05 + 0.70p$
陰性	$0.25p$	$0.95(1-p)$	$0.95 - 0.70p$
	p	$1-p$	1

下段の $p, 1-p$ や右端の $0.05 + 0.70p, 0.95 - 0.70p$ は確率の縦もしくは横の合計である. それらをよく **マージン** (margin, 周辺確率)と呼ぶ.

この確率の2×2の表の部分に対応する確率質量函数は以下のように書ける:

$$\begin{aligned} P(\text{陽性}, \text{病気有}) &= 0.75p, & P(\text{陽性}, \text{病気無}) &= 0.05(1-p), \\ P(\text{陰性}, \text{病気有}) &= 0.25p, & P(\text{陰性}, \text{病気無}) &= 0.95(1-p). \end{aligned}$$

さらに, マージンとして以下の確率も上の表にすでに書き込まれている:

$$\begin{aligned} P(\text{病気有}) &= P(\text{陽性}, \text{病気有}) + P(\text{陰性}, \text{病気有}) = p, \\ P(\text{病気無}) &= P(\text{陽性}, \text{病気無}) + P(\text{陰性}, \text{病気無}) = 1-p, \\ P(\text{陽性}) &= P(\text{陽性}, \text{病気有}) + P(\text{陽性}, \text{病気無}) = 0.05 + 0.70p, \\ P(\text{陰性}) &= P(\text{陰性}, \text{病気有}) + P(\text{陰性}, \text{病気無}) = 0.95 - 0.70p. \end{aligned}$$

上の表から, 陽性または陰性であるという条件によって定まる病気の有無に関する条件付き確率が以下のように計算される:

$$
\begin{aligned}
(\text{陽性的中率}) &:= P(\text{病気有}|\text{陽性}) = \frac{P(\text{陽性}, \text{病気有})}{P(\text{陽性})} = \frac{0.75p}{0.05 + 0.70p}, \\
(\text{偽陽性率}) &:= P(\text{病気無}|\text{陽性}) = \frac{P(\text{陽性}, \text{病気無})}{P(\text{陽性})} = \frac{0.05(1-p)}{0.05 + 0.70p}, \\
(\text{偽陰性率}) &:= P(\text{病気有}|\text{陰性}) = \frac{P(\text{陰性}, \text{病気有})}{P(\text{陰性})} = \frac{0.25p}{0.95 - 0.70p}, \\
(\text{陰性的中率}) &:= P(\text{病気無}|\text{陰性}) = \frac{P(\text{陰性}, \text{病気無})}{P(\text{陰性})} = \frac{0.95(1-p)}{0.95 - 0.70p}.
\end{aligned}
$$

これらの式は, 私には見難く, 上の確率の表を直接見た方がわかり易いように感じられる. しかし, 議論の内容を正確に把握できるようにするためにあえて式も書いてみた.

注意: 伝統的には以上の内容はBayesの定理の応用として説明されることが多い. しかし, 上の説明ではBayesの定理は一切使用する必要がなかった. このことからもBayesの定理が必須ではないことがわかる. 最後に得られた陽性的中立, 偽陽性率, 偽陰性率, 陰性的中率の公式達はちょうどBayesの定理の形をしている. そして, すぐ上でも述べたように見易い結果ではない. 実際の計算ではそれらのBayesの定理を使っても得られる公式を使うよりも, 確率の表を地道に書いて, 表を見ながら直接計算した方が間違う可能性も減り, 直観的な意味も理解し易いだろう.

以上を踏まえて次の問題を解け.

1.7 必修の易しい計算問題: 有病率によって偽陽性率と偽陰性率がどのように変化するか

前節の状況において, 有病率 p が $5\%, 10\%, 20\%, 40\%, 80\%$ の場合の偽陽性率と偽陰性率を求めよ.

解答例: 前節で求めた偽陽性率と偽陰性率の公式に $p = 0.05, 0.1, 0.2, 0.4, 0.8$ を代入すれば求まるが, 以下では地道に確率の表を書いて求めてみよう. (前節で求めた公式を使った人は以下の地道なやり方を採用しても手間がそう増えず, 状況をより把握し易くなっていることを確認して欲しい.)

前節の確率の表を引用しよう:

	病気有	病気無	
陽性	$0.75p$	$0.05(1-p)$	$0.05 + 0.70p$
陰性	$0.25p$	$0.95(1-p)$	$0.95 - 0.70p$
	p	$1-p$	1

$p = 0.05$ の場合:

	病気有	病気無	
陽性	3.75%	4.75%	8.5%
陰性	1.25%	90.25%	91.5%
	5%	95%	100%

$p = 0.1$ の場合:

	病気有	病気無	
陽性	7.5%	4.5%	12%
陰性	2.5%	85.5%	88%
	10%	90%	100%

$p = 0.2$ の場合:

	病気有	病気無	
陽性	15%	4%	19%
陰性	5%	76%	81%
	20%	80%	100%

$p = 0.4$ の場合:

	病気有	病気無	
陽性	30%	3%	33%
陰性	10%	57%	67%
	40%	60%	100%

$p = 0.8$ の場合:

	病気有	病気無	
陽性	60%	1%	61%
陰性	20%	19%	39%
	80%	20%	100%

以上の表を眺めると, 偽陽性率(陽性になった場合の病気無の条件付き確率)と偽陰性率(陰性になった場合の病気有の条件付き確率)は以下の表のようになることがわかる:

有病率	偽陽性率	偽陰性率
5%	$4.75/8.5 \approx 56\%$	$1.25/91.5 \approx 1.4\%$
10%	$4.5/12 \approx 38\%$	$2.5/88 \approx 2.8\%$
20%	$4/19 \approx 21\%$	$5/81 \approx 6.2\%$
30%	$3/33 \approx 9.1\%$	$10/67 \approx 15\%$
40%	$1/61 \approx 1.6\%$	$20/39 \approx 51\%$

有病率が低いと偽陽性率が高くなり, 有病率が高いと偽陰性率が高くなる.

解答終

In [4]:
```
function falsepositive_falsenegative(prevalence_rate, sensitivity, specificity)
    a = positive_besick  = sensitivity       * prevalence_rate
    b = negative_besick  = (1 - sensitivity) * prevalence_rate
    c = positive_notsick = (1 - specificity) * (1 - prevalence_rate)
    d = negative_notsick = specificity       * (1 - prevalence_rate)
    positive = positive_besick + positive_notsick
    negative = negative_besick + negative_notsick
    falsepositive = positive_notsick / positive
    falsenegative = negative_besick  / negative
    falsepositive, falsenegative
end
```

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

In [5]:
```
falsepositive_falsenegative.([0.05, 0.1, 0.2, 0.4, 0.8], 0.75, 0.95) .▷
x → round.(x; digits=4)
```

Out[5]:
```
5-element Vector{Tuple{Float64, Float64}}:
 (0.5588, 0.0137)
 (0.375, 0.0284)
 (0.2105, 0.0617)
 (0.0909, 0.1493)
 (0.0164, 0.5128)
```

1.8 有病率によって偽陽性率と偽陰性率がどのように変化するかの視覚化

In [6]:

```
function plot_falsepositive(prevalence_rate, sensitivity, specificity;
        xsize = 20, ysize = xsize, ms = 5, size = (480, 300), kwargs...)
    N = xsize * ysize
    a = positive_besick  = sensitivity       * prevalence_rate
    b = negative_besick  = (1 - sensitivity) * prevalence_rate
    c = positive_notsick = (1 - specificity) * (1 - prevalence_rate)
    d = negative_notsick = specificity       * (1 - prevalence_rate)
    A = round(Int, a*N)
    B = round(Int, b*N)
    C = round(Int, c*N)
    D = round(Int, d*N)
    xsize_besick  = (A + B + ysize - 1) ÷ ysize
    xsize_notsick = (C + D + ysize - 1) ÷ ysize
    TP = [(mod(i, xsize_besick), -(i ÷ xsize_besick)) for i in 0:A-1]
    FN = [(mod(i, xsize_besick), -(i ÷ xsize_besick)) for i in A:A+B-1]
    FP = [(xsize_besick + mod(i, xsize_notsick), -(i ÷ xsize_notsick)) for i in 0:C-1]
    TN = [(xsize_besick + mod(i, xsize_notsick), -(i ÷ xsize_notsick)) for i in C:C+D-1]
    P = plot(; frame=false, axis=false, tick=false, margin=0Plots.mm)
    scatter!(TP; m=:utriangle, msc=:red, mc=:white, label="true-positive", ms)
    scatter!(FN; m=:diamond, msc=:red, mc=:white, label="false-negative", ms, msw=3)
    scatter!(FP; m=:dtriangle, msc=:blue, mc=:white, label="false-positive", ms, msw=3)
    scatter!(TN; m=:square, msc=:blue, mc=:white, label="true-negative", ms)
    plot!(; legend = :outertopright)
    plot!(; xlim = (-0.5, xsize+0.5), ylim = (-ysize-0.5, 0.5))
    plot!(; size, kwargs...)
 end
```

Out[6]:

```
plot_falsepositive (generic function with 1 method)
```

以下は, 感度(病気有の人が検査陽性になる確率)が 75% で特異度(病気無の人が検査陽性になる確率)が 95% の場合.

赤色は病気有を, 青色は病気無を意味する. すなわち, 赤い上向き三角と赤いひし形は病気Dに罹っている人で, 青の下向き三角と青の正方形は病気Dに罹っていない人である.

三角は検査陽性を意味し, 四角は検査陰性を意味する. すなわち, 赤い上向き三角と青の下向き三角は検査で陽性になった人達で, それ以外のひし形と正方形は陰性になった人達である.

青い下向き三角 は病気Dに罹っていないのに検査で陽性になってしまったので **偽陽性** (**false-positive**). 目立つように太くしてある. (青い正方形は病気有でかつ検査陽性なので真陽性 (true-negative).)

赤いひし形 は病気Dに罹っているのに検査で陰性になってしまったので **偽陰性** (**false-negative**). これも目立つように太くしてある. (青い正方形は病気無でかつ検査陰性なので真陰性 (true-negative).)

以下の図を見れば, 有病率(赤色の割合=赤い上向き三角と赤いひし形を合わせた割合)を変えたときにどのように偽陽性と偽陰性の割合が変化するかが分かる.

In [7]:

```
plot_falsepositive(0.05, 0.75, 0.95)
```

Out[7]:

In [8]:
```
plot_falsepositive(0.10, 0.75, 0.95)
```

Out[8]:

In [9]:
```
plot_falsepositive(0.20, 0.75, 0.95)
```

Out[9]:

In [10]:
```
plot_falsepositive(0.40, 0.75, 0.95)
```

Out[10]:

In [11]:
```
plot_falsepositive(0.80, 0.75, 0.95)
```

Out[11]:

感度(病気有の人が検査陽性になる確率)を 75% から 95% に上げて, 特異度(病気無の人が検査陽性になる確率)が 75% に下がった場合. 当たり前のことだが, 偽陽性(青い下向き三角)が大幅に増える. 検査対象の有病率を 20% まで上げても, 検査陽性者のうち半分以上が病気無になってしまう.

In [12]:
```
plot_falsepositive(0.05, 0.95, 0.75)
```

Out[12]:

In [13]:
```
plot_falsepositive(0.10, 0.95, 0.75)
```

Out[13]:

In [14]:
```
plot_falsepositive(0.20, 0.95, 0.75)
```

Out[14]:

In [15]:
```
plot_falsepositive(0.40, 0.95, 0.75)
```

Out[15]:

In [16]:
```
plot_falsepositive(0.80, 0.95, 0.75)
```

Out[16]:

1.9 おまけの問題: モンティ・ホール問題 (Monty Hall problem)

以下のようなゲームを考える:

(1) 親は $1, 2, 3$ の中から番号 X を無作為に選んで秘密にしておく.

(2) あなたは, $1, 2, 3$ の中から番号 1 を選んでそれを親に告げる.

(3) 親は $1, 2, 3$ の中から 1 と X 以外の番号 Y を無作為に選んで, あなたに告げる.

(4) あなたは $1, 2, 3$ の中から $1, Y$ を除いて残った1つの番号 Z を選ぶ.

あなたが最後に選んだ番号 Z が秘密の番号 X に一致する確率を求めよ.

注意: (2)で 1 を選ばずに $1, 2, 3$ のどれかを無作為に選んでも結果は同じになる.

代表的誤答例: $Z = X$ となる確率は $1/2$ である. **誤答終**

解答例1: すべての場合の確率を書き下す方法. 確率 $P(X = x, Y = y)$ 達を表にすると以下のようになる:

	$Y = 1$	$Y = 2$	$Y = 3$	
$X = 1$	0	1/6	1/6	1/3
$X = 2$	0	0	1/3	1/3
$X = 3$	0	1/3	0	1/3
	0	1/2	1/2	

Y として 1 と X 以外の番号を選ぶので, $Y = 1$ の確率は 0 になる. $X = 1$ となる確率は $1/3$ であり, そのとき, 1 と $X = 1$ 以外の番号は $2, 3$ の2つで, それらを無作為に選ぶのでそれぞれに確率は $1/6$ になる. $X = 2$ の場合には常に $Y = 3$ となり, $X = 3$ の場合には常に $Y = 2$ となり, それぞれの確率は $1/3$ になる. これを表にすると上のようになる.

- $X = 1, Y = 2$ のとき $Z = 3$ となり, $Z \neq X$ となる.
- $X = 1, Y = 3$ のとき $Z = 2$ となり, $Z \neq X$ となる.
- $X = 2, Y = 3$ のとき $Z = 2$ となり, $Z = X$ となる.
- $X = 3, Y = 2$ のとき $Z = 3$ となり, $Z = X$ となる.

ゆえに $P(Z = X) = P(X = 2, Y = 3) + P(X = 3, Y = 2) = 1/3 + 1/3 = 2/3$.

解答終

解答例2: あなたが最後に選んだ番号 Z が X に一致しないことと, あなたが最初に選んだ番号 1 が X に一致することは同値であり, $X = 1$ となる確率は $1/3$ である. ゆえに, その補事象である $Z = X$ となる確率は $2/3$ である.

要するに, 最初に選んだ数字が運悪く X に一致していなければ, 最後に選び直して得た Z は X に一致する. たったこれだけの話である.

解答終

注意: 単に確率を計算すればよいだけの問題である. 「主観確率」「ベイズ確率」のような用語を持ち出す説明を読むときには, 本当は易しい話を無駄に難しく考えて, おかしな考え方に陥っていることを疑った方がよい. 歴史の偶然のせいで世間的に広まってしまった考え方の多くが間違っている.

1.10 連続確率分布の条件付き確率分布

連続変数 (x, y) の確率密度函数 $p(x, y)$ が定める連続分布 D を考え, 変数 y に対応する確率変数を Y と書く:

$$E[f(Y)] = \iint f(y)p(x, y)\,dx\,dy = \int f(y)\left(\int p(x, y)\,dx\right)dy.$$

これより, 確率変数 Y が従う分布は確率密度函数

$$p(y) = \int p(x, y)\,dx$$

を持つことがわかる. 条件 $Y = y$ が定める x に関する **条件付き確率分布** の密度函数 $p(x|y)$ を次のように定める:

$$p(x|y) = \frac{p(x, y)}{p(y)}.$$

このとき,

$$\int p(x|y)\,dx = \frac{\int p(x, y)\,dx}{p(y)} = 1$$

より, $p(x|y)$ がパラメータ y を持つ x の確率密度函数になっていることがわかる.

条件 $Y = y$ という条件のもとでの変数 x に対応する確率変数 X と y 付きの函数 $f(X, y)$ の **条件付き期待値** を次のように定める:

$$E[f(X, Y)|Y = y] = E[f(X, y)|y] = \int f(x, y)p(x|y)\,dx.$$

これと同じ記号法を一般のパラメータ θ 付きの確率密度函数 $p(x|\theta)$ の場合にも使うことにする:

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

注意: 条件付き確率分布の密度函数の定義について.

$p(x|y)$ の定義が確率密度の比になっていることがわかりにくいと感じる人は微小な dx, dy をかけて確率とみなせるようにし, $p(x|y)dx$ を

$$p(x|y)\,dx = \frac{p(x,y)\,dx\,dy}{p(y)\,dy}$$

によって定義していると思っても良い.

$$\frac{(\text{図の } dx\,dy \text{ に } (x,y) \text{ が含まれる確率})}{(\text{図の } dy \text{ に } y \text{ が含まれる確率})} = \frac{p(x,y)\,dx\,dy}{p(y)\,dy}$$

$$\| \qquad\qquad \|$$

$$\left(\begin{array}{l} y \text{ が } dy \text{ に含まれるという条件が定める} \\ \text{条件付き確率分布で} \\ x \text{ が } dx \text{ に含まれる確率} \end{array}\right) = p(x|y)\,dx$$

1.11 条件付き期待値の基本性質(連続分布版)

条件付き期待値 $g(y) = E[f(X,y)|y]$ は y の函数とみなせるので, 確率変数 $E[f(X,Y)|Y] = g(Y)$ が得られる. このとき次が成立することを示せ:

$$E[g(Y)] = E[E[f(X,Y)|Y]] = E[f(X,Y)].$$

解答例: 定義通りに計算しよう:

$$\begin{aligned} E[f(X,Y)|Y] = E[g(Y)] &= \int g(y)p(y)\,dy \\ &= \int E[f(X,y)|y]p(y)\,dy \\ &= \int \left(\int f(x,y)p(x|y)\,dx \right) p(y)\,dy \\ &= \iint f(x,y)p(x|y)p(y)\,dx\,dy \\ &= \iint f(x,y)p(x,y)\,dx\,dy = E[f(X,Y)]. \end{aligned}$$

要するに x, y の同時確率密度函数が $p(x,y) = p(x|y)p(y)$ と書けることから, この問題の結論が出て来る.

解答終

注意: 測度論的確率論ではこれを一般化して条件付き期待値を定義する.

1.12 連続分布版のBayesの定理

以下では $\int \cdots dx$ は定積分を意味するとする.

連続変数 (x,y) の確率密度函数 $P(x,y)$ によって連続確率分布が定義されているとする. 以下では, 変数 x, y に対応する確率変数達 X, Y それぞれ単独の確率密度函数を $p(x), p(y)$ と書くことにする:

$$p(x) = \int p(x,y)\,dy \quad p(y) = \int p(x,y)\,dx.$$

確率密度函数 $p(x)$,p(y)のそれぞれが定める x,y$ の分布を **周辺分布** (marginal distribution)と呼ぶことがある.

このとき, y に対応する確率変数 Y によって $Y = y$ という条件で定義される条件付き確率分布の確率密度函数

$$p(x|y) = \frac{p(x,y)}{p(y)}, \quad p(y) = \int p(x,y)\,dx \tag{1}$$

だけではなく, x に対応する確率変数 X によって $X = x$ という条件で定義される条件付き確率分布の確率密度函数

$$p(y|x) = \frac{p(x,y)}{p(x)}, \quad p(x) = \int P(x,y)\,dy \tag{2}$$

が定義される. このとき

$$\begin{aligned} p(x,y) &= p(x|y)p(y) = p(y|x)p(x), \\ p(x) &= \int p(x,y)\,dy = \int p(x|y)p(y)\,dy, \\ p(y) &= \int p(x,y)\,dx = \int p(y|x)p(x)\,dx \end{aligned} \tag{3}$$

なので,

$$\begin{aligned} p(y|x) &= \frac{p(x|y)p(y)}{\int p(x|y)p(y)\,dy}, \\ p(x|y) &= \frac{p(y|x)p(x)}{\int p(y|x)p(x)\,dx}. \end{aligned} \tag{B}$$

この結果(B)を **Bayesの定理** (ベイズの定理)と呼ぶ.

しかし, 以上の議論を見ればわかるように, Bayesの定理は条件付き確率分布の定義を書き直したものに過ぎず, 価値ある新しい考え方が得られるような結果ではない. しかし, 具体的もしくは理論的な計算で条件付き確率の概念が(B)の形式で使われることがあることは知っておいて損がないと思われるので, このように紹介することにした.

記号法上の注意1: 慣習に従って $p(x)$ と $p(y)$ の違いを独立変数名 x, y の違いで区別するという記号法を採用した. これは簡潔に書くための手抜きであり, $p_X(x)$ や $p_Y(y)$ のように書けば独立変数名の違いで区別する必要はなくなる. 以上では簡潔さを重視するスタイルを採用したので混乱しないように注意して欲しい. $p(x|y)$ と $p(y|x)$ の違いも独立変数名 x, y の違いで区別することになった.

記号法上の注意2: $p(y|x) = \dfrac{p(x|y)p(y)}{\int p(x|y)p(y)\,dy}$ における分子の y は独立変数だが, 分母の y は積分変数であり, 異なる意味を持っている. 違うものを同じ記号で書いている. $\int p(x|y)p(y)\,dy$ は $\int p(x|y')p(y')\,dy'$ と書いた方が混乱を防ぎ易かったかもしれない. しかし, そのようにすると, 変数名を y の1個から y, y' の2個に増やす必要があり, 式の煩雑さが増すので, あえて同じ記号で書くことにした. 混乱しないように注意して欲しい.

注意: Bayes統計ではパラメータ θ 付きの確率密度函数 $p(x|\theta)$ の他にパラメータ θ の確率密度函数 $p(\theta)$ も用意して, x と θ の同時確率密度函数を $p(x,\theta) = p(x|\theta)p(\theta)$ と定義する. この状況のもとでは, $X = x$ という条件で定義される θ の条件付き確率分布 $p(\theta|x)$ は次のように表される:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{\int p(x|\theta)p(\theta)\,d\theta}.$$

Bayes統計の文脈では, $p(\theta)$ は **事前分布** (prior)と呼ばれ, $p(x|\theta)$ は **事後分布** (posterior)と呼ばれる. Bayes統計では新たな道具として事前分布 $p(\theta)$ を利用できるお陰で, それ抜きでは困難な統計分析も可能になる. ただし, Bayes統計がそういう優れた性質を持っている理由は「主観確率」や「ベイズ主義」のような概念抜きに語ることができる. この点に関する詳しい説明は次の教科書にある:

- 渡辺澄夫, ベイズ統計の理論と方法, 2012年, コロナ社, 226ページ.

さらに, 次の論説も参考になる:

- 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)

1.13 問題: 連続分布の条件付き確率分布の例 (分散が固定された正規分布モデルのBayes統計)

この節の式には細かい誤りが多数含まれている可能性がある. 誤りを見付けたら自分で訂正すること. 誤りがなかったら, 拍手喝采して欲しい.

$n, \rho > 0$ を任意に取って固定する. $x, \mu \in \mathbb{R}$ の同時確率密度函数

$$p(x,\mu|n,\rho) = \frac{e^{-n(x-\mu)^2/2}}{\sqrt{2\pi/n}}\frac{e^{-\mu^2/(2\rho^2)}}{\sqrt{2\pi\rho^2}}.$$

によって定義される $(x,\mu)\in\mathbb{R}^2$ の確率分布を考える. その分布において, 変数 x に対応する確率変数 X に関する条件 $X=x$ で定義された μ に関する条件付き確率分布の密度函数が

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

になることを示せ. これは平均と分散がそれぞれ

$$\frac{n\rho^2}{1+n\rho^2}x, \qquad \frac{\rho^2}{1+n\rho^2}$$

の正規分布である.

注意: 特に, $n\to\infty$ とすると, 条件 $X=x$ で定義された μ の条件付き確率分布の期待値は x に収束し, 分散は 0 に収束するので, μ の条件付き確率分布は x の近くに集中するようになる.

注意: この問題の例は正規分布モデル $p(x|n,\mu)=e^{-n(x-\mu)^2/2}/\sqrt{2\pi/n}$ と事前分布 $p(\mu|\rho)=e^{-\mu^2/(2\rho^2)}/\sqrt{2\pi\rho^2}$ に関するBayes統計におけるデータの数値 x が定める事後分布 $p(\mu|x,n,\rho)$ を求める計算になっている. x は標本平均に, n はサンプルサイズに対応しており, $p(x|n,\mu)$ の形は分散 1 の正規分布のサイズ n 標本分布のおいて標本平均の分散が $1/n$ になることを表している. だから, この場合には固定された分散 1 を持つ正規分布の標本分布を統計モデルとして採用している場合になっていると考えて欲しい. $n=0$ のとき事後分布が事前分布に一致することにも注意せよ.

注意: 条件付き確率分布の概念はBayes統計に限らず統計学における最も基本的かつ重要な確率論の道具であるが, 特にBayes統計では条件付き確率分布を使った推論が系統的に利用される.

解答例: X 単独の密度函数 $p(x|\rho)$ を求めよう.

$$\begin{aligned} n(x-\mu)^2+\frac{\mu^2}{\rho^2} &= \frac{1+n\rho^2}{\rho^2}\mu^2-2nx\mu+nx^2 \\ &= \frac{1+n\rho^2}{\rho^2}\left(\mu-\frac{n\rho^2 x}{1+n\rho^2}\right)^2+\frac{nx^2}{1+n\rho^2} \end{aligned} \tag{$*$}$$

より,

$$\begin{aligned} p(x|n,\rho) &= \int_{\mathbb{R}} p(x,\mu|n,\rho)\,d\mu \\ &= \frac{\exp(-nx^2/(2(1+n\rho^2)))}{\sqrt{2\pi/n}\,\sqrt{2\pi\rho^2}}\int_{\mathbb{R}}\exp\left(-\frac{1+n\rho^2}{2\rho^2}\left(\mu-\frac{n\rho^2 x}{1+n\rho^2}\right)^2\right)d\mu \\ &= \frac{\exp(-nx^2/(2(1+n\rho^2)))}{\sqrt{2\pi/n}\,\sqrt{2\pi\rho^2}}\sqrt{\frac{2\pi\rho^2}{1+n\rho^2}} \\ &= \frac{1}{\sqrt{2\pi(1+n\rho^2)/n}}\exp\left(-\frac{x^2}{2(1+n\rho^2)/n}\right). \end{aligned}$$

このように X 単独の分布は平均 0, 分散 $(1+\rho^2)/n$ の正規分布になる.

ゆえに, 条件 $X=x$ で定義される μ の分布の確率密度函数 $p(\mu|x,n,\rho)$ は以下のようになる. すなわち,

$$n(x-\mu)^2+\frac{\mu^2}{\rho^2}-\frac{nx^2}{1+n\rho^2} = \frac{1+n\rho^2}{\rho^2}\left(\mu-\frac{n\rho^2 x}{1+n\rho^2}\right)^2$$

を使うと,

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

解答終

注意: 本当は最初に公式($*$)を得た時点ですべての答えが出ていたと考えられる. 以上では丁寧だが無駄な詳細を含む計算を納得感を高めるためにあえてやってみせたが, 実践的な計算においてはそのように考えるべきである. 以下においても同様に無駄な詳細を含む計算をやってみせる場合があるので注意して欲しい. (要するにこのノートの計算の仕方を実践的な計算で真似しないで欲しいということ.)

In [17]:

```
@vars n ρ positive=true
@vars x μ

expr1 = n*(x - μ)^2 + μ^2/ρ^2
expr2 = (1+n*ρ^2)/ρ^2 * (μ - n*ρ^2/(1+n*ρ^2)*x)^2
expr3 = (expr1 - expr2).factor()
[
    expr1
    expr2
    expr3
]
```

Out[17]:

$$\left[\begin{array}{c} n(x-\mu)^2+\frac{\mu^2}{\rho^2} \\ \frac{\left(n\rho^2+1\right)\left(-\frac{nx\rho^2}{n\rho^2+1}+\mu\right)^2}{\rho^2} \\ \frac{nx^2}{n\rho^2+1} \end{array}\right]$$

1.14 分散が固定された正規分布モデルのBayes統計の事後分布の視覚化

$\mu_0 = 4$, $\rho = 1/2$ とおく.

平均 $\mu_0 = 4$, 分散 1 の正規分布のサイズ $N = 2^{10}$ の標本 $X_1, X_2, \ldots, X_N$ をランダムに生成し, $n = 0, 1, 2, 4, 8, \ldots, 1024 = 2^{10}$ について標本平均 $x = \bar{X}_n = (X_1 + \cdots + X_n)/n$ を求め, 前節で求めた μ の事後分布 $p(\mu|x, n, \rho)$ をプロットしてみよう.

n を大きくすると μ の事後分布はサンプルを生成した分布の平均値 μ_0 に集中して行く.

これはBayes統計の漸近論的な基礎付けの最も簡単な場合になっている.

In [18]:

```
Random.seed!(4649373) # よろしくみなさん

μ₀ = 4
ρ = 1/2
N = 2^10
X = rand(Normal(μ₀, 1), N) # 標本生成
PP = []
prior = Normal(0, ρ)
P = plot(prior, μ₀-6, μ₀+4.5; label="prior", xtick=-10:10)
push!(PP, P)
for k in 0:10
    n = 2^k
    x = mean(@view X[1:n])
    μ_posterior = n*ρ^2*x/(1 + n*ρ^2)
    σ_posterior = √(ρ^2/(1+n*ρ^2))
    posterior = Normal(μ_posterior, σ_posterior)
    P = plot(posterior, μ₀-6, μ₀+4.5; label="n=$n", xtick=-10:10)
    push!(PP, P)
end
plot(PP...; size=(800, 500), layout=(4, 3))
```

Out[18]:

1.15 問題: 2変量の正規分布とその条件付き確率分布の例

$$\Sigma = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$$

は固有値がすべて正の実対称行列であるとし, $\mu = (\mu_x, \mu_y) \in \mathbb{R}^2$ であるとする. このとき,

$$\Sigma^{-1} = \frac{1}{ac - b^2} \begin{bmatrix} c & -b \\ -b & a \end{bmatrix}$$

であり, $(x, y) \in \mathbb{R}^2$ の確率密度函数を

$$\begin{aligned} p(x, y) &= \frac{1}{\det(2\pi\Sigma)^{1/2}} \exp\left(-\frac{1}{2} \left[x - \mu_x,\ y - \mu_y \right] \Sigma^{-1} \begin{bmatrix} x - \mu_x \\ y - \mu_y \end{bmatrix} \right) \\ &= \frac{1}{\sqrt{(2\pi)^2 (ac - b^2)}} \exp\left(-\frac{a(y - \mu_y)^2 - 2b(x - \mu_x)(y - \mu_y) + c(x - \mu_x)^2}{2(ac - b^2)} \right) \end{aligned}$$

と定めることができる. これが $\iint_{\mathbb{R}^2} p(x, y)\, dx\, dy = 1$ を満たすことは, Σ を直交行列で対角すれば示せる. (詳細は略す. 以下では認めて使ってよい.) この確率密度函数が定める確率分布を **2変量正規分布** と呼び,

$$\mathrm{MvNormal}(\mu, \Sigma)$$

と表すことにする. μ を平均(もしくは期待値)と呼び, Σ を **分散共分散行列** (variance-covariance matrix)と呼ぶ. (この記号法は, 2変量の場合に限定せずに, 任意の **多変量正規分布** (multivariate normal distribution)にも使用することにする.)

この場合には条件 $X = x$ が定める y に関する条件付き確率分布が, 平均と分散がそれぞれ

$$\mu = (b/a)(x - \mu_x) + \mu_y, \quad \sigma^2 = (ac - b^2)/a$$

の正規分布になることを示せ.

解答例: 上の2変量正規分布において条件 $X = x$ が定める y に関する条件付き確率分布を求めよう.

$$\begin{aligned}
&a(y - \mu_y)^2 - 2b(x - \mu_x)(y - \mu_y) + c(x - \mu_x)^2 \\
&= a\left(y - \mu_y - \frac{b}{a}(x - \mu_x)\right)^2 - \frac{b^2}{a}(x - \mu_x)^2 + c(x - \mu_x)^2 \\
&= a\left(y - \mu_y - \frac{b}{a}(x - \mu_x)\right)^2 + \frac{ac - b^2}{a}(x - \mu_x)^2.
\end{aligned}$$

ゆえに,

$$\begin{aligned}
&\sqrt{(2\pi)^2(ac - b^2)}\ p(x) \\
&= \int_{\mathbb{R}} \exp\left(-\frac{a\big(y - \mu_y - (b/a)(x - \mu_x)\big)^2 + ((ac - b^2)/a)(x - \mu_x)^2}{2(ac - b^2)}\right) dy \\
&= \sqrt{\frac{2\pi(ac - b^2)}{a}}\ \exp\left(-\frac{(x - \mu_x)^2}{2a}\right)
\end{aligned}$$

なので,

$$p(x) = \frac{1}{\sqrt{2\pi a}} \exp\left(-\frac{(x - \mu_x)^2}{2a}\right).$$

これは x の周辺分布が平均 μ_x, 分散 a の正規分布になることを意味している. (実は多変量正規分布の一般論からこれは言えることでもある.) さらに,

$$\begin{aligned}
&a(y - \mu_y)^2 - 2b(x - \mu_x)(y - \mu_y) + c(x - \mu_x)^2 - \frac{ac - b^2}{a}(x - \mu_x)^2 \\
&= a\left(y - \mu_y - \frac{b}{a}(x - \mu_x)\right)^2
\end{aligned}$$

より,

$$p(y|x) = \frac{p(x, y)}{p(x)} = \frac{1}{\sqrt{2\pi(ac - b^2)/a}} \exp\left(-\frac{a\big(y - \mu_y - (b/a)(x - \mu_x)\big)^2}{2(ac - b^2)}\right).$$

これは条件 $X = x$ が定める y の条件付き確率分布が平均 $(b/a)(x - \mu_x) + \mu_y$, 分散が $(ac - b^2)/a$ の正規分布になることを意味している.

注意: 以上の計算を一般の多変量正規分布に一般化すれば **Gauss過程回帰** が得られる.

注意: 最小二乗法による線形回帰との関係. 条件付き確率分布の平均と分散は $\sigma_x^2 = a$, $\sigma_{xy} = b$, $\sigma_y^2 = c$ と書くと,

$$\frac{b}{a}(x - \mu_x) + \mu_y = \frac{\sigma_{xy}}{\sigma_x^2}(x - \mu_x) + \mu_y, \quad \frac{ac - b^2}{a} = \frac{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2}{\sigma_x^2}$$

と表される. これは「標本分布について」のノートで導出したデータ $(x_1, y_1), \ldots, (x_n, y_n)$ に関する最小二乗法による線形回帰の公式

$$\hat{\alpha} + \hat{\beta}x = \frac{s_{xy}}{s_x^2}(x - \bar{x}) + \bar{y}, \quad \hat{\sigma}^2 = \frac{n - 1}{n}\frac{s_x^2 s_y^2 - s_{xy}^2}{s_x^2}$$

に非常に似ている. ここで

$$
\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \\
s_x^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2, \\
s_{xy} &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).
\end{aligned}
$$

さらに,

$$
\begin{aligned}
\hat{\mu}_x &= \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \hat{\mu}_y = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \\
\hat{\sigma}_x^2 &= \frac{n-1}{n} s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad \hat{\sigma}_y^2 = \frac{n-1}{n} s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2, \\
\hat{\sigma}_{xy} &= \frac{n-1}{n} s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).
\end{aligned}
$$

とおくと,

$$
\hat{\alpha} + \hat{\beta}x = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x^2}(x - \hat{\mu}_x) + \hat{\mu}_y, \quad \hat{\sigma}^2 = \frac{\hat{\sigma}_x^2\hat{\sigma}_y^2 - \hat{\sigma}_{xy}^2}{\hat{\sigma}_x^2}
$$

となり, 式の類似性はさらに増す.

この類似は偶然ではない. 最小二乗法による線形回帰は, データ $(x_1, y_1), \ldots, (x_n, y_n)$ に最もフィットする2変量正規分布を最尤法で構成し($\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}_x^2, \hat{\sigma}_y^2, \hat{\sigma}_{xy}$ がその最尤法の解になっている), 条件 $X = x$ で定められた y に関する条件付き確率分布を求めることと数学的に同じことになっている.

1.16 2変量の正規分布とその条件付き確率分布の例の視覚化

以下では前節の内容を2変量正規分布の乱数を大量に生成することによって数値的に確認してみよう.

In [19]:
```
# 二変量正規分布
μ = [10, 20]
Σ = [
    6 2
    2 3
]
mvnormal = MvNormal(μ, Σ)
```

Out[19]:
```
FullNormal(
dim: 2
μ: [10.0, 20.0]
Σ: [6.0 2.0; 2.0 3.0]
)
```

In [20]:
```
# X, Y 単独の分布
xdist = Normal(μ[1], √Σ[1,1])
ydist = Normal(μ[2], √Σ[2,2])
@show xdist ydist;
```

```
xdist = Normal{Float64}(μ=10.0, σ=2.449489742783178)
ydist = Normal{Float64}(μ=20.0, σ=1.7320508075688772)
```

In [21]:
```
# 条件X=xが定めるyの条件付き確率分布の平均値 (≈ 回帰直線 (regression line))
f(x, μ, Σ) = (Σ[1,2]/Σ[1,1])*(x - μ[1]) + μ[2]
```

Out[21]:
```
f (generic function with 1 method)
```

In [22]:

```
# テストデータの生成: 二変量正規分布の分布は楕円状になる.
n = 10^4
XY = rand(mvnormal, n)
X, Y = XY[1,:], XY[2,:]
scatter(X, Y; ma=0.3, ms=2, msw=0, label="", title="data")
```

Out[22]:

In [23]:

```
# データでのX単独の分布
stephist(X; norm=true, label="data X")
plot!(xdist; label="true", ls=:dash)
```

Out[23]:

In [24]:

```
# データでのY単独の分布
stephist(Y; norm=true, label="data Y")
plot!(ydist; label="true", ls=:dash)
```

Out[24]:

In [25]:
```
# μの最尤推定
μ̂ = mean(XY, dims=2)
```

Out[25]:
```
2×1 Matrix{Float64}:
  9.986931890736575
 19.987253275638245
```

In [26]:
```
# Σの最尤推定
Σ̂ = cov(XY'; corrected=false)
```

Out[26]:
```
2×2 Matrix{Float64}:
 6.04319  1.9508
 1.9508   2.95852
```

In [27]:
```
# 条件X=xが定めるyの条件付き確率分布の期待値と回帰直線の比較
xlim = (μ[1] - 4√Σ[1,1], μ[1] + 4√Σ[1,1])
plot(x -> f(x, μ̂, Σ̂), xlim...; label="linear regression", c=:red)
plot!(x -> f(x, μ, Σ), xlim...; label="true", ls=:dash, c=:blue)
plot!(; legend=:topleft)
```

Out[27]:

In [28]:
```
# サンプルとまとめてプロット
scatter(X, Y; ma=0.3, ms=2, msw=0, label="", title="data")
plot!(x -> f(x, μ̂, Σ̂); label="linear regression", c=:red)
plot!(x -> f(x, μ, Σ); label="true", ls=:dash, c=:blue)
plot!(; legend=:topleft)
```

Out[28]:

In [29]:
```
# 条件X=xが定めるyの条件付き確率分布の分散とその推定値の比較
σ² = det(Σ)/Σ[1,1]
σ̂² = det(Σ̂)/Σ̂[1,1]
@show σ² σ̂²;
```

```
σ² = 2.3333333333333335
σ̂² = 2.328779918715761
```

以下では, x, y の同時確率密度函数 $p(x, y)$ (これは平均 μ と分散共分散行列 Σ を持つ二変量正規分布の密度函数)と x で条件付けられた y に関する条件付き確率分布の密度函数 $p(y|x) = p(x, y)/p(x)$ のヒートマップ(値が大きなところほど明るくプロットされる)を比較してみよう.

In [30]:

```
f(x, y) = pdf(mvnormal, SVector(x, y)) # p(x,y)
f(x) = pdf(xdist, x) # p(x)
g(x, y) = f(x, y)/f(x)

x = range(0, 20, 300)
y = range(10, 30, 300)

P1 = heatmap(x, y, f; colorbar=false, title="p(x,y)")
P2 = heatmap(x, y, g; colorbar=false, title="p(y|x)")
plot(P1, P2; size=(600, 250))
```

Out[30]:

$p(x, y)$ は xy 平面上の確率密度函数になっている.

$p(y|x)$ は各 x ごとに y 軸に平行な直線上の確率密度函数になっている.

その様子の違いの典型例が上のグラフからわかる.

$p(y|x)$ のヒートマップを見れば, パラメータ x を持つ y の確率密度函数は x を y に対応させる函数の一般化(x に対応する y の値が確率的に揺らぐ)になっていることもわかる.

2 2×2の分割表の分布

実践的によく使われている2×2の分割表のモデルになる確率分布を紹介しよう.

2×2の分割表とは数学的には0以上の整数を成分とする2×2の行列

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad a_{ij} \in \mathbb{Z}_{\geq 0}$$

のことである. 以下では2×2の分割表の集合に順次制限して狭くして行きながら, 4種類の確率分布を紹介する.

それら4種の確率分布達は条件付け確率分布の階層をなしている.

2.1 4つのPoisson分布の積

$\lambda_{ij} \geq 0$ であるとし,

$$\lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}$$

とおく. 制限のない分割表 $A = [a_{ij}]$ 達の確率質量函数

$$P(A|\lambda) = \prod_{i,j=1}^{2} \left(e^{-\lambda_{ij}} \frac{\lambda_{ij}^{a_{ij}}}{a_{ij}!} \right) \qquad (a_{ij} \in \mathbb{Z}_{\geq 0})$$

で定まる分割表の確率分布を **4つのPoisson分布の積** と呼ぶことにする.

分割表の4つのPoisson分布の積モデルはある一定期間のあいだ収集したサンプルを2つの条件 X, Y を使って次のように分類することによって得られた2×2の分割表のモデルになっていると考えられる:

- a_{11} はサンプル中の X でかつ Y であるものの個数で, a_{12} は X でかつ Y でないものの個数で, a_{21} は X でなくて Y であるものの個数で, a_{22} は X でも Y でもないものの個数であるとする.

以上の設定のもとで4つのPoisson分布の期待値は

$$\lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}$$

に一致する.

2.2 四項分布

$\theta_{ij} \geq 0$, $\sum_{i,j=1}^{2} \theta_{ij} = 1$ であるとし,

$$\theta = \begin{bmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{bmatrix}$$

とおく. $n \in \mathbb{Z}_{\geq 0}$ を任意に取って固定し, 成分の総和が n になるという条件

$$\sum_{i,j=1}^{2} a_{ij} = n$$

を満たす分割表 $A = [a_{ij}]$ に関する確率分布が次の確率質量函数によって定められる:

$$P(A|n,\theta) = n! \prod_{i,j=1}^{4} \frac{\theta_{ij}^{a_{ij}}}{a_{ij}!} \qquad (a_{11} + a_{12} + a_{21} + a_{22} = n)$$

これは **四項分布** である.

分割表の四項分布モデルは, 最初に決めておいたサイズ n のサンプルを収集して, 2つの条件 X, Y を使って前節を同じようにサンプルを分類することによって得られた分割表のモデル化になっていると考えられる.

この四項分布の期待値は

$$n\theta = \begin{bmatrix} n\theta_{11} & n\theta_{12} \\ n\theta_{21} & n\theta_{22} \end{bmatrix}$$

になる.

2.3 2つの二項分布の積

$0 \leq \eta_i \leq 1$ $(i = 1, 2)$であるとし, $\eta = (\eta_1, \eta_2)$ とおく. $r_1, r_2 \in \mathbb{Z}_{\geq 0}$ を任意に取って固定し, $r = (r_1, r_2)$ とおく. このとき, 横方向の合計(横方向のマージン)が

$$\sum_{j=1}^{n} a_{ij} = r_i \quad (i = 1, 2)$$

と固定されているような分割表の確率分布を次の確率質量函数によって定めることができる:

$$P(A|r,\eta) = \binom{r_1}{a_{11}} \eta_1^{a_{11}} (1-\eta_2)^{a_{12}} \cdot \binom{r_2}{a_{21}} \eta_1^{a_{21}} (1-\eta_2)^{a_{22}} \quad (a_{i1} + a_{i2} = r_i)$$

これを **2つの二項分布の積** と呼ぶ.

分割表の2つの二項分布の積モデルは, 前もって決めておいたサイズ r_1, r_2 の条件 X を満たす群と満たさない群のサンプルを取得して, 条件 Y によって分類することによって得られた分割表のモデル化になっていると考えられる.

この設定のもとで, 2つの二項分布の積の期待値は

$$r_1\eta_1 r_1(1-\eta_1)$$
$$r_2\eta_2 r_2(1-\eta_2)$$

になる.

2.4 Fisherの非心超幾何分布

$\omega > 0$ であるとする. $r_1, r_2, s_1, s_2, n \in \mathbb{Z}_{\geq 0}$ で

$$n = r_1 + r_2 = s_1 + s_2$$

という条件を満たすものを任意に取って固定し, $r = (r_1, r_2)$, $s = (s_1, s_2)$ とおく. このとき, 横方向の合計と縦方向の合計の両方(全マージン)が

$$\sum_{j=1}^{2} a_{ij} = r_i \quad (i = 1, 2), \qquad \sum_{i=1}^{2} a_{ij} = s_j \quad (j = 1, 2)$$

と固定されているような分割表の確率分布を次の形の確率質量函数によって定めることができる:

$$\begin{aligned}
&P(A|r, s, \omega) \\
&= Z(r, s, \omega)^{-1} \frac{\omega^{a_{11}}}{\prod_{i,j=1}^{2} a_{ij}!} = \frac{\omega^{a_{11}}}{\widetilde{Z}(r, s, \omega)} \binom{n}{r_1}^{-1} \binom{s_1}{a_{11}} \binom{s_2}{a_{12}} \\
&= \frac{\omega^{a_{11}}}{\widetilde{Z}(r, s, \omega)} \binom{r_1}{a_{11}} \frac{s_1(s_1 - 1) \cdots (s_1 - a_{11} + 1)\, s_2(s_2 - 1) \cdots (s_2 - (r_1 - a_{11}) + 1)}{n(n-1) \cdots (n - r_1 + 1)}. \\
&(a_{i1} + a_{i2} = r_i,\ a_{1j} + a_{2j} = s_j)
\end{aligned}$$

これを **Fisherの非心超幾何分布** (Fisher's noncentral hypergeometric distribution)と呼ぶ.

パラメータの函数達 $Z(r, s, \omega)$, $\widetilde{Z}(r, s, \omega)$ は確率の総和が 1 になるという条件で決める. 一般にそれらの定数の具体形はシンプルに書けないが, $\omega = 1$ の場合は例外的に,

$$\widetilde{Z}(r, s, 1) = 1, \qquad Z(r, s, 1) = \frac{n!}{r_1! r_2! s_1! s_2!}$$

とシンプルな形になり,

$$P(A|r, s, 1) = \binom{n}{r_1}^{-1} \binom{s_1}{a_{11}} \binom{s_2}{a_{12}} = \frac{r_1! r_2! s_1! s_2!}{n! a_{11}! a_{12}! a_{21}! a_{22}!}$$

となる. この $\omega = 1$ の場合のFisherの非心超幾何分布は単に **超幾何分布** と呼ばれている.

注意: 超幾何分布の期待値は

$$\begin{bmatrix} r_1 s_1/n & r_1 s_2/n \\ r_2 s_1/n & r_2 s_2/n \end{bmatrix}$$

になる. Fisherの非心超幾何分布の期待値についてはこのようなシンプルな公式はない.

注意: 超幾何分布の確率質量函数の上の表示は以下のような解釈を持つ.

袋の中に s_1 個の赤い玉と s_2 個の白い玉が入っているとし, その中から r_1 個の玉を一度にまとめて無作為に取り出す. 取り出した r_1 個の玉の中の赤い玉の個数を a_{11} とし, 白い玉の個数を $a_{12} = r_1 - a_{11}$ と書き, 袋の中に残された r_2 個の玉の中の赤い玉の個数を $a_{21} = s_1 - a11$ と書き, 白い玉の個数を $a_{22} = s_2 - a_{12} = s_2 - r_1 + a_{11}$ と書き, 2×2の分割表 $A = [a_{ij}]$ を作る. a_{11} だけから残りの $a_{12}, a_{21}. a_{22}$ の値が決定される.

このとき, A の分布は前節の超幾何分布に従う. なぜならば, 袋の中の $n = s_1 + s_2$ 個の玉の中から r_1 個を一度に取り出す組み合わせ全体の個数は $\binom{n}{r_1}$ であり(分母), 袋の中の s_1 個の赤い玉の中から a_{11} 個選ぶ組み合わせ全体の個数と s_2 個の白い玉の中から a_{12} 個選ぶ組み合わせ全体の個数の積は $\binom{s_1}{a_{11}}\binom{s_2}{a_{12}}$ になる(分子)からである.になる.

注意: 超幾何分布と非復元抽出の関係. 超幾何分布の確率質量函数は次のようにも表される:

$$P(A|r, s, 1) = \binom{r_1}{a_{11}} \frac{s_1(s_1 - 1) \cdots (s_1 - a_{11} + 1)\, s_2(s_2 - 1) \cdots (s_2 - (r_1 - a_{11}) + 1)}{n(n-1) \cdots (n - r_1 + 1)}.$$

この公式は以下のような解釈を持つ. 袋の中に s_1 個の赤い玉と s_2 個の白い玉が入っているとし, $s_1 + s_2 = n$ であるとする. 袋の中から無作為に1個ずつ玉を取り出す. ただし, 取り出した玉は袋に戻さないとする(非復元抽出). そのように玉を r_1 個取り出したとき, その中に赤い玉が a_{11} 個含まれる確率が超幾何分布における確率の値 $P(A|r, s, 1)$ になる.

注意: 以上の状況のもとで a_{11} が動ける範囲はちょうど次になる:

$$\max(0, r_1 - s_2) \le a_{11} \le \min(s_1, r_1).$$

この条件は以下の条件を整理することによって得られる:

$$
\begin{aligned}
a_{11} &\geq 0, \\
a_{11} &= r_1 - a_{12} \leq r_1, \\
a_{11} &= s_1 - a_{21} \leq s_1, \\
a_{11} &= r_1 - s_2 + a_{22} \geq r_1 - s_2.
\end{aligned}
$$

これらの等式部分は以下から得られる:

$$
\begin{aligned}
a_{12} &= r_1 - a_{11}, \\
a_{21} &= s_1 - a_{11}, \\
a_{22} &= s_2 - a_{12} = s_2 - r_1 + a_{11}.
\end{aligned}
$$

注意: 2×2の分割表のFisherの非心超幾何分布およびその特別な場合である超幾何分布においては a_{11} の値を決めれば他の a_{12}, a_{21}, a_{22} の値が決まるので, それらは

$$\max(0, r_1 - s_2) \leq a_{11} \leq \min(s_1, r_1)$$

を満たす整数 a_{11} のみに関する確率分布だとみなされる. a_{11} のみに関するFisherの非心超幾何分布を

$$\mathrm{FisherNoncentralHypergeometric}(s_1, s_2, r_1, \omega)$$

と書き, 超幾何分布を

$$\mathrm{Hypergeometric}(s_1, s_2, r_1)$$

と書くことにする. $\mathrm{Hypergeometric}(s_1, s_2, r_1)$ は s_1 個の赤い玉と s_2 個の白い玉が入っている袋から取り出した r_1 個の玉の中の赤い玉の個数 a_{11} の分布になっている.

袋の中の赤い玉の個数 s_1 が取り出す玉の個数 r_1 より小さい場合には取り出した r_1 個の玉がすべて赤い玉になること ($a_{11} = s_1$) はありえず, 袋の中の白い玉の個数 s_2 が取り出す玉の個数 r_1 より小さい場合には取り出した r_1 個の玉がすべて白い玉になること($a_{11} = 0$)となることはありえない. このことに注意すれば a_{11} の動く範囲についても理解し易いだろう.

2.5 Fisherの非心超幾何分布の正規分布近似の視覚化

Fisherの非心超幾何分布 $\mathrm{FisherNoncentralHypergeometric}(s, f, n)$ も s, f, n が十分大きければ正規分布で近似される.

In [31]:
```
function plot_fnch(s₁, s₂, r₁, ω)
    fnch = FisherNoncentralHypergeometric(s₁, s₂, r₁, ω)
    μ, σ = mean(fnch), std(fnch)
    xlim = (max(-1.0, μ-4σ), min(maximum(fnch)+1.0, μ+4σ))
    x = range(xlim..., 1000)
    plot(x -> mypdf(fnch, x), xlim...; label="")
    plot!(Normal(μ, σ); label="", ls=:dash)
    title!("s₁=$s₁, s₂=$s₂, r₁=$r₁, ω=$ω")
end
```

Out[31]: plot_fnch (generic function with 1 method)

In [32]:
```
plot(plot_fnch(6, 14, 10, 0.5), plot_fnch(6, 14, 10, 2.0); size=(800, 250))
```

Out[32]:

In [33]:
```
plot(plot_fnch(12, 28, 20, 0.5), plot_fnch(12, 28, 20, 2.0); size=(800, 250))
```

Out[33]:

In [34]:
```
plot(plot_fnch(24, 56, 40, 0.5), plot_fnch(24, 56, 40, 2.0); size=(800, 250))
```

Out[34]:

In [35]:
```
plot(plot_fnch(48, 112, 80, 0.5), plot_fnch(48, 112, 80, 2.0); size=(800, 250))
```

Out[35]:

2.6 四項分布は4つのPoisson分布の積の条件付き確率分布

オッズ比 ω を次のように定める:

$$\omega = \frac{\lambda_{11}/\lambda_{12}}{\lambda_{21}/\lambda_{22}} = \frac{\lambda_{11}\lambda_{22}}{\lambda_{12}\lambda_{21}}.$$

4つのPoisson分布の積の確率質量函数を $\sum_{i,j=1} a_{ij} = n$ が固定された場合に制限してみよう.

$$\theta_{ij} = \frac{\lambda_{ij}}{\sum_{i',j'=1}^{n} \lambda_{i'j'}}$$

とおくと,

$$\omega = \frac{\theta_{11}/\theta_{12}}{\theta_{21}/\theta_{22}} = \frac{\theta_{11}\theta_{22}}{\theta_{12}\theta_{21}}$$

でかつ,

$$P(A|\lambda) = \prod_{i,j=1}^{2}\left(e^{-\lambda_{ij}}\frac{\lambda_{ij}^{a_{ij}}}{a_{ij}!}\right) = \underbrace{\frac{1}{n!}e^{-\sum_{i,j=1}^{2}\lambda_{ij}}\left(\sum_{i,j=1}^{2}\lambda_{ij}\right)^{n}}_{\text{independent of } a_{ij}} n!\prod_{i,j=1}^{2}\frac{\theta_{ij}^{a_{ij}}}{a_{ij}!}.$$

ゆえに, 四項分布における確率の総和が1であること

$$\sum_{\sum_{i,j=1}^{2}a_{ij}=n} n!\prod_{i,j=1}^{2}\frac{\theta_{ij}^{a_{ij}}}{a_{ij}!} = 1$$

より,

$$P(n|\lambda) = \sum_{\sum_{i,j=1}^{2}a_{ij}=n} P(A|\lambda) = \frac{1}{n!}e^{-\sum_{i,j=1}^{2}\lambda_{ij}}\left(\sum_{i,j=1}^{2}\lambda_{ij}\right)^{n}.$$

ゆえに, $\sum_{i,j=1}^{2} a_{ij} = n$ という条件で定義される条件付き確率分布の確率質量函数は

$$\frac{P(A|\lambda)}{P(n|\lambda)} = n!\prod_{i,j=1}^{2}\frac{\theta_{ij}^{a_{ij}}}{a_{ij}!}$$

と, 四項分布の確率質量函数になる.

2.7 2つの二項分布の積は四項分布は4つの条件付き確率分布

四項分布の確率質量函数を $\sum_{j=1}^{2} a_{ij} = r_i, r_1 + r_2 = n$ の場合に制限してみよう.

$$\eta_i = \frac{\theta_{i1}}{\theta_{i1}+\theta_{i2}}$$

とおくと,

$$\omega = \frac{\eta_1/(1-\eta_1)}{\eta_2/(1-\eta_2)} = \frac{\eta_1(1-\eta_2)}{(1-\eta_1)\eta_2}$$

でかつ,

$$P(A|n,\theta) = n!\prod_{i,j=1}^{4}\frac{\theta_{ij}^{a_{ij}}}{a_{ij}!} = \underbrace{n!\prod_{i=1}^{2}(\theta_{i1}+\theta_{i2})^{r_i}}_{\text{independent of } a_{ij}} \cdot \frac{\eta_1^{a_{11}}(1-\eta_1)^{r_1-a_{11}}}{a_{11}!(r_1-a_{11})!}\frac{\eta_2^{a_{21}}(1-\eta_2)^{r_1-a_{21}}}{a_{21}!(r_2-a_{21})!}$$

$$= \underbrace{\binom{n}{r_1}\prod_{i=1}^{2}(\theta_{i1}+\theta_{i2})^{r_i}}_{\text{independent of } a_{ij}} \cdot \binom{r_1}{a_{11}}\eta_1^{a_{11}}(1-\eta_1)^{r_1-a_{11}} \cdot \binom{r_2}{a_{21}}\eta_2^{a_{21}}(1-\eta_2)^{r_1-a_{21}}$$

これを $a_{11} + a_{12} = r_1, a_{21} + a_{22} = r_2$ を満たす分割表について和を取ると, 2つの二項分布の積の部分の和が 1 になる. ゆえに

$$P(r|n,\theta) = \binom{n}{r_1}\prod_{i=1}^{2}(\theta_{i1}+\theta_{i2})^{r_i} = \binom{n}{r_1}(\theta_{11}+\theta_{12})^{r_1}(1-(\theta_{11}+\theta_{12}))^{n-r_1}.$$

したがって, 条件 $a_{11} + a_{12} = r_1, a_{21} + a_{22} = r_2$ が定める条件付き確率分布の確率質量函数は次の形になる:

$$\frac{P(A|n,\theta)}{P(r|n,\theta)} = \binom{r_1}{a_{11}}\eta_1^{a_{11}}(1-\eta_1)^{r_1-a_{11}} \cdot \binom{r_2}{a_{21}}\eta_2^{a_{21}}(1-\eta_2)^{r_1-a_{21}}.$$

これは2つの二項分布の積の確率質量函数である.

2.8 Fisherの非心超幾何分布は2つの二項分布の積の条件付き確率分布

2つの二項分布の積の確率質量函数をさらに $\sum_{i=1}^{2} a_{ij} = s_j, s_1 + s_2 = n$ の場合に制限してみよう. そのとき,

$$a_{21} = s_1 - a_{11}, \quad a_{22} = s_2 - r_1 + a_{11}, \quad \omega = \frac{\eta_1(1-\eta_2)}{(1-\eta_1)\eta_2}$$

であることを使うと,

$$\begin{aligned}
P(A|r,\theta) &= \binom{r_1}{a_{11}}\eta_1^{a_{11}}(1-\eta_1)^{r_1-a_{11}} \cdot \binom{r_2}{a_{21}}\eta_2^{a_{21}}(1-\eta_2)^{r_2-a_{21}} \\
&= \frac{r_1!}{a_{11}!a_{12}!}\eta_1^{a_{11}}(1-\eta_1)^{r_1-a_{11}} \cdot \frac{r_2!}{a_{21}!a_{22}!}\eta_2^{s_1-a_{11}}(1-\eta_2)^{s_2-r_1+a_{11}} \\
&= \underbrace{r_1!r_2!(1-\eta_1)^{r_1}\eta_2^{s_1}(1-\eta_2)^{s_2-r_1}}_{\text{independent of } a_{ij}} \frac{\omega^{a_{11}}}{a_{11}!a_{12}!a_{21}!a_{22}!}.
\end{aligned}$$

これより, 2つの二項分布の積における条件 $\sum_{i=1}^{2} a_{ij} = s_j$ に関する条件付き確率分布がFisherの非心超幾何分布になることがわかる.

2.9 独立性条件

以上における条件付き確率分布の構成でオッズ比 ω が保たれて行くことに注意せよ. 以上で扱った4種類の確率分布は $\omega = 1$ を満たすとき **独立である**, もしくは, **独立性条件を満たす** という.

2つの二項分布の積においては独立であることと $\eta_1 = \eta_2$ であることは同値である. η_i は **比率** と呼ばれることがあり, $\eta_1 = \eta_2$ は比率が等しいという条件になる.

「比率が等しい」という仮説の検定で $\omega = 1$ の統計モデルが使用される.

たとえば, タバコを吸っているかどうかの条件を X と書き, 肺がんになるかどうかの条件を Y と書くとき, その調査結果から作った分割表

	$Y=1$	$Y=0$	
$X=1$	a人	b人	r_1人
$X=0$	c人	d人	r_2人
	s_1人	s_2人	n人

に関する独立性は条件 X と条件 Y が無関係であることを意味する.

2.10 Pearsonのχ²統計量

以上で紹介した2×2の分割表

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad a_{ij} \in \mathbb{Z}_{\geq 0}$$

の4つの確率分布について, 独立性条件が成立しているならば, 次の式で定義される **Pearsonのχ²統計量**

$$X^2 = \frac{(a+b+c+d)(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

は近似的に自由度 1 のχ²分布に従うことが知られている. 以下でそれを視覚化によって確認しておこう.

In [36]:

```julia
function plot_pearson_chisq(randfunc, A, B, C, D; L=10^5)
    # 仮の期待値 A, B, C, D から独立性条件を満たす期待値を作成
    N = A + B + C + D
    Ā = (A+B)*(A+C)/N
    B̄ = (A+B)*(B+D)/N
    C̄ = (C+D)*(A+C)/N
    D̄ = (C+D)*(B+D)/N
    # Pearsonのχ² 統計量を配列 X² に格納
    X² = Vector{Float64}(undef, L)
    Threads.@threads for i in 1:L
        a, b, c, d = randfunc(Ā, B̄, C̄, D̄)
        X²[i] = safediv((a+b+c+d)*(a*d-b*c)^2, (a+b)*(c+d)*(a+c)*(b+d))
    end
    ecdf_chisq = ecdf(X²) # Pearsonのχ² 統計量の経験累積分布函数
    plot(; legend=:bottomright)
    plot!(x -> ecdf_chisq(x), -0.2, 6.5; label="ecdf of Pearson's χ²")
    plot!(x -> cdf(Chisq(1), x); label="cdf of Chisq(1)", ls=:dash)
    title!("$randfunc($Ā, $B̄, $C̄, $D̄)")
end

function rand_4poisson(A, B, C, D)
    rand(Poisson(A)), rand(Poisson(B)), rand(Poisson(C)), rand(Poisson(D))
end

function rand_quadnomial(A, B, C, D)
    N = Int(A + B + C + D)
    rand(Multinomial(N, [A/N, B/N, C/N, D/N]))
end

function rand_2binomial(A, B, C, D)
    R₁, R₂ = Int.((A+B, C+D))
    a, c = rand(Binomial(R₁, A/(A+B))), rand(Binomial(R₂, C/(C+D)))
    b, d = R₁-a, R₂-c
    a, b, c, d
end

function rand_hypergeometric(A, B, C, D)
    A, B, C, D = Int.((A, B, C, D))
    a = rand(Hypergeometric(A+C, B+D, A+B))
    b = A+B-a
    c = A+C-a
    d = D-A+a
    a, b, c, d
end
```

Out[36]: rand_hypergeometric (generic function with 1 method)

以下では期待値が独立性条件を満たす

$$\begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix}$$

の場合のPearsonのχ²統計量の分布の累積分布函数をプロットし, 自由度 1 のχ²分布と比較している.

In [37]:

```julia
plot_pearson_chisq(rand_4poisson, 4, 6, 6, 9)
```

Out[37]:

In [38]:
```
plot_pearson_chisq(rand_quadnomial, 4, 6, 6, 9)
```

Out[38]:

In [39]:
```
plot_pearson_chisq(rand_2binomial, 4, 6, 6, 9)
```

Out[39]:

In [40]:
```
plot_pearson_chisq(rand_hypergeometric, 4, 6, 6, 9)
```

Out[40]:

最後の超幾何分布の場合にのみ自由度 1 のχ²分布による近似が大幅に悪化することに注意せよ. それ以外の場合にはχ²分布による近似の精度はかなりよいように見える.

3 尤度 (ゆうど)と推定

3.1 尤度に関するまとめ

「尤度」は「ゆうど」と読むが, 正しい読み方を知っても統計学的実践に関する有用な知見は得られない.

尤度の定義は **統計モデル内でデータと同じ数値が生成される確率もしくは確率の密度** である.

尤度は統計モデルの「もっともらしさ」(尤もらしさ)の指標では **ない**. (尤度の英語名は「もっともらしさ」を意味する likelihood なのでこの点には注意を要する.)

尤度は「証拠の強さ」の指標でも **ない**. (尤度は証拠の強さの指標として不適切であるにもかかわらず, 歴史的な事情で evidence (証拠)という別名が付いているので, この点にも注意が必要である.)

尤度は **統計モデルのデータへの数値への適合度**(フィッティングの良さ)の指標の1つ(他にも沢山ある)として基本的かつ有用である.

尤度をそのような道具として使用することは, 統計モデルのデータの数値への適合度(フィッティングの良さ)をモデル内でデータと同じ数値がどれだけ生成され易いかで測ることを意味している.

尤度が高い(＝モデルがデータの数値によくフィットしている)にも関わらず役に立たない統計モデルを作ってしまうことはよくある. そうなってしまうことを **オーバーフィッティング** (overfitting, 過剰適合)という.

3.2 尤度函数と尤度の定義と諸注意

パラメータ $\theta = (\theta_1, \dots, \theta_d)$ を持つ確率質量函数 $P(x|\theta)$ または確率密度函数 $p(x|\theta)$ が現実から取得するデータ(数値列) $x = (x_1, \dots, x_n)$ の生成法則のモデル化になっているとき, それらを **統計モデル** と呼ぶ.

データの数値 x が得られたとき, パラメータ θ の函数 $\theta \mapsto P(x|\theta)$ または $\theta \mapsto p(x|\theta)$ をデータ x に関する統計モデルの **尤度函数** (ゆうどかんすう, likelihood function)と呼ぶ.

さらにパラメータの値 θ が与えられたとき, そこでの尤度函数の値 $P(x|\theta)$ または $p(x|\theta)$ を **尤度** (ゆうど, likelihood)と呼ぶ.

すなわち, **統計モデル内においてデータ x と同じ数値が生成される確率 $P(x|\theta)$ または確率の密度 $p(x|\theta)$ を尤度(ゆうど)と呼ぶ**.

尤度は **統計モデルのデータの数値への適合度の指標の1つ**(フィッティングの良さの指標の1つ, 他にも沢山ある)として使われる.

データの数値 x が与えられたとき, 統計モデル $p(x|\theta)$ をデータの数値 x に最も適合するようなパラメータ θ の値は, 尤度を適合度の指標として採用したとき, 尤度函数 $\theta \mapsto p(x|\theta)$ を最大化する $\theta = \hat{\theta}$ として得られる.

この方法を **最尤法** (maximum likelihood method)と呼び, **最尤法の解** $\hat{\theta}$ を **最尤推定値** (maximum likelihood estimate)と呼び, $\theta = \hat{\theta}$ での尤度函数の値 $p(x|\hat{\theta})$ をデータの数値 x に関するモデルの **最大尤度** と呼ぶ.

実際の計算や理論的な取り扱いでは尤度函数そのものではなく, その対数である **対数尤度函数** (log-likelihood function) $\theta \mapsto \log p(x|\theta)$ を使った方が便利なことが多い. 実際にはその -1 倍の **負の対数尤度函数** (negative log-likelihood function) $\theta \mapsto -\log p(x|\theta)$ が使われることが多い. 最尤法の解は負の対数尤度函数を最小化することによって得られる. (χ^2分布との関係で -1 倍ではなく, -2 倍することも多い.)

注意: 尤度(ゆうど)は英語では「もっともらしさ」(尤もらしさ)を意味する likelihood という単語で表されるが, 統計学における専門用語としての尤度は「もっともらしさ」と呼ぶには適さない量になっているので注意が必要である. (非常に残念なことに, 多くの教科書で「尤度はもっともらしさを意味する」のような誤誘導的な解説が書いてある.)

注意: 尤度は「もっともらしさ」の指標としては不適切だが, 統計モデルのデータへの適合度(フィットの度合い)の指標の1つとして基本的かつ有用である.

注意: 尤度函数 $\theta \mapsto P(x|\theta)$ または $\theta \mapsto p(x|\theta)$ は一般に θ の確率密度函数にはなっていない.

注意: 「尤度函数は一般には確率密度函数ではない」を「尤度は確率ではない」と要約すると誤りになるので注意が必要である. データ x に関するパラメータ θ の離散分布モデルにおける尤度 $P(x|\theta)$ は確率の値になっている.

注意: 繰り返しになるが, **尤度はモデルのデータへの適合度の指標の1つ**(他にも沢山ある)として基本的かつ有用である.

注意: 尤度をデータへのモデルの適合度の指標として使うことは, モデルのデータへの適合度をモデル内でどれだけデータと同じ数値が生成され易いかでモデルのデータへの適合度を測るということになる.

注意: データの数値を得てから, 後付けで統計モデルを作れば, いくらでも尤度の高い統計モデルを作れる. さらに, データの数値を見てから後付けで統計モデルを作らなくても, 事前に統計モデルに大量のパラメータを入れておけば, 後付けで採用しそうなモデルもその中に含まれているようにでき, 事前に決めておいた方法でパラメータをデータの数値に合わせて調節することによって, 統計モデルの尤度を高めることも可能である. 何らかの理由で尤度が高い(すなわち既知のデータへの適合度が高い)のにこれから得られるデータの予測に役に立たないモデルができてしまうことを **オーバーフィッティング** (overfitting, 過剰適合)と呼ぶ. 過剰適合の問題があるので, 尤度を単純に「もっともらしさ」の指標として安易に採用することは危険である.

3.3 例: Bernoulli試行の尤度函数と最尤法

当たりが出る確率 p が未知のルーレットを n 回まわして得た当たりとはずれの長さ n 列 $x_1, x_2, \dots, x_n$ がデータとして得られる状況を考える. 当たりは 1 ではずれは 0 で表しておくことにする. 例えば, ルーレットを $n = 5$ 回まわして, 当たり, はずれ, はずれ, 当たり, はずれが順に出たとき, $(x_1, x_2, x_3, x_4, x_5) = (1, 0, 0, 1, 0)$ となるとする.

その状況をBernoulli試行でモデル化する. すなわち, 次の確率質量函数を統計モデルとして採用する:

$$P(x_1, \dots, x_n|p) = p^{x_1+\dots+x_n}(1-p)^{n-(x_1+\dots+x_n)} \quad (x_i = 1, 0).$$

このとき, データの数値 $(x_1, \dots, x_n)$ が得られたとき, この統計モデルの尤度函数は

$$L(p) = p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)}$$

になる. これの右辺の式の形は上の確率質量函数と同じである. しかし, 尤度函数の側では $x_1, \ldots, x_n$ が具体的なデータの数値になっている点が異なる.

記号の簡単のため, データ $x_1, \ldots, x_n$ の中の当たり(=1)の個数を k と書くことにする:

$$k = x_1 + \cdots + x_n.$$

このとき, 尤度函数は次のように書かれる:

$$L(p) = p^k(1-p)^{n-k}$$

実際の計算では尤度函数そのものではなく, その対数である **対数尤度函数** (log-likelihood function)

$$\ell(p) = \log L(p) = k\log p + (n-k)\log(1-p)$$

の方が便利であることが多い. 実際にはさらにこれの -1 倍(**負の対数尤度函数** (negative log-likelihood function)

$$-\ell(p) = -\log L(p) = -k\log p - (n-k)\log(1-p)$$

が使われることが多い. 最尤法の解は負の対数尤度函数を最小化することによって得られる. (χ²分布との関係で -1 倍ではなく, -2 倍することも多い.)

パラメータの値 p の尤度 $L(p)$ もしくはその対数 $\log L(p)$ はBernoulli試行モデルのデータ $x_1, \ldots, x_n$ への適合度の指標の1つとみなされる. だから, Bernoulli試行モデルのデータ $x_1, \ldots, x_n$ への適合度を最大化することは, 対数尤度 $\log L(p)$ を最大化するパラメータ値 p を採用することに等しいと考えることができる. このようにしてモデルのパラメータを決定することを **最尤法** (maximum likelihood method)や **最尤推定法** (maximum likelihood estimation)と呼ぶ. そして, そのようにして求めたパラメータ値を **最尤推定値** (maximum likelihood esitimate, **最尤法の解**) と呼び, 最尤推定値における尤度函数の値を **最大尤度** (maximum likelihood)と呼ぶ. **この段落の定義はBernoulli試行モデル以外の場合にもそのまま一般化される.**

最尤法の解 $p = \hat{p}$ を求めてみよう. そのために対数尤度函数を微分してみる:

$$\ell'(p) = \frac{k}{p} - \frac{n-k}{1-p} = \frac{k-np}{p(1-p)}$$

なので, $\ell'(p)$ は $p < k/n$ で正になり, $p > k/n$ で負になる. ゆえに $\ell(p)$ は $p < k/n$ で単調増加し, $p > k/n$ で単調減少する. ゆえに, 対数尤度函数 $\ell(p)$ を最大化する $p = \hat{p}$ は

$$\hat{p} = \frac{k}{n}$$

になる. Bernoulli試行モデルの「n 回中 k 回が当たりが出た」というデータに関する尤度は成功確率パラメータの値を $p = \hat{p} = k/n$ としたときに最大になり, Bernoulli試行モデルがデータに最もよくフィットするようになる.

この結果を見て, 「n 回中 k 回が当たりが出た」というデータによれば, そのルーレットで当たりが出る確率は k/n であるに違いないと推測することは **誤り** である. 以上で行った計算は単にモデルがデータに最もよくフィットするパラメータ値を求めただけであり, データを生成したルーレットで当たりが出る確率を求めてはいない.

ルーレットを n 回まわしたときに当たりが出る回数 k は, ルーレットを n 回まわす操作を行うごとに違う値になるだろう. データは確率的に揺らぐ. ルーレットで当たりが出る確率(これは未知)を適切に推測するためには, データの確率的揺らぎについても考慮する必要がある. この点については様々な考え方があるが, 後で我々は区間推定(信頼区間)の考え方を学ぶことになるだろう.

3.4 Bernoulli試行モデルの尤度函数の視覚化

前節で求めた「n 回中当たりが k 回出た」というデータに関するBernoulli試行モデルの尤度函数

$$L(p) = p^k(1-p)^{n-k}$$

のグラフを描いてみよう.

しかし, 浮動小数点数での計算では $0 < p < 1$ を満たす p の高次のべきは容易に 0 になってしまう. そこで最大値が 1 になるように定数をかけた次の函数のグラフをプロットすることにしよう:

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

In [41]:
```
function plot_berlik(n, k)
    if k < 0 || k > n
        return plot(; axis=false, tick=false, frame=false, margin=0Plots.mm)
    end
    L(p) = if 0 < p < 1
        exp(-(xlogy(k, safediv(k/n, p))+xlogy(n-k, safediv(1-k/n, 1-p))))
    else
        zero(p)
    end
    plot(L, -0.02, 1.02; label="", title="n=$n, k=$k")
    plot!(; xtick=0:0.1:1, tickfontsize=6)
end
```

Out[41]: plot_berlik (generic function with 1 method)

$n = 10$ を固定した場合

In [42]:
```
n = 10
plot(plot_berlik.(n,(0,1,2))...; size=(800,150), layout=(1,3))
```

Out[42]:

In [43]:
```
plot(plot_berlik.(n,(3,4,5))...; size=(800,150), layout=(1,3))
```

Out[43]:

In [44]:
```
n = 10
plot(plot_berlik.(n,(6,7,8))...; size=(800,150), layout=(1,3))
```

Out[44]:

In [45]:
```
n = 10
plot(plot_berlik.(n,(9,10,11))...; size=(800,150), layout=(1,3))
```

Out[45]:

k/n を固定して n を大きくした場合には, 尤度函数 $L(p)$ の台は k/n に集中して行く.

In [46]:
```
plot(plot_berlik.((10, 100, 1000), (3, 30, 300))...; size=(800,150), layout=(1,3))
```

Out[46]:

このグラフから, データに基づくパラメータの推定において, パラメータの推定値としてどれだけの幅を持たせるかについて, 尤度函数がなにがしかの情報を持っていそうなこともわかる.

3.5 問題: 尤度が高くても全然もっともらしくない例(オーバーフィッティングの例)

データは当たり(1 と書く)とはずれ(0 と書く)がランダムに出るルーレットを n 回まわして得た 1 と 0 の列 $x_1, x_2, \dots, x_n$ として得られるとする.

統計モデルとして以下の2つを考える.

モデル0: Bernoulli試行モデル: このモデルのパラメータ $0 \leq p \leq 1$ を持つ確率質量函数は次のように書けるのであった:

$$P_0(x_1, \dots, x_n | p) = p^{x_1 + \dots + x_n}(1-p)^{n-(x_1+\dots+x_n)} \quad (x_i = 1, 0).$$

モデル1: ルーレットをまわすごとに当たりが出る確率が変わってもよいと想定したモデル: このモデルは n 個のパラメータ $0 \leq p_1, \dots, p_n \leq 1$ を持つ次の確率質量函数によって与えられる:

$$P_1(x_1, \dots, x_n | p_1, \dots, p_n) = \prod_{i=1}^{n} \left(p_i^{x_i}(1-p_i)^{1-x_i}\right) \quad (x_i = 1, 0).$$

データ(1 と 0 の列) $x_1, \dots, x_n$ について, モデル0の最尤法の解 $p = \hat{p}$ は $\hat{p} = k/n$ $(k = x_1 + \dots + x_n)$ になるのであった. 以下の2つを同時に示せ:

(1) モデル1の最尤法の解 $p_i = \hat{p}_i$ は $\hat{p}_i = x_i$ になる.

(2) どのようなデータ $x_1, x_2, \dots, x_n$ についてもモデル1の最大尤度は離散分布モデルの場合に可能な最大の値 1 になる.

注意: 確率と違って確率密度の値はいくらでも大きくなりえるので, 尤度が確率密度の値になる連続分布モデルにおいては尤度はいくらでも大きな値を取り得る. 離散分布モデルの場合には尤度は確率の値になるのでその最大値は 1 になる.

注意: この例はオーバーフィッティング(過剰適合)のシンプルな例になっている. パラメータの個数が多いモデルではオーバーフィッティングし易くなる.

注意: この例を知っていれば, 「尤度が高ければもっともらしい」と考えることは単純に誤りであることがわかる. 尤度を常識的な意味でのもっともらしさと繋げるためには特別な条件が必要になる.

解答例: モデル1において, データ $x_1, x_2, \dots, x_n$ (1 と 0 の列)について, $p_i = x_i$ とおくと,

$$x_i^{x_i}(1-x_i)^{1-x_i} = 1 \quad (x_i = 1, 0)$$

なので($0^0 = 1$ と約束しておく), その尤度は

$$P_1(x_1, \dots, x_n | x_1, \dots, x_n) = \prod_{i=1}^{n} \left(x_i^{x_i}(1-x_i)^{1-x_i}\right) = 1$$

となり, 可能な最大の値になる. ゆえに, 最尤推定値は $\hat{p} = x_i$ となる.

解答終

3.6 問題: 正規分布の標本分布モデルの尤度函数と最尤法

データは n 個の実数達 $x_1, \dots, x_n$ であるとし, 統計モデルとして平均 μ, 分散 σ^2 の正規分布のサイズ n の標本分布を考える. その統計モデルの確率密度函数 $p(x_1, \dots, x_n | \mu, \sigma^2)$ は,

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

とおくとき次のように表される:

$$p(x_1,\dots,x_n|\mu,\sigma^2) = \prod_{i=1}^n p(x_i|\mu,\sigma^2).$$

データ $x_1,\dots,x_n$ に関するこのモデルの尤度函数を最大化する $\mu=\hat{\mu}$, $\sigma^2=\hat{\sigma}^2$ はそれぞれ次になることを示せ:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\hat{\mu})^2.$$

これは, 正規分布の標本分布モデルの最尤法の解が標本平均と(不偏にするための補正無しの)標本分散になることを意味している.

注意: この結果はデータの標本平均と標本分散を求めることは, 正規分布の標本分布モデルがデータに最もよくフィットするような μ と σ^2 を求めることに等しいことがわかる. この点は次の節の要約統計の説明でも触れる.

解答例: 記号の簡単のため以下のように書くことにする:

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

このとき

$$\begin{aligned}\sum_{i=1}^n (x_i-\mu)^2 &= \sum_{i=1}^n ((x_i-\bar{x})-(\mu-\bar{x}))^2 \\ &= \underbrace{\sum_{i=1}^n (x_i-\bar{x})^2}_{=nv^2} - 2(\mu-\bar{x})\underbrace{\sum_{i=1}^n (x_i-\bar{x})}_{=0} + \sum_{i=1}^n (\mu-\bar{x})^2 \\ &= n(v^2+(\mu-\bar{x})^2)\end{aligned}$$

なので, 正規分布の標本分布の密度函数は以下のように表される:

$$\begin{aligned}p(x_1,\dots,x_n|\mu,\sigma^2) &= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2\right) \\ &= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{n}{2\sigma^2}(v^2+(\mu-\bar{x})^2)\right).\end{aligned}$$

ゆえにデータ $x_1,\dots,x_n$ に関するこれの尤度函数を $\ell(\mu,\sigma)$ と書くと,

$$-\frac{2}{n}\ell(\mu,\sigma) = \frac{(\mu-\bar{x})^2+v^2}{\sigma^2} + \log\sigma^2 + \log(2\pi).$$

これを最小化する $\mu=\hat{\mu}$ と $\sigma^2=\hat{\sigma}^2$ が $\hat{\mu}=\bar{x}$, $\hat{\sigma}^2=v^2$ となることを示せばよい.

σ^2 を固定するとき, 上の式を最小化する $\mu=\hat{\mu}$ が $\hat{\mu}=\bar{x}$ になることはすぐにわかる.

$\mu=\bar{x}$ のとき

$$-\frac{2}{n}\ell(\bar{x},\sigma) = \frac{v^2}{\sigma^2} + \log\sigma^2 + \text{const.} \tag{$\sharp$}$$

右辺を σ^2 を1つの変数だとみなして, σ^2 で微分すると

$$-\frac{v^2}{(\sigma^2)^2} + \frac{1}{\sigma^2} = \frac{\sigma^2-v^2}{(\sigma^2)^2}$$

であることから, これは σ^2 の函数として, $\sigma^2 < v_2$ のとき単調減少し, $\sigma^2 > v^2$ のとき単調増加するので, ($\sharp$)を最小化する $\sigma^2=\hat{\sigma}^2$ が $\hat{\sigma}^2=v^2$ であることがわかる.

解答終

3.7 正規分布の標本分布モデルの尤度函数の視覚化

値が大きな点ほど明るくプロットされるヒートマップによって正規分布の標本分布モデルの尤度函数をプロットしてみよう. 尤度函数を与えるためにはデータ $x_1, \ldots, x_n$ が必要になるが, データは簡単のため標準正規分布に従う乱数で生成することにする.

In [47]:
```
function plot_normallik(X, xlim, ylim; kwargs...)
    n = length(X)
    mle = fit_mle(Normal, X)
    μ̂, σ̂² = mean(mle), var(mle)
    f(μ, σ²) = exp(
        loglikelihood(Normal(μ, √σ²), X) -
        loglikelihood(Normal(μ̂, √σ̂²), X)
    )
    x = range(xlim..., 200)
    y = range(ylim..., 200)
    heatmap(x, y, f; colorbar=false)
    plot!(; xguide="μ", yguide="σ²", tickfontsize=7)
    title!("n = $n")
    plot!(; fmt=:png, size=(250, 250))
end
```

Out[47]: plot_normallik (generic function with 1 method)

In [48]:
```
Random.seed!(4649373) # よろしくみなさん
plot((plot_normallik(randn(10), (-1.5, 1.5), (0, 4)) for _ in 1:3)...;
    size=(750,250), layout=(1,3), leftmargin=4Plots.mm, bottommargin=4Plots.mm)
```

Out[48]:

In [49]:
```
plot((plot_normallik(randn(40), (-1.5, 1.5), (0, 4)) for _ in 1:3)...;
    size=(750,250), layout=(1,3), leftmargin=4Plots.mm, bottommargin=4Plots.mm)
```

Out[49]:

In [50]:
```
plot((plot_normallik(randn(160), (-1.5, 1.5), (0, 4)) for _ in 1:3)...;
    size=(750,250), layout=(1,3), leftmargin=4Plots.mm, bottommargin=4Plots.mm)
```

Out[50]:

In [51]:
```
plot((plot_normallik(randn(640), (-1.5, 1.5), (0, 4)) for _ in 1:3)...;
    size=(750,250), layout=(1,3), leftmargin=4Plots.mm, bottommargin=4Plots.mm)
```

Out[51]:

n を大きくすると, サンプルを生成した分布の期待値と分散の値の近くに尤度函数の台が集中して行く.

以下はサンプルを生成する分布を正規分布とは大きく違う指数分布にした場合

In [52]:
```
Random.seed!(4649373) # よろしくみなさん
plot((plot_normallik(rand(Exponential(), 10), (-0.5, 2.5), (0, 4)) for _ in 1:3)...;
    size=(750,250), layout=(1,3), leftmargin=4Plots.mm, bottommargin=4Plots.mm)
```

Out[52]:

In [53]:
```
plot((plot_normallik(rand(Exponential(), 40), (-0.5, 2.5), (0, 4)) for _ in 1:3)...;
    size=(750,250), layout=(1,3), leftmargin=4Plots.mm, bottommargin=4Plots.mm)
```

Out[53]:

In [54]:
```
plot((plot_normallik(rand(Exponential(), 160), (-0.5, 2.5), (0, 4)) for _ in 1:3)...;
    size=(750,250), layout=(1,3), leftmargin=4Plots.mm, bottommargin=4Plots.mm)
```

Out[54]:

In [55]:
```
plot((plot_normallik(rand(Exponential(), 640), (-0.5, 2.5), (0, 4)) for _ in 1:3)...;
    size=(750,250), layout=(1,3), leftmargin=4Plots.mm, bottommargin=4Plots.mm)
```

Out[55]:

3.8 問題: 最小二乗法がある種の正規分布モデルの最尤法に一致すること

データは2つの実数の組が n 個 $(x_1, y_1), \ldots, (x_n, y_n)$ の形式であるとし, 記号の簡単のために次のようにおく:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad v_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,$$
$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \quad v_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2, \quad v_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).$$

統計モデルとして, パラメータ $x_1, \ldots, x_n, \alpha, \beta$ と $\sigma^2 > 0$ を持つ次の確率密度函数で定まる $y_1, \ldots, y_n$ に関する確率分布を考える:

$$p(y_1, \ldots, y_n | x_1, \ldots, x_n, \alpha, \beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - (\alpha + \beta x_i))^2\right).$$

このモデルのデータ $(x_1, y_1), \ldots, (x_n, y_n)$ に関する尤度函数

$$L(\alpha,\beta,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-(\alpha+\beta x_i))^2\right)$$

を最大化する $\alpha=\hat{\alpha}$, $\beta=\hat{\beta}$, $\sigma^2=\hat{\sigma}^2$ は次になることを示せ:

$$\hat{\beta}=\frac{v_{xy}}{v_x^2},\quad \hat{\alpha}=\bar{y}-\hat{\beta}\bar{x},\quad \hat{\sigma}^2=\frac{v_x^2v_y^2-v_{xy}^2}{v_x^2}.$$

注意: この公式は最小二乗法の公式と一致する.

解答例: 尤度函数の形を整理し直そう.

$$\begin{aligned}(y_i-(\alpha+\beta x_i))^2&=(y_i-\bar{y}+\bar{y}-(\alpha+\beta\bar{x})+(\alpha+\beta\bar{x})-(\alpha+\beta x_i))^2\\&=((y_i-\bar{y})-\beta(x_i-\bar{x})+(\bar{y}-(\alpha+\beta\bar{x})))^2\\&=\beta^2(x_i-\bar{x})^2-2\beta(x_i-\bar{x})(y_i-\bar{y})+(y_i-\bar{y})^2\\&+(\bar{y}-(\alpha+\beta\bar{x}))^2+2(y_i-\bar{y})(\bar{y}-(\alpha+\beta\bar{x}))-2\beta(x_i-\bar{x})(\bar{y}-(\alpha+\beta\bar{x}))\end{aligned}$$

より

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}(y_i-(\alpha+\beta x_i))^2&=v_x^2\beta^2-2v_{xy}\beta+v_y^2+(\bar{y}-(\alpha+\beta\bar{x}))^2\\&=v_x^2\left(\beta-\frac{v_{xy}}{v_x^2}\right)^2+\frac{v_x^2v_y^2-v_{xy}^2}{v_x^2}+(\bar{y}-(\alpha+\beta\bar{x}))^2.\end{aligned}$$

ゆえに, 尤度函数の対数の $-2/n$ 倍を $f(\alpha,\beta,\sigma^2)$ と書くと,

$$\begin{aligned}f(\alpha,\beta,\sigma^2)&=-\frac{2}{n}\log L(\alpha,\beta,\sigma^2)\\&=\frac{1}{\sigma^2}\left(v_x^2\left(\beta-\frac{v_{xy}}{v_x^2}\right)^2+\frac{v_x^2v_y^2-v_{xy}^2}{v_x^2}+(\bar{y}-(\alpha+\beta\bar{x}))^2\right)+\log\sigma^2+\log(2\pi).\end{aligned}$$

これを最小化する $\alpha=\hat{\alpha}$, $\beta=\hat{\beta}$, $\sigma^2=\hat{\sigma}^2$ が

$$\hat{\beta}=\frac{v_{xy}}{v_x^2},\quad \hat{\alpha}=\bar{y}-\hat{\beta}\bar{x},\quad \hat{\sigma}^2=\frac{v_x^2v_y^2-v_{xy}^2}{v_x^2}.$$

になることを示せばよい. σ^2 を固定したとき, $f(\alpha,\beta,\sigma^2)$ を最小化する $\alpha=\hat{\alpha}$, $\beta=\hat{\beta}$ が

$$\hat{\beta}=\frac{v_{xy}}{v_x^2},\quad \hat{\alpha}=\bar{y}-\hat{\beta}\bar{x}$$

になることはすぐにわかる. このとき

$$f(\hat{\alpha},\hat{\beta},\sigma^2)=\frac{1}{\sigma^2}\frac{v_x^2v_y^2-v_{xy}^2}{v_x^2}+\log\sigma^2+\log(2\pi)$$

であり, 前節と同様に σ^2 を1つの変数だとみなして σ^2 で微分することによって, これを最小化する $\sigma^2=\hat{\sigma}^2$ が

$$\hat{\sigma}^2=\frac{v_x^2v_y^2-v_{xy}^2}{v_x^2}$$

になることがわかる.

解答終

3.9 最尤法の観点から見た十分統計量

$p(x|\theta)$ はパラメータ θ を持つ x に関する確率密度函数であるとする(確率質量函数の場合も同様である). $p(x|\theta)$ をデータ x に関する統計モデルの密度函数である場合について考える.

このとき, Fiher-Neymanの分解定理より, x の函数 $t(x)$ が **十分統計量** (sufficient statistic)であることと, モデルの密度函数が 0 以上の値を持つ函数たち $a(x)$, $b(t|\theta)$ によって

$$p(x|\theta)=a(x)b(t(x)|\theta)$$

と表されることは同値である. すなわちこの条件を十分統計量の定義として採用してもよい. この条件はデータ x の尤度函数

$$L(\theta) = a(x)b(t(x)|\theta)$$

が θ に関する定数 $a(x)$ 倍の違いを除けば, $t(x)$ と θ だけで尤度函数が表されることと同値である. 尤度 $L(\theta)$ を最大化することと $b(t(x)|\theta)$ を最大化することは同値なので, 最尤法の解は, データ x そのものの値ではなく, データ x から計算される $t(x)$ の値だけで決まることになる.

十分統計量はデータの要約を与えるためにも有用である.

3.9.1 Bernoulli試行モデルの十分統計量

Bernoulli試行モデルのパラメータ p を持つ確率質量函数は

$$P(x_1, \dots, x_n|p) = p^{x_1+\cdots+x_n}(1-p)^{n-(x_1+\cdots+x_n)} \quad (x_i = 1, 0).$$

と書けるので,

$$t(x_1, \dots, x_n) = x_1 + \cdots + x_n$$

は十分統計量になっている. $x_1, \dots, x_n$ そのものの情報を忘れても, $k = x_1 + \cdots + x_n$ の情報さえあれば最尤法を実行できる.

3.9.2 正規分布の標本分布モデルの十分統計量

正規分布の標本分布モデルのパラメータ μ, σ^2 を持つ確率密度函数は

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

とおくと,

$$p(x_1, \dots, x_n|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\Big(-\frac{n}{2\sigma^2}(v^2 + (\mu - \bar{x})^2)\Big)$$

と書けるので,

$$t(x_1, \dots, x_n) = (\bar{x}, v^2) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$$

は十分統計量になっている. すなわち, 標本平均と標本分散の組み合わせは正規分布の標本分布モデルの十分統計量になっている. 補正無しの標本分散 v^2 を不偏分散 $s^2 = (n/(n-1))v^2$ で置き換えても十分統計量のままである.

3.9.3 最小二乗法を与える正規分布モデルの尤度函数を記述する統計量

最小二乗法を与える正規分布モデルのパラメータ $x_1, \dots, x_n, \alpha, \beta, \sigma^2$ を持つ確率密度函数は

$$\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i, \quad v_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \\
\bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i, \quad v_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2, \quad v_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}).
\end{aligned}$$

とおくと,

$$\begin{aligned}
&p(y_1, \dots, y_n|x_1, \dots, x_n, \alpha, \beta, \sigma^2) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\Big(-\frac{n}{2\sigma^2}\big(v_x^2\beta^2 - 2v_{xy}\beta + v_y^2 + (\bar{y} - (\alpha + \beta\bar{x}))^2\big)\Big) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{n}{2\sigma^2}\left(v_x^2\left(\beta - \frac{v_{xy}}{v_x^2}\right)^2 + \frac{v_x^2 v_y^2 - v_{xy}^2}{v_x^2} + (\bar{y} - (\alpha + \beta\bar{x}))^2\right)\right)
\end{aligned}$$

と書けるので,

$$t(y_1, \dots, y_n|x_1, \dots, x_n) = (\bar{x}, \bar{y}, v_x^2, v_y^2, v_{xy})$$

は少し拡張した意味での十分統計量だと言ってよいだろう.

注意: $(\bar{x}, \bar{y}, v_x^2, v_y^2, v_{xy})$ は2変量正規分布の標本分布モデルの十分統計量になっている.

4 記述統計 (要約統計)

$x_1, \ldots, x_n$ の形式のデータや $(x_1, y_1), \ldots, (x_n, y_n)$ の形式のデータが得られたとき, それらの情報を要約するためによく使われる量がある.

それらについて簡単にまとめておこう.

4.1 n 個の 1 と 0 からなるデータ $x_1, \ldots, x_n$ の要約

n 個の 1 と 0 からなるデータ $x_1, \ldots, x_n$ の要約として, $x_1, \ldots, x_n$ の中の 1 の個数 $k = x_1 + \cdots + x_n$ がよく使われる. 報告の形式は「n 個中 k 個が 1 である」になる. その k はBernoulli試行モデルの十分統計量になっているのであった.

そのようなデータの要約の仕方は $x_1, \ldots, x_n$ が並んでいる順序が重要な場合には適さない.

「n 個中の k 個が 1 である」と報告するときには n と k の両方の数値を報告することが重要である. なぜならばデータのサイズ n は統計分析の精度を見積もる上で非常に重要な情報だからである.

n の情報を省いて「1 の割合は 30% であった」のように k/n の情報だけを伝えることは好ましくない.

4.2 より一般の離散的なデータの要約

n 個の $1, 2, \ldots, r$ からなるデータ $x_1, \ldots, x_n$ の要約として, $x_1, \ldots, x_n$ の中含まれる $a = 1, 2, \ldots, r$ の個数 k_a がまとめた $(k_1, \ldots, k_r)$ がよく使われる. ($(k_1, \ldots, k_r)$ はカテゴリカル分布の標本分布モデルの十分統計量になっている.)

そのようなデータの要約の仕方は $x_1, \ldots, x_n$ が並んでいる順序が重要な場合には適さない.

データの要約(summary)を報告する場合には, 必ずデータのサイズがわかるように報告しなければいけない. なぜならばデータのサイズ n は統計分析の精度を見積もる上で非常に重要な情報だからである.

例えば「$1, 2, 3, 4$ の割合がそれぞれ $10\%, 20\%, 30\%, 40\%$ であった」のように k_a/n の情報だけを伝えることは好ましくない. 全体の割合の情報だけではなく, 全体の個数の情報も必ず報告しなければいけない.

4.3 n 個の実数からなるデータ $x_1, \ldots, x_n$ の要約

n 個の実数からなるデータ $x_1, \ldots, x_n$ の要約としては以下がよく使われている:

- 標本平均 $\bar{x}$ と不偏分散 s^2
- 中央値
- 四分位数

標本平均と不偏分散の組み合わせは正規分布の標本分布モデルの十分統計量になっており, それだけではなく, その最尤法の解にもなっているのであった.

標本平均だけではなく, 中央値の情報も報告されているとき, それらに大きなずれが存在するならば, 分布に非対称性があることもわかる.

四分位数は外れ値がある場合の分布の散らばり方の要約として頑健である.

以上で挙げた, 標本平均, 不偏分散, 中央値, 四分位数はデータの分布が単峰型の場合にはデータの要約の仕方として適切だが, 山が2以上ある分布の場合には適さない.

この型のデータの様子は, ヒストグラムや経験累積分布函数(x に x 以下の x_i 達の割合を対応させる函数)をプロットして, **データ全体の様子を視覚的に必ず確認した方がよい.**

4.4 n 個の実数の対からなるデータ $(x_1, y_1), \ldots, (x_n, y_n)$ の要約

n 個の実数の対からなるデータ $(x_1, y_1), \ldots, (x_n, y_n)$ の要約としてよく使われているのは,

- 標本平均 $\bar{x}, \bar{y}$ と不偏分散と不偏共分散 s_x^2, s_y^2, s_{xy}

である.　この5つの量を合わせたものは二変量正規分布の標本分布モデルの十分統計量になっている.

それらの統計量を使って線形回帰の結果を表示することもできる. 線形回帰もデータの要約の仕方だとみなされる.

このようなデータを取得する場合には x_i と y_i のあいだの関係がどうなっているかに興味がある.

そのときに上の5つの要約統計量(や線形回帰の結果)だけしか見ないと, x_i と y_i のあいだの重要な関係を見逃してしまうことがある.

この型のデータを扱う場合には, **散布図** (scatter plot)を描くなどして、**データ全体の様子を視覚的に必ず確認した方がよい.**

4.5 Anscombeの例(アンスコムの例, 再)

以下の散布図(scatter plot)を見れば, 要約統計量だけを見てはいけないことがわかる.

In [56]:

```
using RDatasets
anscombe = dataset("datasets", "anscombe")

SA1 = scatter(anscombe.X1, anscombe.Y1; title="Anscombe 1", label="", msw=0)
SA2 = scatter(anscombe.X2, anscombe.Y2; title="Anscombe 2", label="", msw=0)
SA3 = scatter(anscombe.X3, anscombe.Y3; title="Anscombe 3", label="", msw=0)
SA4 = scatter(anscombe.X4, anscombe.Y4; title="Anscombe 4", label="", msw=0)
plot(SA1, SA2, SA3, SA4; size=(500, 450), layout=(2,2),
    xlim=(3, 20), ylim=(2, 14), xlabel="x", ylabel="y")
```

Out[56]:

In [57]:

```julia
function plot_ols(x, y, α̂, β̂; xlim=nothing,
        title="data and regression line", kwargs...)
    a, b = extrema(x)
    isnothing(xlim) && (xlim = (a - 0.1(b-a), b + 0.1(b-a)))
    scatter(x, y; label="", msw=0)
    plot!(x -> α̂ + β̂*x, xlim...; label="", lw=2)
    plot!(; xlabel="x", ylabel="y", msw=0)
    plot!(; size=(400, 300))
    plot!(; title, kwargs...)
end

function solve_anscombe(x, y;
        title="", xlim=(3, 20), ylim=(2, 14), size=(250, 225))
    @show x̄ = mean(x)
    @show ȳ = mean(y)
    @show sx² = var(x)
    @show sy² = var(y)
    @show sxy = cov(x, y)
    β̂ = sxy/sx²
    α̂ = ȳ - β̂*x̄
    σ̂ = √((n - 1)/n*(sx²*sy² - sxy^2)/sx²)
    @show α̂ β̂ σ̂

    plot_ols(x, y, α̂, β̂; title, xlim, ylim, size)
end

println("---------- Anscombe 1")
PA1 = solve_anscombe(anscombe.X1, anscombe.Y1, title="Anscombe 1")
println("\n---------- Anscombe 2")
PA2 = solve_anscombe(anscombe.X2, anscombe.Y2, title="Anscombe 1")
println("\n---------- Anscombe 3")
PA3 = solve_anscombe(anscombe.X3, anscombe.Y3, title="Anscombe 1")
println("\n---------- Anscombe 4")
PA4 = solve_anscombe(anscombe.X4, anscombe.Y4, title="Anscombe 1")

plot(PA1, PA2, PA3, PA4; size=(500, 450), layout=(2,2))
```

```
---------- Anscombe 1
x̄ = mean(x) = 9.0
ȳ = mean(y) = 7.500909090909093
sx² = var(x) = 11.0
sy² = var(y) = 4.127269090909091
sxy = cov(x, y) = 5.501
α̂ = 3.0000909090909103
β̂ = 0.5000909090909091
σ̂ = 1.1129429904536887

---------- Anscombe 2
x̄ = mean(x) = 9.0
ȳ = mean(y) = 7.500909090909091
sx² = var(x) = 11.0
sy² = var(y) = 4.127629090909091
sxy = cov(x, y) = 5.5
α̂ = 3.000909090909091
β̂ = 0.5
σ̂ = 1.1134927848074192

---------- Anscombe 3
x̄ = mean(x) = 9.0
ȳ = mean(y) = 7.500000000000001
sx² = var(x) = 11.0
sy² = var(y) = 4.12262
sxy = cov(x, y) = 5.497
α̂ = 3.0024545454545466
β̂ = 0.49972727272727274
σ̂ = 1.1126802162509963

---------- Anscombe 4
x̄ = mean(x) = 9.0
ȳ = mean(y) = 7.50090909090909
sx² = var(x) = 11.0
sy² = var(y) = 4.12324909090909
sxy = cov(x, y) = 5.499
α̂ = 3.0017272727272726
β̂ = 0.49990909090909086
σ̂ = 1.112125937113239
```

Out[57]:

4.6 問題: DataSaurusの例

以下のデータは

- http://www.thefunctionalart.com/2016/08/download-datasaurus-never-trust-summary.html (http://www.thefunctionalart.com/2016/08/download-datasaurus-never-trust-summary.html)

から取得したものである.

- CSVファイル (https://raw.githubusercontent.com/genkuroki/Statistics/master/2022/data/Datasaurus_data.csv)
- x座標のみ (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_X.txt), y座標のみ (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_Y.txt)
- x座標のみコンマ付き (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_X_with_commas.txt), y座標のみコンマ付き (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_Y_with_commas.txt)

以下のデータの左列はx座標のデータであり, 右列はy座標のデータである.

(1) 標本平均 $\bar{x}$, $\bar{y}$ と不偏分散と不偏共分散 s_x^2, s_y^2, s_{xy} を小数点以下2桁以上求めよ.

(2) 何らかの方法で散布図(scatter plot)を描け.

さらに datasaurus same stats についてインターネットで検索して統計分析で注意するべきことを学べるサイトを見つけて内容を理解するように努力せよ.

ヒント: (1),(2)についてはインターネット上で適切に検索すれば以上の問題を解くことができるサイトを発見できる.

In [58]:

```python
datasaurus = [
    55.3846 97.1795
    51.5385 96.0256
    46.1538 94.4872
    42.8205 91.4103
    40.7692 88.3333
    38.7179 84.8718
    35.6410 79.8718
    33.0769 77.5641
    28.9744 74.4872
    26.1538 71.4103
    23.0769 66.4103
    22.3077 61.7949
    22.3077 57.1795
    23.3333 52.9487
    25.8974 51.0256
    29.4872 51.0256
    32.8205 51.0256
    35.3846 51.4103
    40.2564 51.4103
    44.1026 52.9487
    46.6667 54.1026
    50.0000 55.2564
    53.0769 55.6410
    56.6667 56.0256
    59.2308 57.9487
    61.2821 62.1795
    61.5385 66.4103
    61.7949 69.1026
    57.4359 55.2564
    54.8718 49.8718
    52.5641 46.0256
    48.2051 38.3333
    49.4872 42.1795
    51.0256 44.1026
    45.3846 36.4103
    42.8205 32.5641
    38.7179 31.4103
    35.1282 30.2564
    32.5641 32.1795
    30.0000 36.7949
    33.5897 41.4103
    36.6667 45.6410
    38.2051 49.1026
    29.7436 36.0256
    29.7436 32.1795
    30.0000 29.1026
    32.0513 26.7949
    35.8974 25.2564
    41.0256 25.2564
    44.1026 25.6410
    47.1795 28.7180
    49.4872 31.4103
    51.5385 34.8718
    53.5897 37.5641
    55.1282 40.6410
    56.6667 42.1795
    59.2308 44.4872
    62.3077 46.0256
    64.8718 46.7949
    67.9487 47.9487
    70.5128 53.7180
    71.5385 60.6410
    71.5385 64.4872
    69.4872 69.4872
    46.9231 79.8718
    48.2051 84.1026
    50.0000 85.2564
    53.0769 85.2564
    55.3846 86.0256
    56.6667 86.0256
    56.1538 82.9487
    53.8462 80.6410
    51.2821 78.7180
    50.0000 78.7180
    47.9487 77.5641
    29.7436 59.8718
```

```
    29.7436 62.1795
    31.2821 62.5641
    57.9487 99.4872
    61.7949 99.1026
    64.8718 97.5641
    68.4615 94.1026
    70.7692 91.0256
    72.0513 86.4103
    73.8462 83.3333
    75.1282 79.1026
    76.6667 75.2564
    77.6923 71.4103
    79.7436 66.7949
    81.7949 60.2564
    83.3333 55.2564
    85.1282 51.4103
    86.4103 47.5641
    87.9487 46.0256
    89.4872 42.5641
    93.3333 39.8718
    95.3846 36.7949
    98.2051 33.7180
    56.6667 40.6410
    59.2308 38.3333
    60.7692 33.7180
    63.0769 29.1026
    64.1026 25.2564
    64.3590 24.1026
    74.3590 22.9487
    71.2821 22.9487
    67.9487 22.1795
    65.8974 20.2564
    63.0769 19.1026
    61.2821 19.1026
    58.7179 18.3333
    55.1282 18.3333
    52.3077 18.3333
    49.7436 17.5641
    47.4359 16.0256
    44.8718 13.7180
    48.7179 14.8718
    51.2821 14.8718
    54.1026 14.8718
    56.1538 14.1026
    52.0513 12.5641
    48.7179 11.0256
    47.1795  9.8718
    46.1538  6.0256
    50.5128  9.4872
    53.8462 10.2564
    57.4359 10.2564
    60.0000 10.6410
    64.1026 10.6410
    66.9231 10.6410
    71.2821 10.6410
    74.3590 10.6410
    78.2051 10.6410
    67.9487  8.7180
    68.4615  5.2564
    68.2051  2.9487
    37.6923 25.7692
    39.4872 25.3846
    91.2821 41.5385
    50.0000 95.7692
    47.9487 95.0000
    44.1026 92.6923
];
```

注意: この例についての詳しい解説が

- Justin Matejka, George Fitzmaurice. Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing. Honourable Mention, ACM SIGCHI Conference on Human Factors in Computing Systems, 2017 [link (https://www.autodesk.com/research/publications/same-stats-different-graphs)]

にある. この解説は非常に面白い. データ全体の様子がわかるようなグラフを作成することが統計分析では必須なことがよくわかる解説になっている.

このサイトは "datasaurus same stats" をインターネットで検索すればすぐに見つかる.

4.6.1 Julia言語を使う解答例

In [59]:
```
# (1)
x, y = datasaurus[:,1], datasaurus[:,2]
@show x̄ = mean(x)
@show ȳ = mean(y)
@show sx² = var(x)
@show sy² = var(y)
@show sxy = cov(x,y);
```

```
x̄ = mean(x) = 54.26327323943663
ȳ = mean(y) = 47.832252816901395
sx² = var(x) = 281.06998759176105
sy² = var(y) = 725.5159610042119
sxy = cov(x, y) = -29.11393267347017
```

In [60]:
```
# (2)
scatter(x, y; label="", tirle="Datasaurus")
```

Out[60]:

こんなことになっていることは要約統計量たちを眺めてもわからない!

データを得たら, 必ず全体の様子を何らかの視覚化によって確認しなければいけない.

これは絶対に必要なことである.

4.6.2 Julia言語の使い方

(1) https://julialang.org/downloads/ (https://julialang.org/downloads/) から current stable release のインストーラをダウンロードして公式バイナリをパソコンに入れる.

(2) juliaを起動して, 適当にコードを打ち込んで計算できることを確認する.

```
julia> sinpi(1/6)
0.5
```

(3) 以下を実行してPlots.jlパッケージ (https://github.com/JuliaPlots/Plots.jl)をJuliaの環境に入れる.

```
julia> ]
pkg> add Plots
ここにインストールの様子が表示される.
pkg> ここでバックスペースキーを押す.
julia>
```

(4) 以下を実行して, パッケージを読み込む.

```
julia> using Statistics
julia> using Plots
```

(4) `datasaurus = [ ... ]` のセルの内容を `julia>` と表示されている画面にコピー&ペーストする.

(5) 以下を順次実行する.

```
julia> x, y = datasaurus[:,1], datasaurus[:,2]
julia> mean(x)
julia> mean(y)
julia> var(x)
julia> var(y)
julia> cov(x,y)
```

これで問題の(1)の答えが得られる.

(6) 次を実行すれば問題(2)の答えが得られる.

```
julia> scatter(x, y)
```

注意: 複雑な計算を十分にできるソフトを自分のパソコンに入れておいた方がよい. もちろんそれはJulia言語である必要はない. 統計学を学習するという目的のためにはR言語やPython言語もお勧めである.

4.6.3 インターネット上の各種計算サイトを使う方法

jp.foxcalculators.com (https://jp.foxcalculators.com/)とkeisan.casio.jp (https://keisan.casio.jp)を利用してみよう.

(1) 平均と不偏分散と不偏共分散はjp.foxcalculators.comの共分散計算機 (https://jp.foxcalculators.com/math/3003.html)で計算できる. データの入力は

- x座標のみコンマ付き (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_X_with_commas.txt), y座標のみコンマ付き (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_Y_with_commas.txt)

からコピー＆ペーストすればよい. 不偏分散が不偏共分散の特別な場合であることに注意せよ. (x_i と x_i の不偏共分散は x_i の不偏分散になる.)

- $\bar{x} \approx 54.263$
- $s_x^2 \approx 281.07$

- $\bar{y} \approx 47.832$
- $s_y^2 \approx 725.516$

共分散計算機

Enter Value of X=

```
25.7692,
25.3846,
41.5385,
95.7692,
95.0000,
92.6923
```

Enter the Numbers with Comma separated(,)

Enter Value of Y=

```
25.7692,
25.3846,
41.5385,
95.7692,
95.0000,
92.6923
```

Enter the Numbers with Comma separated(,)

Calculate

Result:

No.of Inputs	142
X Mean	47.832
Y Mean	47.832
Covariance(X,Y)	725.516

$s_{xy} \approx -29.114$

(2) 散布図はkeisan.casio.jpでの散布図の描画 (https://keisan.casio.jp/exec/system/1412059302)で描ける. データの入力は

- x座標のみ (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_X.txt), y座標のみ (https://github.com/genkuroki/Statistics/blob/master/2022/data/Datasaurus_Y.txt)

からx座標とy座標を別々にコピー＆ペーストすればよい.

散布図の描画

気温別のコーヒーとアイスの売上から散布図で描画します。

	X	Y
1	55.3846	97.1795
2	51.5385	96.0256
3	46.1538	94.4872
4	42.8205	91.4103
5	40.7692	88.3333
6	38.7179	84.8718
7	35.6410	79.8718
8	33.0769	77.5641
9	28.9744	74.4872
10	26.1538	71.4103

In [ ]:
```
1
```