確率分布達の解釈

- 黒木玄
- 2022-04-11〜2022-06-08, 2022-08-10, 2023-03-18, 2024-01-06

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, leftrightarrow, "U=\frac{P}{1-P}" left, "P=\frac{U}{1+U}" right] & \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[d, leftrightarrow, "\alpha=\frac{\nu_1}{2} \parallel \beta=\frac{\nu_2}{2}"] & \mathrm{Gamma}(\alpha,\theta) \arrow[d, leftrightarrow, "\alpha=\frac{\nu_1}{2} \parallel"] \\
\mathrm{Gamma}\left(\frac{\nu_1}{2},2\right)\times\mathrm{Gamma}\left(\frac{\nu_2}{2},2\right) \arrow[r, "(X,Y)\mapsto U=\frac{X}{Y}"] \arrow[d, equal] \arrow[u, shift right, "\text{一般化}"'] & \mathrm{BetaPrime}\left(\frac{\nu_1}{2},\frac{\nu_2}{2}\right) \arrow[r, "\nu_2\to\infty"] \arrow[d, leftrightarrow, "\text{スケール変換}"] & \mathrm{Gamma}\left(\frac{\nu_1}{2},\theta\right) \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[d, shift right] \arrow[u, shift right, "\text{一般化}"'] & \mathrm{FDist}(1,\nu) \arrow[r, "\nu\to\infty"] \arrow[d, shift right] \arrow[u, shift right, "\text{一般化}"'] & \mathrm{Chisq}(1) \arrow[d, shift right] \\
\mathrm{Normal}(0,1)\times\mathrm{Chisq}(\nu=n-1) \arrow[r, "(Z,Y)\mapsto T=\frac{Z}{\sqrt{Y/\nu}}"] \arrow[d, leftrightarrow, "\text{アフィン変換}" '] \arrow[u, "Z^2" '] & \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}\left(\frac{n-1}{2},\frac{2\sigma^2}{n-1}\right) & & \\
\mathrm{Normal}(\mu,\sigma)^n \arrow[u, "\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i"] \arrow[u, shift right, "S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})"'] & &
\end{tikzcd}
$$

標本分布

ベータ分布	$\mathrm{Beta}(\alpha,\beta)$	$\frac{1}{B(\alpha,\beta)}p^{\alpha-1}(1-p)^{\beta-1}dp$	$p=\frac{u}{1+u}$
ベータプライム分布	$\mathrm{BetaPrime}(\alpha,\beta)$	$\frac{1}{B(\alpha,\beta)}\frac{u^{\alpha-1}}{(1+u)^{\alpha+\beta}}du$	$u=\frac{p}{1-p}$

これはほぼF分布

$$
\begin{tikzcd}
\boxed{\text{Bernoulli試行}} \arrow[r, "t=n/L", "L\to\infty"'] & \boxed{\text{連続時間極限}} \\
\begin{array}{l} n\text{回中平均 } np\text{回} \\ K\sim\mathrm{Binomial}(n,p) \end{array} \arrow[r, "n=L", "p=\frac{\lambda}{L}"'] \arrow[d, "p\sim\mathrm{Beta}(\alpha,\beta)"'] & \begin{array}{l} \text{単位時間に平均 } \lambda\text{回} \\ K\sim\mathrm{Poisson}(\lambda) \end{array} \arrow[d, "\lambda\sim\mathrm{Gamma}(\alpha,\theta)"] \\
K\sim\mathrm{BetaBinomial}(n,\alpha,\beta) \arrow[r, "n=L", "\beta=\frac{L}{\theta}"'] & K\sim\mathrm{NegativeBinomial}\left(\alpha,\frac{1}{1+\theta}\right) \\
N-k\sim\mathrm{BetaNegativeBinomial}(k,\alpha,\beta) \arrow[r, "\beta=\frac{L}{\theta}"] & T\sim\frac{1}{\theta}\mathrm{BetaPrime}(k,\alpha) \\
N-k\sim\mathrm{NegativeBinomial}(k,p) \arrow[u, "p\sim\mathrm{Beta}(\alpha,\beta)"] \arrow[r, "N=LT", "p=\frac{\lambda}{L}"'] & T\sim\mathrm{Gamma}\left(k,\frac{1}{\lambda}\right) \arrow[u, "\lambda\sim\mathrm{Gamma}(\alpha,\theta)"'] \\
\text{ちょうど}k\text{回イベントが起こるまでN回} & \text{ちょうど}k\text{回イベントが起こるまでの時間がT} \\
\boxed{\text{Bernoulli試行}} \arrow[r, "L\to\infty", "\text{単位時間にL回試行}"'] & \boxed{\text{連続時間極限}}
\end{tikzcd}
$$

$\boxed{\begin{array}{l}\text{Pólya's urn: } p=\alpha/(\alpha+\beta)\text{で} \\ \text{当たりが出ると } (\alpha,\beta)\to(\alpha+1,\beta) \\ \text{はずれが出ると } (\alpha,\beta)\to(\alpha,\beta+1)\end{array}}$

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v分の1に従う正規分布は自由度v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 問題: 0からθのあいだの一様分布の上限θの不偏推定量
6.5 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
6.6 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
6.7 おまけ: 二項分布とベータ分布の関係のBayes統計への応用
6.7.1 平坦事前分布とimproper事前分布Beta(0,1)の事後分布の比較
6.8 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
6.8.1 様々な事前分布の事後分布のプロット
6.9 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
6.10 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
6.11 おまけ: 二項分布と負の二項分布のP値のBayes統計的解釈への応用
6.12 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)
6.13 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]:

```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 https://github.com/jverzani/SymPyCore.jl/blob/main/src/SymPy/show_sympy.jl#L
@eval SymPy begin
function Base.show(io::IO,  ::MIME"text/latex", x::SymbolicObject)
    out = _sympy_.latex(↓(x), mode="inline",fold_short_frac=false)
    out = replace(out, r"\\frac{"⇒"\\dfrac{")
    print(io, string(out))
end
end
```

In [3]:

```julia
mybinomial(a, k) = exp(loggamma(a+1) - loggamma(k+1) - loggamma(a-k+1))
```

Out[3]: mybinomial (generic function with 1 method)

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.

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 [4]:
```
@syms 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[4]: 0

以下のように素朴に計算することもできる.

In [5]:
```
@syms a::positive b::positive
@syms x y z t
expr = x^2/a + y^2/b
```

Out[5]: $\frac{y^2}{b} + \frac{x^2}{a}$

In [6]:
```
expr = expr(y⇒z-x).expand()
```

Out[6]: $\frac{x^2}{b} - \frac{2xz}{b} + \frac{z^2}{b} + \frac{x^2}{a}$

In [7]:
```
A, B = 1/a+1/b, z/b
expr = sympy.poly(expr(x ⇒ t + B/A), t)
```

Out[7]: $\mathrm{Poly}\left(\frac{a+b}{ab}t^2 + \frac{z^2}{a+b}, t, domain = \mathbb{Z}(z,a,b)\right)$

In [8]:
```
# 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=fals
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,

$$\mu - c\sigma \leq X \leq \mu + c\sigma \iff -c \leq Z = \frac{X - \mu}{\sigma} \leq c$$

なので, 標準正規分布に従う確率変数 Z について $P(-c \leq Z \leq c)$ 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, $P(-c \leq Z \leq c) = 1 - \alpha$ となることと, $1 - P(Z \leq c) = \alpha/2$ すなわち

$$P(Z \leq c) = 1 - \alpha/2$$

となることは同値である. 下の図を見よ.

In [9]:
```
var"P(μ-cσ ≤ X ≤ μ+cσ) = 1-α ⇔ 1 - P(Z ≤ c) = α/2"
```

Out[9]:

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/)では以下のように計算できる. (ただし, `using SpecialFunctions, Distributions` が必要.)

In [10]:
```
@show √2 * erfinv(0.95)
@show √2 * erfinv(0.99);
```

```
√2 * erfinv(0.95) = 1.9599639845400576
√2 * erfinv(0.99) = 2.5758293035489053
```

In [11]:
```
@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:

$$\begin{aligned}
tx-\frac{(x-\mu)^2}{2\sigma^2} &= -\frac{(x-\mu)^2-2\sigma^2tx}{2\sigma^2} = -\frac{x^2-2(\mu+\sigma^2t)x+\mu^2}{2\sigma^2} \\
&= -\frac{(x-(\mu+\sigma^2t))^2+\mu^2-(\mu+\sigma^2t)^2}{2\sigma^2} \\
&= -\frac{(x-(\mu+\sigma^2t))^2-2\sigma^2\mu t+\sigma^4t^2}{2\sigma^2} \\
&= -\frac{(x-(\mu+\sigma^2t))^2}{2\sigma^2}+\mu t+\sigma^2\frac{t^2}{2}
\end{aligned}$$

より

$$\begin{aligned}
E[e^{tX}] &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{tx}e^{-(x-\mu)^2)/(2\sigma^2)}\,dx \\
&= \frac{e^{\mu t+\sigma^2t^2/2}}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} e^{-(x-(\mu+\sigma^2t))^2)/(2\sigma^2)}\,dx = e^{\mu t+\sigma^2t^2/2}.
\end{aligned}$$

こらから $\log E[e^{tX}]=\mu t+\sigma^2t^2/2$ 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: $e^{tX}=1+Xt+X^2t^2/2+O(t^3)$ 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 [12]:
```
plot(LogNormal(0, 1), 0, 8; label="", title="LogNormal(0, 1)")
```

Out[12]:

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

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 [13]:

```julia
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[13]: plot_nct (generic function with 1 method)

In [14]:
```
plot(plot_nct(1), plot_nct(2); size=(800, 250))
```

Out[14]:

In [15]:
```
plot(plot_nct(3), plot_nct(4); size=(800, 250))
```

Out[15]:

In [16]:
```
plot(plot_nct(10), plot_nct(20); size=(800, 250))
```

Out[16]:

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 [17]:
```
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[17]: plot_gamoαθ (generic function with 2 methods)

In [18]:
```
plot(plot_gamoαθ(10), plot_gamoαθ(30), plot_gamoαθ(100); size=(800, 200), layout=(1, 3))
```

Out[18]:

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

$$P(-c \leq T \leq c)$$

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,

- $P(-c \le T \le c) = 95\%$ となる $c \ge 0$ は $c \approx 2.26$
- $P(-c \le T \le c) = 99\%$ となる $c \ge 0$ は $c \approx 3.25$

$\nu = 19$ のとき,

- $P(-c \le T \le c) = 95\%$ となる $c \ge 0$ は $c \approx 2.09$
- $P(-c \le T \le c) = 99\%$ となる $c \ge 0$ 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 [19]:

```
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 [20]:

```
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) = 2.2621571627982044
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) = 2.045229642132704
quantile(TDist(ν), 0.995) = 2.7563859036706053

ν = Inf
quantile(TDist(ν), 0.975) = 1.9599639845400576
quantile(TDist(ν), 0.995) = 2.5758293035489053
```

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),
- quantile(StudentTDistribution(9), 0.995) (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 [21]:

```
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[21]:

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 [22]:

```
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[22]:

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 [23]:

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

In [24]:

```
plot(plot_binpoi(1, 5), plot_binpoi(1, 15), plot_binpoi(1, 50); size=(800, 200), layout
```

Out[24]:

In [25]:

```
plot(plot_binpoi(3, 15), plot_binpoi(3, 45), plot_binpoi(3, 150); size=(800, 200), layo
```

Out[25]:

In [26]:
```
plot(plot_binpoi(10, 50), plot_binpoi(10, 150), plot_binpoi(10, 500); size=(800, 200),
```

Out[26]:

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 [27]:

```julia
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[27]:

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,

$$\frac{k-\lambda}{\lambda} = \frac{x}{\sqrt{\lambda}} + o(\lambda^{-1/2}) = O(\lambda^{-1/2}), \quad \frac{(k-\lambda)^2}{\lambda^2} = \frac{x^2}{\lambda} + o(\lambda^{-1}) = O(\lambda^{-1}), \quad \text{etc.}$$

なので

$$\frac{(k-\lambda)^2}{\lambda} = x^2 + o(1), \quad \frac{(k-\lambda)^3}{\lambda^2} = O(\lambda^{-1/2}) = o(1), \quad \text{etc.}$$

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 [28]:

```
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="Negative
end
```

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

In [29]:

```
plot(plot_gampoi(1, 3), plot_gampoi(3, 1); size=(800, 250), layout=(1, 2))
```

Out[29]:

In [30]:

```
plot(plot_gampoi(6, 2), plot_gampoi(10, 1); size=(800, 250), layout=(1, 2))
```

Out[30]:

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 [31]:

```
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[31]:

In [32]:

```
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[32]:

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(χ^2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χ^2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χ^2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χ^2分布に従うことを意味している.

解答終

In [33]:

```
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[33]:

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: $Z_1, \ldots, Z_n$ 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, $Z_1^2, \ldots, Z_n^2$ 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 [34]:

```
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[34]:

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:

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:

$$p_{\text{approx}}(t|\alpha, \theta, L) = P\Big(\text{round}(Lt-\alpha) \;\Big|\; \alpha, 1/(L\theta)\Big) L.$$

$T = N/L$ 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 $\text{round}(x)$ 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 [35]:

```
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))
    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[35]: plot_nbgam (generic function with 1 method)

In [36]:

```
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[36]:

In [37]:

```
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, 2.8284271
247461903)
(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, 2.8284271
247461903)
(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, 2.8284271
247461903)
```

Out[37]:

In [38]:
```
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, 8.94427190
999916)
(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, 8.94427190
999916)
(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, 8.944271909
99916)
```

Out[38]:

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:

$$\sum_{k=0}^{\kappa-1} \frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!} = \frac{\int_\tau^\infty e^{-\lambda t} t^{\kappa-1}\,dt}{\Gamma(\kappa)/\lambda^\kappa} \qquad (\lambda, \tau > 0,\ \kappa = 1, 2, \ldots)$$

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 [39]:

```
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[39]: plot_cdfpoigam (generic function with 1 method)

In [40]:

```
plot(plot_cdfpoigam(3, 1), plot_cdfpoigam(3, 5); size=(800, 250), bottommargin=4Plots.m
```

Out[40]:

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, χ^2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

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,

- $P(Y \le c) = 95\%$ となる c は $c \approx 3.84$
- $P(Y \le c) = 99\%$ となる c は $c \approx 6.63$

$\nu = 3$ のとき,

- $P(Y \le c) = 95\%$ となる c は $c \approx 5.99$
- $P(Y \le c) = 99\%$ となる c は $c \approx 9.21$

$\nu = 3$ のとき,

- $P(Y \le c) = 95\%$ となる c は $c \approx 7.81$
- $P(Y \le c) = 99\%$ となる c は $c \approx 11.34$

解答終

In [41]:

```
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 [42]:

```
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, $P(Y \le y) = 2Q^{-1}(1-p|\alpha)$ 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.

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,

$$P(-c \le Z \le c) = 0.95, \qquad P(Y \le d) = 0.95$$

を満たす c, d の値はそれぞれ

$$c \approx 1.96, \qquad d \approx 3.84$$

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 [43]:

```
(196//100)^2 ▷ float
```

Out[43]: 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,

$$\begin{aligned} P(-c \le Z \le c) &= \int_{-c}^{c} \frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz = 2\int_0^c \frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz \\ &= 2\int_0^{c^2} \frac{e^{-y/2}}{\sqrt{2\pi}}\frac{1}{2}y^{-1/2}\,dy = \int_0^{c^2} \frac{e^{-y/2}y^{1/2-1}}{2^{1/2}\Gamma(1/2)}\,dy = \\ &= P(Y \le c^2). \end{aligned}$$

つまり $P(-c \le Z \le c) = P(y \le c^2)$ なので, $P(-c \le Z \le c) = 1 - \alpha$ となる c の二乗と $P(Y \le d) = 1 - \alpha$ 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 $T_1, T_2, \dots, T_n$ 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. $T_1, \dots, T_n$ 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 $T_1, \dots, T_n$ とする. (2) $T_1, \dots, T_n$ 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 $1, 2, \dots, n$ を並び替えたもの $i_1, i_2, \dots, i_n$ について,

$$P(T_{i_1}, \dots, T_{i_{k-1}} < t < T_{i_k} < t + dt < T_{i_{k+1}}, \dots, T_{i_n}) = t^{k-1}\, dt\, (1 - (t + dt))^{n-k}.$$

さらに, $1, 2, \dots, n$ を $k-1$ 個の $i_1, \dots, i_{k-1}$ と1個の i_k と $n-k$ 個の $i_{k+1}, \dots, i_n$ 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 $T_1, \dots, T_n$ 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, $T_1, \dots, T_n$ 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

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 [44]:

```julia
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[44]:

```
plot_sim_uniformbeta (generic function with 2 methods)
```

In [45]:
```
plot_sim_uniformbeta(9; size=(800, 500), layout=(3, 3), tickfontsize=6)
```

Out[45]:

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 問題: 0からθのあいだの一様分布の上限θの不偏推定量

$a < b$ と仮定する. a から b のあいだの **一様分布**

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

が確率密度函数

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

によって定義される. $a < x < b$ における不等号は $\leq$ でもよい.

$\theta > 0$ であるとし, 独立な確率変数達 $X_1, \ldots, X_n$ の各々は 0 から θ のあいだの一様分布に従っていると仮定する. $X_1, \ldots, X_n$ の中で k 番目に小さなものを $X_{(k)}$ と書く. 例えば $X_{(n)} = \max(X_1, \ldots, X_n)$ となる. 定数 c_k で次を満たすものを求めよ:

$$E[c_k X_{(k)}] = \theta.$$

このとき $c_k X_{(k)}$ はパラメータ θ の **不偏推定量** (unbiases estimator)であるという.

解答例: $T_i = X_i/\theta$, $T_{(k)} = X_{(k)}/\theta$ とおくと, $T_i \sim \mathrm{Uniform}(0, 1)$ なので, $T_{(k)} \sim \mathrm{Beta}(k, n-k+1)$ となる. ゆえに, $X_{(k)} = \theta T_{(k)}$ と $\mathrm{Beta}(k, n-k+1)$ の平均が $k/(n+1)$ であることより,

$$E[X_{(k)}] = \theta E[T_{(k)}] = \frac{k}{n+1}\theta.$$

したがって,

$$c_k = \frac{n+1}{k}$$

とおくと, $E[c_k X_{(k)}] = \theta$ が成立する.

解答終

注意: 例えば,

$$E\left[\frac{n+1}{n}\max(X_1, \ldots, X_n)\right] = \theta.$$

In [46]:

```
θ = 10
dist = Uniform(0, θ)
n = 10
tmp = zeros(n)
θ̂ = [(n+1)/n*maximum(rand!(dist, tmp)) for _ in 1:10^6]
@show mean(θ̂)
@show std(θ̂)
@show std(n/(n+1)*θ̂; mean=θ)

histogram(θ̂; norm=true, alpha=0.3, bin=100,
    label="unbiased estimator")
vline!([θ]; label="true θ")
plot!(; legend=:topleft, xtick=0:20)
```

```
mean(θ̂) = 9.999618792219303
std(θ̂) = 0.9135079986628124
std((n / (n + 1)) * θ̂; mean = θ) = 1.2315616724759624
```

Out[46]:

In [47]:

```
[
    [mean((n+1)/k * Beta(k, n-k+1)) for k in 1:n];;
    [std((n+1)/k * Beta(k, n-k+1)) for k in 1:n];;
]
```

Out[47]:

```
10×2 Matrix{Float64}:
 1.0  0.912871
 1.0  0.612372
 1.0  0.471405
 1.0  0.381881
 1.0  0.316228
 1.0  0.263523
 1.0  0.218218
 1.0  0.176777
 1.0  0.136083
 1.0  0.0912871
```

6.5 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,

$$P(T_{(k)} \le p) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}.$$

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:

$$\begin{aligned} &K \sim \mathrm{Binomial}(n, p), \quad T \sim \mathrm{Beta}(k, n-k+1) \\ &\implies P(K \ge k) = P(T \le p), \quad P(K \le k-1) = 1 - P(T \le p). \end{aligned}$$

k を $k+1$ に置き換えると,

$$\begin{aligned} &K \sim \mathrm{Binomial}(n, p), \quad T \sim \mathrm{Beta}(k+1, n-k) \\ &\implies P(K \ge k+1) = P(T \le p), \quad P(K \le k) = 1 - P(T \le p). \end{aligned}$$

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,

$$P(T_{(k)} \le p) = \frac{\int_0^p t^{k-1}(1-t)^{n-k}\,dt}{B(k, n-k+1)}$$

が得られるので, 平易に得られる

$$P(T_{(k)} \le p) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j}$$

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 [48]:

```
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[48]: `plot_cdfbinbeta (generic function with 1 method)`

In [49]:

```
plot(plot_cdfbinbeta(20, 0.2), plot_cdfbinbeta(20, 0.4); size=(800, 250), bottommargin=
```

Out[49]:

6.6 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.

6.7 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,

$$\begin{aligned} T \sim \mathrm{Beta}(k, n-k+1) &\implies P(K \ge k) = P(T \le p_0), \\ T \sim \mathrm{Beta}(k+1, n-k) &\implies P(K \le k) = P(T \ge p_0). \end{aligned}$$

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「$p \le p_0$」という仮説が成立する確率は二項分布モデルにおける「$p \le p_0$」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「$p \ge p_0$」という仮説が成立する確率は二項分布モデルにおける「$p \ge p_0$」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. 「$p \le p_0$」や「$p \ge p_0$」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」の問題における, 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

6.7.1 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「$p \le p_0$」という仮説が成立する確率は二項分布モデルにおける「$p \le p_0$」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 [50]:
```
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[50]: plot_2betas (generic function with 1 method)

In [51]:
```
plot(plot_2betas(20, 5), plot_2betas(40, 10), plot_2betas(80, 20);
    size=(800, 200), layout=(1, 3), tickfontsize=7)
```

Out[51]:

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.

6.8 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 $P \leq 0.3$ 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,

$$\begin{aligned} P(k|n,p)\varphi(p) &= C\,p^k(1-p)^{n-k}\cdot p^{a-1}(1-p)^{b-1} = C\,p^{a+k-1}(1-p)^{b+n-k-1} \\ \int_0^1 P(k|n,p)\varphi(p)\,dp &= C\int_0^1 p^{a+k-1}(1-p)^{b+n-k-1}\,dp = C\,B(a+k,b+n-k). \end{aligned}$$

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 [52]:

```
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[52]: `plot_betaposterior (generic function with 1 method)`

6.8.1 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 [53]:

```
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[53]:

In [54]:

```
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[54]:

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 [55]:

```
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[55]:

In [56]:

```
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[56]:

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 [57]:

```
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[57]:

In [58]:

```
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[58]:

6.9 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:

$$\begin{aligned} &N - k \sim \mathrm{NegativeBinomial}(k, p), \quad T \sim \mathrm{Beta}(k, n-k+1) \\ &\implies P(N \le n) = P(T \le p), \quad P(N > n) = 1 - P(T \le p). \end{aligned}$$

In [59]:

```
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), p)")
    plot!(; legend=:bottomright, xlabel="m")
    title!("α=$α, p=$p")
end
```

Out[59]: `plot_cdfnegbinbeta (generic function with 1 method)`

In [60]:
```
plot(plot_cdfnegbinbeta(5, 0.7), plot_cdfnegbinbeta(5, 0.3);
    size=(800, 250), bottommargin=4Plots.mm)
```

Out[60]:

6.10 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 [61]:
```
cdf(NegativeBinomial(3.14159, 0.7), 2), cdf(Beta(3.14159, 2+1), 0.7)
```

Out[61]: (0.8243946882545398, 0.8243946882545398)

In [62]:

```
plot(plot_cdfnegbinbeta(3.14159, 0.7),plot_cdfnegbinbeta(3.14159, 0.3);
    size=(800, 250), bottommargin=4Plots.mm)
```

Out[62]:

6.11 おまけ: 二項分布と負の二項分布のP値のBayes統計的解釈への応用

二項分布の場合と同様に, 負の二項分布の場合にもベータ分布は共役事前分布になっている.

以下では二項分布と負の二項分布とベータ分布の関係から, 二項分布モデルおよび負の二項分布モデルの片側検定のP値がBayes統計的な解釈を持つことを説明しよう. ただし, P値や検定などに関する解説はここでは省略し, 結果だけを簡潔に記す.

ベータ分布の累積分布函数と二項分布・負の二項分布の累積分布函数達の関係より,

$$
\begin{aligned}
\frac{\int_0^p t^{k-1}(1-t)^{n-k}\,dt}{B(k,n-k+1)} &= \sum_{i=k}^{n}\binom{n}{i}p^i(1-p)^{n-i} = \sum_{j=k}^{n}\binom{j-1}{k-1}p^k(1-p)^{j-k},\\
\frac{\int_p^1 t^{k}(1-t)^{n-k-1}\,dt}{B(k+1,n-k)} &= \sum_{i=0}^{k}\binom{n}{i}p^i(1-p)^{n-i} = \sum_{j=n+1}^{\infty}\binom{j-1}{(k+1)-1}p^{k+1}(1-p)^{j-(k+1)},\\
\frac{\int_p^1 t^{k-1}(1-t)^{n-k-1}\,dt}{B(k,n-k)} &= \sum_{i=0}^{k-1}\binom{n-1}{i}p^i(1-p)^{(n-1)-i} = \sum_{j=n}^{\infty}\binom{j-1}{k-1}p^k(1-p)^{j-k}.
\end{aligned}
$$

すなわち,

$$
\begin{aligned}
\mathrm{cdf}(\mathrm{Beta}(k,n-k+1),p) &= \mathrm{ccdf}(\mathrm{Binomial}(n,p),k-1)\\
&= \mathrm{cdf}(\mathrm{NegativeBinomial}(k,p),n-k),\\
\\
\mathrm{ccdf}(\mathrm{Beta}(k+1,n-k),p) &= \mathrm{cdf}(\mathrm{Binomial}(n,p),k)\\
&= \mathrm{ccdf}(\mathrm{NegativeBinomial}(k+1,p),n-(k+1)),\\
\\
\mathrm{ccdf}(\mathrm{Beta}(k,n-k),p) &= \mathrm{cdf}(\mathrm{Binomial}(n-1,p),k-1)\\
&= \mathrm{ccdf}(\mathrm{NegativeBinomial}(k,p),(n-1)-k).
\end{aligned}
$$

これは「試行回数 n, 成功回数 k」というデータについて, 以下のようになることを含む:

$$
\begin{aligned}
&(\text{二項分布 } \mathrm{Binomial}(n,p) \text{ での仮説 } p \le p_0 \text{ の片側検定のP値})\\
&= (\text{負の二項分布 } \mathrm{NegativeBinomial}(k,p) \text{ での仮説 } p \le p_0 \text{ の片側検定のP値})\\
&= (\text{ベータ分布 } \mathrm{Beta}(k,n-k+1) \text{ 内で仮説 } p \le p_0 \text{ が成立する確率})\\
\\
&(\text{二項分布 } \mathrm{Binomial}(n,p) \text{ での仮説 } p \ge p_0 \text{ の片側検定のP値})\\
&= (\text{ベータ分布 } \mathrm{Beta}(k+1,n-k) \text{ 内で仮説 } p \ge p_0 \text{ が成立する確率})\\
\\
&(\text{負の二項分布 } \mathrm{NegativeBinomial}(k,p) \text{ での仮説 } p \ge p_0 \text{ の片側検定のP値})\\
&= (\text{ベータ分布 } \mathrm{Beta}(k,n-k) \text{ 内で仮説 } p \ge p_0 \text{ が成立する確率}).
\end{aligned}
$$

これは以下が成立していることを意味している:

(1) 二項分布 $\mathrm{Binomial}(n,p)$ と負の二項分布 $\mathrm{NegativeBinomial}(k,p)$ のどちらにおいても, 仮説 $p \le p_0$ の片側検定のP値は, improper事前分布 $\mathrm{Beta}(0,1)$ から定まる事後分布内でその仮説が成立する確率に等しい.

(2) 二項分布 $\mathrm{Binomial}(n,p)$ での仮説 $p \ge p_0$ の片側検定のP値は, improper事前分布 $\mathrm{Beta}(1,0)$ から定まる事後分布内でその仮説が成立する確率に等しい.

(3) 負の二項分布 $\mathrm{NegativeBinomial}(k,p)$ での仮説 $p \ge p_0$ の片側検定のP値は, improper事前分布 $\mathrm{Beta}(0,0)$ から定まる事後分布内で仮説 $p \ge p_0$ が成立する確率に等しい.

このとき (2)と(3)が一致していないことに注意せよ

In [63]:
```
rd(x) = round(x; digits=10)
n, k, p = 10, 1:9, 0.2
```

Out[63]:
```
(10, 1:9, 0.2)
```

In [64]:
```
[
    @. cdf(Beta(k, n-k+1), p);;
    @. ccdf(Binomial(n, p), k-1);;
    @. cdf(NegativeBinomial(k, p), n-k);;
] .▷ rd
```

Out[64]:
```
9×3 Matrix{Float64}:
 0.892626     0.892626     0.892626
 0.62419      0.62419      0.62419
 0.3222       0.3222       0.3222
 0.120874     0.120874     0.120874
 0.0327935    0.0327935    0.0327935
 0.00636938   0.00636938   0.00636938
 0.000864358  0.000864358  0.000864358
 7.79264e-5   7.79264e-5   7.79264e-5
 4.1984e-6    4.1984e-6    4.1984e-6
```

In [65]:
```
[
    @. ccdf(Beta(k+1, n-k), p);;
    @. cdf(Binomial(n, p), k);;
    @. ccdf(NegativeBinomial(k+1, p), n-(k+1));;
] .▷ rd
```

Out[65]:
```
9×3 Matrix{Float64}:
 0.37581   0.37581   0.37581
 0.6778    0.6778    0.6778
 0.879126  0.879126  0.879126
 0.967207  0.967207  0.967207
 0.993631  0.993631  0.993631
 0.999136  0.999136  0.999136
 0.999922  0.999922  0.999922
 0.999996  0.999996  0.999996
 1.0       1.0       1.0
```

In [66]:
```
[
    @. ccdf(Beta(k, n-k), p);;
    @. cdf(Binomial(n-1, p), k-1);;
    @. ccdf(NegativeBinomial(k, p), (n-1)-k);;
] .▷ rd
```

Out[66]:
```
9×3 Matrix{Float64}:
 0.134218  0.134218  0.134218
 0.436208  0.436208  0.436208
 0.738198  0.738198  0.738198
 0.914358  0.914358  0.914358
 0.980419  0.980419  0.980419
 0.996934  0.996934  0.996934
 0.999686  0.999686  0.999686
 0.999981  0.999981  0.999981
 0.999999  0.999999  0.999999
```

6.12 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:

$$K \sim \mathrm{Binomial}(n, p), \quad T \sim \mathrm{Beta}(k, n-k+1), \quad M = N - k \sim \mathrm{NegativeBinomial}(k, p)$$

のとき,

$$\begin{aligned} P(K \geq k) &= P(T \leq p) = P(M \leq n-k) = P(N \leq n), \\ P(K < k) &= P(T > p) = P(M > n-k) = P(N > n). \end{aligned}$$

これを具体的な式で書くと,

$$\sum_{i=k}^{n} \binom{n}{i} p^i (1-p)^{n-i} = \frac{\int_0^p t^{k-1}(1-t)^{n-k}}{B(k, n-k+1)}$$

$$= \sum_{i=0}^{n-k} \binom{k+i-1}{i} p^k (1-p)^i = \sum_{j=k}^{n} \binom{j-1}{j-k} p^k (1-p)^{j-k},$$

$$\sum_{i=0}^{k-1} \binom{n}{i} p^i (1-p)^{n-i} = \frac{\int_p^1 t^{k-1}(1-t)^{n-k}}{B(k, n-k+1)}$$

In [67]:

```
k, p = 3.14, 0.5772
mycdf_negbin(k, p, m) = sum(mybinomial(k+i-1,i)*p^k*(1-p)^i for i in 0:m)
[(m, mycdf_negbin(k, p, m), cdf(Beta(k, m+1), p)) for m in 0:10]
```

Out[67]:

```
11-element Vector{Tuple{Int64, Float64, Float64}}:
 (0, 0.17805929618323893, 0.1780592961832391)
 (1, 0.4144493933217376, 0.4144493933217378)
 (2, 0.6213370607769629, 0.6213370607769628)
 (3, 0.7712059353810816, 0.7712059353810814)
 (4, 0.8684705352614055, 0.8684705352614053)
 (5, 0.9271948544617901, 0.9271948544617898)
 (6, 0.9608790456560383, 0.960879045656038)
 (7, 0.9794746055099702, 0.9794746055099699)
 (8, 0.9894399474401324, 0.9894399474401322)
 (9, 0.994655134192169, 0.9946551341921688)
 (10, 0.9973319810761049, 0.9973319810761047)
```

In [68]:

```
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="ccdf(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
    plot!(; xlabel="k", xtick)
    title!("n=$n, p=$p")
end
```

Out[68]:

```
plot_cdf_bin_beta_nb (generic function with 1 method)
```

In [69]:

```
plot(plot_cdf_bin_beta_nb(10, 0.3), plot_cdf_bin_beta_nb(30, 0.3);
    size=(800, 250), bottommargin=4Plots.mm)
```

Out[69]:

In [70]:

```
plot(plot_cdf_bin_beta_nb(100, 0.3; xtick=0:5:100), plot_cdf_bin_beta_nb(300, 0.3; xtic
    size=(800, 250), bottommargin=4Plots.mm)
```

Out[70]:

6.13 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 [71]:
```
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[71]: plot_cdf_poi_gam (generic function with 1 method)

In [72]:
```
plot(plot_cdf_poi_gam(1, 1), plot_cdf_poi_gam(1, 3; xtick=0:5:100);
    size=(800, 250), bottommargin=4Plots.mm)
```

Out[72]:

In [73]:
```
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[73]:

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 [74]:

```
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="Monte Carlo")
    plot!(x, x -> pdf(BetaBinomial(n, α, β), round(Int, x)); label="BetaBin(n,α,β)", lw
    plot!(x, x -> pdf(Binomial(n, α/(α+β)), round(Int, x)); label="Bin(n,α/(α+β))", ls=
    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[74]: plot_beta (generic function with 1 method)

In [75]:
```
plot(plot_bb(20, 3, 7), plot_bb(20, 9, 21); size=(800, 250))
```

Out[75]:

In [76]:
```
plot(plot_bb(20, 30, 70), plot_bb(20, 90, 210); size=(800, 250))
```

Out[76]:

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($\mathrm{BetaBin}(\alpha, \beta)$)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 $\mathrm{BetaBinomial}(n, \alpha, \beta)$ は 分布 $\mathrm{Binomial}(n, \alpha/(\alpha+\beta)$ 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 $\mathrm{Beta}(\alpha, \beta)$ が $\alpha/(\alpha+\beta)$ に集中するようになるからである.

In [77]:
```
plot(plot_beta(3, 7), plot_beta(9, 21); size=(800, 250))
```

Out[77]:

In [78]:
```
plot(plot_beta(30, 70), plot_beta(90, 210); size=(800, 250))
```

Out[78]:

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

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,

$$E[K] = \frac{n\alpha}{\alpha+\beta}, \quad E[K(K-1)] = \frac{n(n-1)\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}.$$

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 [79]:

```
@syms 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[79]:

$$\frac{abn\,(a+b+n)}{(a+b)^2\,(a+b+1)}$$

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

$$P(k|n,\alpha,\beta) = \binom{n}{k}\frac{B(\alpha+k,\beta+n-k)}{B(\alpha,\beta)} \qquad (k=0,1,\dots,n)$$

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:

$$\begin{aligned} P(k|n,\alpha,\beta) &= \binom{n}{k}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+n)}\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)}\frac{\Gamma(\beta+n-k)}{\Gamma(\beta)} \\ &= \frac{n!}{k!(n-k)!}\frac{\alpha(\alpha+1)\cdots(\alpha+k-1)\cdot\beta(\beta+1)\cdots(\beta+n-k-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+n-1)} \\ &= \frac{\alpha(\alpha+1)\cdots(\alpha+k-1)}{k!}\frac{\beta(\beta+1)\cdots(\beta+n-k-1)}{(n-k)!} \\ &\quad\times\frac{n!}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+n-1)} \\ &= \binom{\alpha+k-1}{k}\binom{\beta+(n-k)-1}{n-k}\Big/\binom{\alpha+\beta+n-1}{n} \end{aligned}$$

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) 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:

$$P(k|n,\alpha,\beta,\delta)=\binom{n}{k}\frac{\alpha(\alpha+\delta)\cdots(\alpha+(k-1)\delta)\cdot\beta(\beta+\delta)\cdots(\beta+(n-k-1)\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)\cdots(\alpha+\beta+(n-1)\delta)}\quad(k=0,1,\dots,n).$$

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 [80]:

```julia
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[80]: 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 [81]:

```julia
plot(plot_hg_bin_bb(10, 3, 9), plot_hg_bin_bb(10, 10, 30); size=(800, 250))
```

Out[81]:

In [82]:

```julia
plot(plot_hg_bin_bb(10, 30, 90), plot_hg_bin_bb(10, 100, 300); size=(800, 250))
```

Out[82]:

超幾何分布，二項分布，ベータ二項分布の期待値

$$P(k|n,\alpha,\beta,\delta) = \binom{n}{k}\frac{\alpha(\alpha+\delta)\cdots(\alpha+(k-1)\delta)\ \beta(\beta+\delta)\cdots(\beta+(n-k-1)\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)\cdots(\alpha+\beta+n\delta)}$$ とおく.

赤玉 α 個, 白玉 β 個が入った袋を考える.

(1) 同時に n 個ランダムに玉を取り出したとき, 出た赤玉の個数が k 個である確率は

$$\binom{n}{k}\frac{\alpha(\alpha-1)\cdots(\alpha-k+1)\ \beta(\beta-1)\cdots(\beta-n+k+1)}{(\alpha+\beta)(\alpha+\beta-1)\cdots(\alpha+\beta-n+1)} = P(k|n,\alpha,\beta,-1).$$

(2) 玉をランダムに取り出して戻すことを n 回くり返したとき, 赤玉が取り出された回数が k 回である確率は

$$\binom{n}{k}\frac{\overbrace{\alpha\alpha\cdots\alpha}^{k個}\ \overbrace{\beta\beta\cdots\beta}^{n-k個}}{\underbrace{(\alpha+\beta)(\alpha+\beta)\cdots(\alpha+\beta)}_{n個}} = P(k|n,\alpha,\beta,0).$$

(3) 玉をランダムに取り出し, 出た玉と同じ色の玉を2個袋に戻すことを n 回くり返したとき, (毎回袋の中の玉の個数は1個増える), その n 回中に赤玉が取り出された回数が k 回である確率は

$$\binom{n}{k}\frac{\alpha(\alpha+1)\cdots(\alpha+k-1)\ \beta(\beta+1)\cdots(\beta+n-k-1)}{(\alpha+\beta)(\alpha+\beta+1)\cdots(\alpha+\beta+n-1)} = P(k|n,\alpha,\beta,1).$$

以上の (1), (2), (3) はそれぞれ $\delta=-1,0,1$ の場合に対応しており, そのどの場合でも,

$$(*)\qquad \sum_k P(k|n,\alpha,\beta,\delta) = 1.$$

(1), (2), (3) のどの場合にも k の期待値が $\frac{n\alpha}{\alpha+\beta}$ となることを示そう. $\delta=-1,0,1$ のとき,

$$\begin{aligned}
(k\text{の期待値}) &= \sum_k k\,P(k|n,\alpha,\beta,\delta)\\
&= \sum_k k\,\frac{n!}{k!(n-k)!}\,\frac{\alpha(\alpha+\delta)\cdots(\alpha+(k-1)\delta)\ \beta(\beta+\delta)\cdots(\beta+(n-k-1)\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)\cdots(\alpha+\beta+n\delta)}\\
&= \frac{n\alpha}{\alpha+\beta}\sum_k \frac{(n-1)!}{(k-1)!(n-k)!}\,\frac{(\alpha+\delta)\cdots(\alpha+(k-1)\delta)\ \beta(\beta+\delta)\cdots(\beta+(n-k-1)\delta)}{(\alpha+\beta+\delta)\cdots(\alpha+\beta+n\delta)}\\
&= \frac{n\alpha}{\alpha+\beta}\sum_k P(k-1|n-1,\alpha+\delta,\beta,\delta) \overset{(*)}{=} \frac{n\alpha}{\alpha+\beta}.
\end{aligned}$$

注意 同様にして $k(k-1)\cdots(k-m+1)$ の期待値が $\frac{n(n-1)\cdots(n-m+1)\,\alpha(\alpha+\delta)\cdots(\alpha+(m-1)\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)\cdots(\alpha+\beta+(m-1)\delta)}$ となることを示せる. $m=1,2$ の場合から k の分散を求めることができる. □

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,

$$\begin{aligned}
&E[K(K-1)\cdots(K-m+1)] \\
&= \sum_k k(k-1)\cdots(k-m+1)\binom{n}{k}\frac{\alpha(\alpha+\delta)\cdots(\alpha+(k-1)\delta)\cdot\beta(\beta+\delta)\cdots(\beta+(n-k-1)\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)\cdots(\alpha+\beta+(n-1)\delta)} \\
&= n(n-1)\cdots(n-m+1)\frac{\alpha(\alpha+\delta)\cdots(\alpha+(m-1)\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)\cdots(\alpha+\beta+(m-1)\delta)} \\
&\times \underbrace{\sum_k \binom{n-m}{k-m}\frac{(\alpha+m\delta)(\alpha+(m+1)\delta)\cdots(\alpha+(k-1)\delta)\cdot\beta(\beta+\delta)\cdots(\beta+(n-k-1)\delta)}{(\alpha+\beta+m)(\alpha+\beta+(m+1)\delta)\cdots(\alpha+\beta+(n-1)\delta)}}_{=\sum_k P(k|n-m,\alpha+m,\beta)=1} \\
&= n(n-1)\cdots(n-m+1)\frac{\alpha(\alpha+\delta)\cdots(\alpha+(m-1)\delta)}{(\alpha+\beta)(\alpha+\beta+\delta)\cdots(\alpha+\beta+(m-1)\delta)}.
\end{aligned}$$

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).

In [83]:

```
@syms 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[83]:

$$\frac{abn\,(a+b+dn)}{(a+b)^2\,(a+b+d)}$$

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

$$P(m|L,\alpha,\beta) = \binom{L}{m}\frac{B(\alpha+m,\beta+L-m)}{B(\alpha,\beta)} \quad (m = 0,1,\ldots,L)$$

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,

$$
\begin{aligned}
P(m|L,\alpha,L/\theta) &= \binom{L}{m}\frac{B(\alpha+m,L/\theta+L-m)}{B(\alpha,L/\theta)} \\
&= \frac{L(L-1)\cdots(L-m+1)}{m!}\frac{(L/\theta)^{\alpha}}{((1+\theta)/\theta)L)^{\alpha+m}} \\
&\times \frac{(((1+\theta)/\theta)L)^{\alpha+m}B(\alpha+m,((1+\theta)/\theta)L-m)}{(L/\theta)^{\alpha}B(\alpha,L/\theta)} \\
&= \underline{L(L-1)\cdots(L-m+1)}
\end{aligned}
$$

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:

$$
\begin{aligned}
\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\,dp &= \frac{(\lambda/L)^{\alpha-1}(1-\lambda/L)^{L/\theta-1}}{B(\alpha,L/\theta)}\frac{d\lambda}{L} \\
&= \frac{\lambda^{\alpha-1}(1-\lambda/L)^{L/\theta-1}}{L^{\alpha}B(\alpha,L/\theta)}\,d\lambda \\
&= \frac{1}{\theta^{\alpha}}\underbrace{\frac{1}{(L/\theta)^{\alpha}B(\alpha,L/\theta)}}_{\to\Gamma(\alpha)}\underbrace{\left(1-\frac{\lambda}{L}\right)^{L/\theta-1}}_{\to\exp(-\lambda/\theta)}\lambda^{\alpha-1}\,d\lambda \\
&\to \frac{1}{\theta^{\alpha}\Gamma(\alpha)}e^{-\lambda/\theta}\lambda^{\alpha-1}\,d\lambda.
\end{aligned}
$$

$L\to\infty$ 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 [84]:

```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[84]: plot_bbnb (generic function with 1 method)

In [85]:

```julia
plot(plot_bbnb.((10, 30), 3, 2)...; size=(800, 250))
```

Out[85]:

In [86]:

```julia
plot(plot_bbnb.((100, 300), 3, 2)...; size=(800, 250))
```

Out[86]:

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
- 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=5536931452160381057)]

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:

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

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 [87]:

```
@syms 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[87]: $\dfrac{k\beta\,(k+\alpha-1)\,(\alpha+\beta-1)}{(\alpha-2)\,(\alpha-1)^2}$

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 [88]:

```
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 [89]:

```
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 [90]:

```
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 [91]:

```
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)) = 0.9999999999999936
sum((m * pdf_betanegbin(m, k, α, β) for m = 0:mmax)) = 5.142857142857027
μ = mean_betanegbin(k, α, β) = 5.142857142857143
sum(((m - μ) ^ 2 * pdf_betanegbin(m, k, α, β) for m = 0:mmax)) = 23.26530612244749
var_betanegbin(k, α, β) = 23.26530612244898
√(sum(((m - μ) ^ 2 * pdf_betanegbin(m, k, α, β) for m = 0:mmax))) = 4.823412290323883
σ = std_betanegbin(k, α, β) = 4.823412290324038
```

Out[91]:

In [92]:

```
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[92]: plot_bnb_bp (generic function with 1 method)

In [93]:

```
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[93]:

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

$$\mathrm{Pólya}(n, \alpha, \beta)$$

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 $\mathrm{Beta}(\alpha, \beta)$ 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 $p \sim \mathrm{Beta}(\alpha, \beta)$ に関するBernoulli試行 $\mathrm{Bernoulli}(p)^n$ 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. Bernoulli試行と一様乱数生成の繰り返しとベータ分布とPólyaの壺の関係については以下のリンク先も参照せよ:

- https://twitter.com/genkuroki/status/1529044220887310336 (https://twitter.com/genkuroki/status/1529044220887310336)