

Bernoulli試行と関連確率分布

- 黒木玄
- 2022-04-09~2022-05-31

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。

目次

[1 WolframAlphaの使い方](#)

[2 一様乱数の生成の繰り返し](#)

[3 Bernoulli試行](#)

▼ [4 離散分布](#)

[4.1 離散分布と確率質量関数](#)

[4.2 離散確率変数](#)

[4.3 離散確率変数の期待値](#)

[4.4 離散確率変数の期待値を取る操作の基本性質](#)

[4.5 確率変数の一般的な定義](#)

[4.6 離散確率変数に関する確率](#)

[4.7 離散確率変数の分散と標準偏差](#)

[4.8 問題: 期待値と分散の特徴付け](#)

[4.9 問題: 平均二乗誤差のバイアス・バリアンس分解](#)

[4.10 カテゴリカル分布の定義](#)

[4.11 問題: ルーレットの例の賞金の期待値と分散](#)

[4.12 Bernoulli分布の定義](#)

[4.13 問題: Bernoulli分布の期待値と分散](#)

[4.14 オッズの定義と例](#)

[4.15 オッズとロジットおよびロジスティック変換](#)

[4.16 Bernoulli試行の確率分布](#)

[4.17 問題: \$n=3\$, \$p=0.3\$ の場合のBernoulli試行の確率分布](#)

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

[4.19 二項定理](#)

[4.20 二項分布の定義](#)

[4.21 問題: 二項分布の期待値と分散](#)

[4.22 問題: 当たりが30%の確率で出るくじを100回引いた場合](#)

[4.23 幾何分布の定義](#)

[4.24 幾何分布の期待値と分散](#)

[4.25 問題: 当たりが出る確率が1%のガチャを当たりが出るまで回す場合](#)

[4.26 負の二項展開](#)

[4.27 負の二項分布の定義](#)

[4.28 負の二項分布の期待値と分散](#)

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

[4.30 多項分布の定義](#)

▼ [5 連続分布など](#)

[5.1 連続分布と確率密度関数](#)

[5.2 一様分布の定義](#)

[5.3 正規分布の定義](#)

[5.4 連続確率変数](#)

[5.5 連続確率変数に関する確率とその期待値](#)

[5.6 確率変数の確率密度関数](#)

[5.7 確率変数の確率質量関数](#)

[5.8 連続分布における確率変数の期待値を取る操作の基本性質](#)

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

1 WolframAlphaの使い方

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

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

は次のように入力すれば計算できる:

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

二項係数を含む式

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

は次のように入力すれば計算できる:

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

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

積分

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

は次のように入力すれば計算できる:

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

これはほぼ $5\% = 0.05$ になる。

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

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

```
In [2]: 1 # Override the Base.show definition of SymPy.jl:
2 # https://github.com/JuliaPy/SymPy.jl/blob/29c5bfd1d10ac53014fa7fef468bc8deccadc2fc/src/types.
3
4 @eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::SymbolicObject)
5     print(io, as_markdown("\displaystyle " * sympy.latex(x, mode="plain", fold_short_frac=false)))
6 end
7 @eval SymPy function Base.show(io::IO, ::MIME"text/latex", x::AbstractArray{Sym})
8     function toeqnarray(x::Vector{Sym})
9         a = join(["\displaystyle " * sympy.latex(x[i]) for i in 1:length(x)], "\\")
10        """\left[ \begin{array}{r}$a\end{array} \right]"""
11    end
12    function toeqnarray(x::AbstractArray{Sym,2})
13        sz = size(x)
14        a = join([join("\displaystyle " .* map(sympy.latex, x[i,:]), "&") for i in 1:sz[1]],
15        "\left[ \begin{array}{c} " * repeat("r",sz[2]) * "}" * a * "\end{array}\right]")
16    end
17    print(io, as_markdown(toeqnarray(x)))
18 end
```

2 一様乱数の生成の繰り返し

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

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

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

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

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

上と同じ結果を与えるコードは次のようにもっとシンプルに書ける。

```
In [4]: 1 rand(10)
```

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

3 Bernoulli試行

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

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

```
In [5]: 1 @show rand(Bernoulli(0.3), 20);
```

```
rand(Bernoulli(0.3), 20) = Bool{0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1}
```

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

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

```
In [6]: 1 ccdf(Binomial(20, 0.3), 10)
```

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

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

```
In [7]: 1 T = rand(20)
        2 for i in eachindex(T) @printf "%.16f\n" T[i] end
```

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

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

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

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

Bernoulli試行に限らず、何らかの確率分布に従う乱数の生成の独立試行は、0以上1未満の一樣乱数の生成を経由して作ることができる。その意味でコンピュータでのプログラミングでおなじみの `rand()` 関数は基本的である。

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

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

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

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

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

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

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

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

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

4 離散分布

4.1 離散分布と確率質量関数

以下, a_i 達は互いに異なると仮定する.

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

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

確率分布を単に分布と呼ぶことがよくある.

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

$$P(a_1) = p_1, P(a_2) = p_2, P(a_3) = p_3, \dots$$

このとき,

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

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

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

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

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

を扱うことになる.

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

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

という確率の値の表で定まる離散分布で自然にモデル化される.

4.2 離散確率変数

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

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

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

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

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

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

このとき,

$$X \sim P(x)$$

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

$$X \sim \text{Dist}$$

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

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

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

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

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

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

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

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

4.3 離散確率変数の期待値

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

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

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

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

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

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

例えば、上のルーレットの例では、

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

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

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

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

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

4.4 離散確率変数の期待値を取る操作の基本性質

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

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

証明: 確率変数 $f(X), g(X)$ と定数 α, β について,

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

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

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

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

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

証明終.

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

4.5 確率変数の一般的な定義

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

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

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

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

4.6 離散確率変数に関する確率

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

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

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

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

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

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

と定めると,

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

と期待値で書けるので, 期待値の概念は確率の概念を含んでいるといえる.

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

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

となる. さらに,

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

となるので,

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

となって、上の結果と一致する。

4.7 離散確率変数の分散と標準偏差

離散確率変数 $Y = f(X)$ の期待値を

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

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

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

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

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

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

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

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

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

確率変数の分散については

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

という公式もよく使われる。この公式は以下のようにして証明される:

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

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

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

4.8 問題: 期待値と分散の特徴付け

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

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

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

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

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

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

解答終

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

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

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

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

4.9 問題: 平均二乗誤差のバイアス・バリエンス分解

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

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

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

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

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

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

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

たし算の順序を交換すれば欲しい公式が得られる。

解答終

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

4.10 カテゴリカル分布の定義

番号の集合 $\{1, 2, \dots, r\}$ 上の確率質量関数

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

が定める有限離散分布を **カテゴリカル分布** と呼び、次のように表す:

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

確率の表は次のようになる:

x	1	2	...	r
$P(x)$	p_1	p_2	...	p_r

多項分布の確率質量関数は次のようにも書ける:

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

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

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

```
In [10]: 1 categ = Categorical(0.1, 0.2, 0.3, 0.4)
```

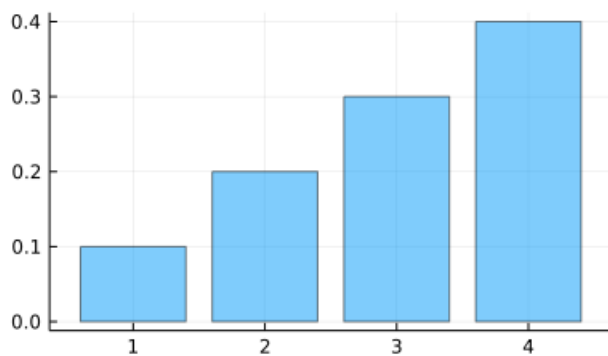
```
Out[10]: Categorical{Float64, Vector{Float64}}(support=Base.OneTo(4), p=[0.1, 0.2, 0.3, 0.4])
```

```
In [11]: 1 @show rand(categ, 20);
```

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

```
In [12]: 1 bar(support(categ), a → pdf(categ, a); label="", title="Categorical(0.1, 0.2, 0.3, 0.4)", alp
```

```
Out[12]: Categorical(0.1, 0.2, 0.3, 0.4)
```



4.11 問題: ルーレットの例の賞金の期待値と分散

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

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

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

解答例: X の期待値, 分散, 標準偏差はそれぞれ

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

$g(X)$ の期待値, 分散, 標準偏差はそれぞれ

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

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

解答終

X の期待値と分散と標準偏差:

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

```
μ_X = 3.0
```

X の分散と標準偏差:

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

```
σ²_X = 1.0
σ_X = 1.0
```

$g(X)$ の期待値:

```
In [15]: 1 μ_g = 10000 * 0.1 + 3000 * 0.2 + 1000 * 0.3 + 300 * 0.4 ▷ Int
          2 @show μ_g;
```

$\mu_g = 2020$

$g(X)$ の分散と標準偏差:

```
In [16]: 1 σ²_g = (10000 - 2020)^2 * 0.1 + (3000 - 2020)^2 * 0.2 + (1000 - 2020)^2 * 0.3 + (300 - 2020)^2 * 0.4
          2 σ_g = √σ²_g
          3 @show σ²_g round(σ_g; digits=1);
```

$\sigma^2_g = 8055600$

$\text{round}(\sigma_g; \text{digits} = 1) = 2838.2$

以下はJulia言語のより高級な機能を使った計算法である.

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

$\mu_X = 3$

$\sigma^2_X = 1$

$\sigma_X = 1$

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

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

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

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

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

$\mu_X = 3$

$\sigma^2_X = 1$

$\sigma_X = 1$

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

$\mu_g = 2020$

$\sigma^2_g = 8055600$

$\text{round}(\sigma_g; \text{digits} = 1) = 2838.2$

4.12 Bernoulli分布の定義

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

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

この分布の確率質量関数は次のようになる:

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

4.13 問題: Bernoulli分布の期待値と分散

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

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

$$E[X] = \sum_{x=1,0} xp(x) = 1p + 0(1-p) = p,$$

$$E[X^2] = \sum_{x=1,0} x^2p(x) = 1^2p + 0^2(1-p) = p,$$

$$\text{var}(X) = E[X^2] - E[X]^2 = p - p^2 = p(1-p).$$

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

解答終

```
In [21]: 1 p = 0.3
         2 bern = Bernoulli(p)
```

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

```
In [22]: 1 @show mean(bern) var(bern);
```

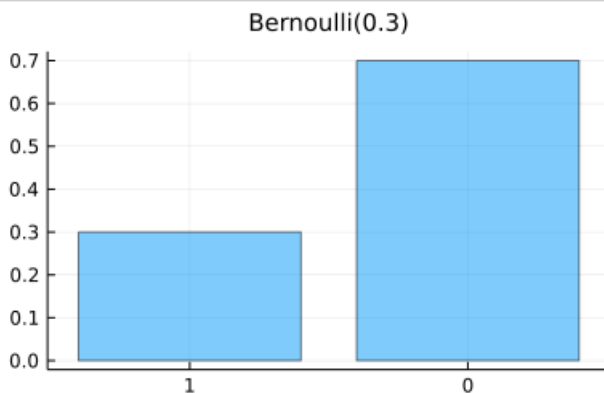
```
mean(bern) = 0.3
var(bern) = 0.21
```

```
In [23]: 1 @show rand(bern, 20);
```

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

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

Out[24]:



4.14 オッズの定義と例

成功確率 p に対して, 失敗確率分の成功確率

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

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

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

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

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

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

4.15 オッズとロジットおよびロジスティック変換

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

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

によって、

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

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

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

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

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

4.16 Bernoulli試行の確率分布

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

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

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

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

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

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

の積

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

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

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

とシンプルに書来直される。

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

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

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

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

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

確率質量関数の全体は以下のように記述される:

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

これら8通りの確率の和は

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

と確かにぴったり 1 になっている.

解答終

In [25]: 1 27//1000 + 3*63//1000 + 3*147//1000 + 343//1000 ▶ Int

Out[25]: 1

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

Bernoulli試行から以下の確率分布が得られる:

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

これらの連続極限として,

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

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

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

4.19 二項定理

n 個から k 個を選ぶ組み合わせの数を

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

と書き, **二項係数** と呼ぶ.

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

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

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

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

特に $y = 1$ のとき,

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

これらの結果を **二項定理** と呼ぶ.

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

4.20 二項分布の定義

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

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

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

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

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

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

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

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

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

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

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

4.21 問題: 二項分布の期待値と分散

K は二項分布に従う確率変数であるとする:

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

K の期待値と分散を求めよ.

解答例: 二項定理を使う. K の期待値は次のようにして得られる:

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

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

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

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

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

解答終

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

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

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

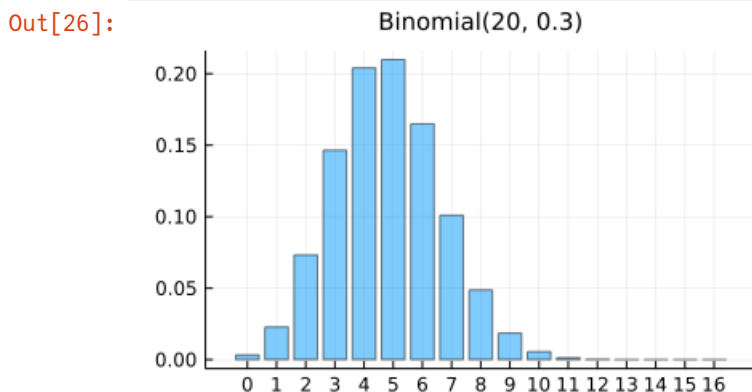
分散についても

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

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

二項分布における確率質量関数のグラフは以下のようになる。

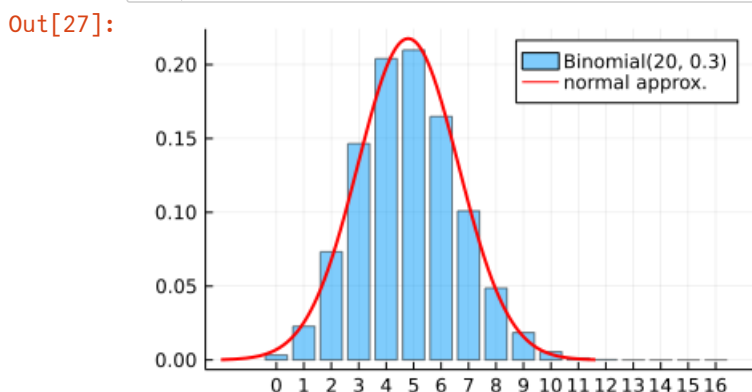
```
In [26]: 1 bin = Binomial(16, 0.3)
2 x = support(bin)
3 bar(x, x → pdf(bin, x); label="", title="Binomial(20, 0.3)", alpha=0.5)
4 plot!(; xtick=x)
```



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

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

```
In [27]: 1 bin = Binomial(16, 0.3)
2 μ, σ = mean(bin), std(bin)
3 x = support(bin)
4 bar(x, x → pdf(bin, x); label="Binomial(20, 0.3)", alpha=0.5)
5 plot!(; xtick=x)
6 plot!(Normal(μ, σ); label="normal approx.", lw=2, c=:red)
```



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

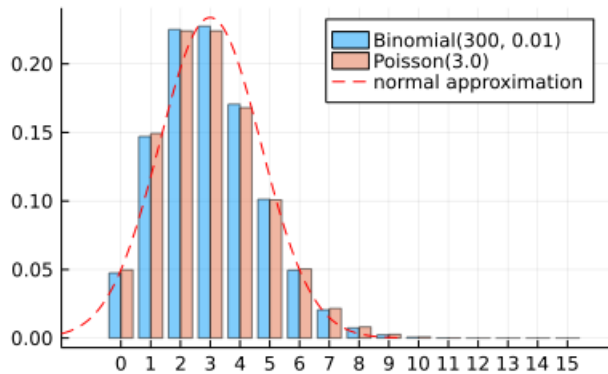
より正確に言うと、固定された $\lambda > 0$ について、 $p = \lambda/n$ とおくと、二項分布 $\text{Binomial}(n, p)$ は $n \rightarrow \infty$ でPoisson分布 $\text{Poisson}(\lambda)$ に近づく。Poisson分布(ポアソン分布)の定義を知らない人はこれをPoisson分布の定義だと思ってよい。この点については後で説明する予定である。

```

In [28]: 1 λ, n = 3, 100
2 p = λ/n
3 bin = Binomial(n, p)
4 μ, σ = mean(bin), std(bin)
5 x = 0:5λ
6 groupedbar(x, [pdf.(bin, x) pdf.(Poisson(n*p), x)]);
7     label=["Binomial(300, 0.01)" "Poisson($ (n*p))"], alpha=0.5)
8 plot!(; xtick=x, xlim=(-2, 16.5))
9 plot!(Normal(μ, σ); label="normal approximation", lw=1, c=:red, ls=:dash)

```

Out[28]:

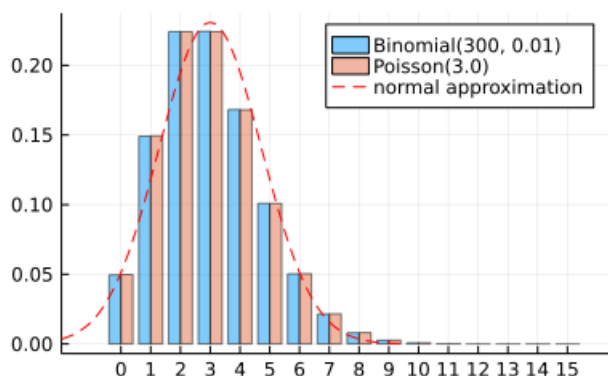


```

In [29]: 1 λ, n = 3, 1000
2 p = λ/n
3 bin = Binomial(n, p)
4 μ, σ = mean(bin), std(bin)
5 x = 0:5λ
6 groupedbar(x, [pdf.(bin, x) pdf.(Poisson(n*p), x)]);
7     label=["Binomial(300, 0.01)" "Poisson($ (n*p))"], alpha=0.5)
8 plot!(; xtick=x, xlim=(-2, 16.5))
9 plot!(Normal(μ, σ); label="normal approximation", lw=1, c=:red, ls=:dash)

```

Out[29]:



4.22 問題: 当たりが30%の確率で出るくじを100回引いた場合

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

- (1) 当たりの回数 K の期待値と分散と標準偏差を求めよ。
- (2) 当たりの回数 K が25回以下になる確率と20回以下になる確率を求めよ。
- (3) 当たりの回数 K が35回以上になる確率と40回以上になる確率を求めよ。

有効桁3桁で求めよ。

解答例: $p = 0.3, n = 100$ とおく。

- (1) K の期待値, 分散, 標準偏差:

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

(2) $K \leq 25$ となる確率:

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

$K \leq 30$ となる確率:

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

(2) $K \geq 35$ となる確率:

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

$K \geq 40$ となる確率:

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

解答終

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

- [sum_{k=0}^{25}binom\(100,k\)0.3^k\(1-0.3\)^{100-k}](https://www.wolframalpha.com/input?i=sum_{k=0}^{25}binom(100,k)0.3^k(1-0.3)^{100-k}) (https://www.wolframalpha.com/input?i=sum_{k=0}^{25}binom(100,k)0.3^k(1-0.3)^{100-k})

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

```
In [30]: 1 logP(n, p, k) = logabsbinomial(n, k)[1] + k*log(p) + (n-k)*log(1-p)
2
3 p = 0.3
4 n = 100
5 @show n*p
6 @show n*p*(1-p)
7 @show sqrt(n*p*(1-p))
8 for Kmax in (25, 20)
9     @eval @show exp(logsumexp(logP(n, p, k) for k in 0:$Kmax))
10 end
11 for Kmin in (35, 40)
12     @eval @show exp(logsumexp(logP(n, p, k) for k in $Kmin:n))
13 end
```

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

```
In [31]: 1 p = 0.3
2 n = 100
3 @eval @show bin = Binomial($n, $p)
4 @show mean(bin)
5 @show var(bin)
6 @show std(bin)
7 for Kmax in (25, 20)
8     @eval @show cdf(bin, $Kmax)
9 end
10 for Kmin in (35, 40)
11     @eval @show ccdf(bin, $Kmin - 1)
12 end
```

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

4.23 幾何分布の定義

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

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

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

に等しいので,

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

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

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

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

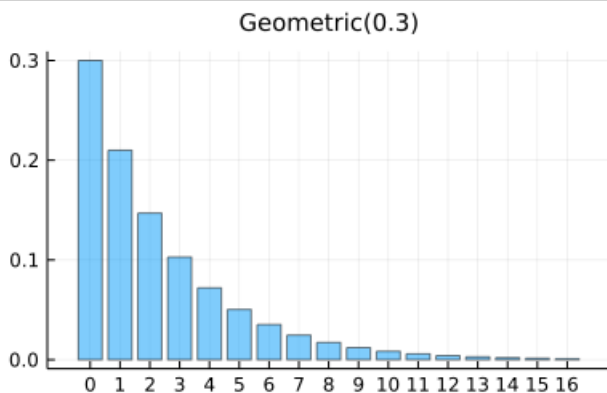
$$\sum_{m=0}^{\infty} P(m) = p \sum_{m=0}^{\infty} (1 - p)^m = \frac{p}{1 - (1 - p)} = 1.$$

$1 - (1 - p)$ が p になることがこれ以後頻繁に現れる.

幾何分布の確率質量関数の形は次のようになる.

```
In [32]: 1 geom = Geometric(0.3)
2 x = 0:16
3 bar(x, x → pdf(geom, x); label="", title="Geometric(0.3)", alpha=0.5)
4 plot!(; xtick = x)
```

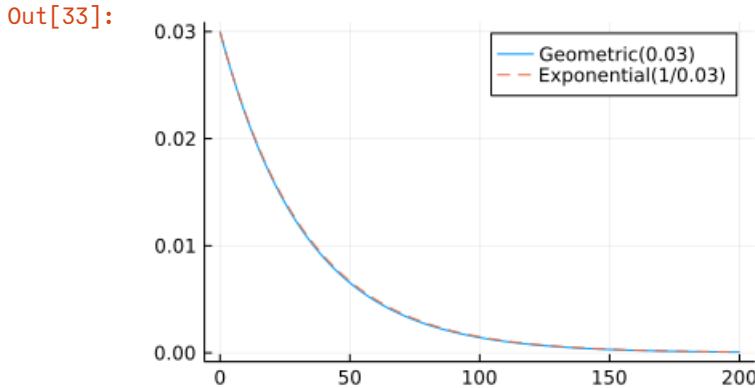
Out[32]:



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

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

```
In [33]: 1 geom = Geometric(0.03)
2 p = succprob(geom)
3 x = 0:200
4 plot(x, x → pdf(geom, x); label="Geometric($p)")
5 plot!(Exponential(1/p), 0, 200; ls=:dash, label="Exponential(1/$p)")
```



4.24 幾何分布の期待値と分散

幾何分布に従う確率変数 M の期待値と分散は次のようになる:

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

詳しくは以下のように計算される. 上で経験した等比級数の和の計算から,

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

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

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

これらの公式を使うと,

$$E[M] = \sum_{m=0}^{\infty} mP(m) = \sum_{m=1}^{\infty} mp(1-p)^m = p(1-p) \sum_{m=1}^{\infty} m(1-p)^{m-1} = p(1-p) \frac{1}{p^2} = \frac{1-p}{p},$$

$$\begin{aligned} E[M(M-1)] &= \sum_{m=0}^{\infty} m(m-1)P(m) = \sum_{m=2}^{\infty} m(m-1)p(1-p)^m \\ &= p(1-p)^2 \sum_{m=2}^{\infty} m(m-1)(1-p)^{m-2} = p(1-p)^2 \frac{2}{p^3} = \frac{2(1-p)^2}{p^2} \end{aligned}$$

$$E[M(M-1)] = E[M^2] - E[M],$$

$$E[M^2] = E[M(M-1)] + E[M] = \frac{2(1-p)^2}{p^2} + \frac{1-p}{p},$$

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

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

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

になる. N の期待値と分散は上の結果より, 次のようになる:

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

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

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

4.25 問題: 当たりが出る確率が1%のガチャを当たりが出るまで回す場合

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

(1) N の期待値と分散と標準偏差を求めよ.

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

確率や標準偏差の値は有効桁3桁以上が合っていれば正解とする.

解答例: $p = 0.01$ とおく.

(1) N の期待値と分散と標準偏差:

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

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

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

N が200回以上になる確率 = 最初の199回分がすべて外れの確率:

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

N が300回以上になる確率 = 最初の299回分がすべて外れの確率:

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

解答終

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

注意: p が小さな幾何分布の標準偏差は

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

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

In [34]:

```
1 p = 0.01
2 @show 1/p
3 @show (1 - p)/p^2
4 @show sqrt((1 - p)/p^2)
5 for Nmin in (150, 200, 300)
6     @eval @show (1 - p)^($Nmin - 1)
7 end

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

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

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

Monte Carlo 法による近似計算:

```
In [36]: 1 N = rand(Geometric(0.01), 10^8) .+ 1
2 @show mean(N)
3 @show var(N)
4 @show std(N)
5 @show mean(>= (150), N)
6 @show mean(>= (200), N)
7 @show mean(>= (300), N);
```

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

4.26 負の二項展開

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

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

と定める. たとえば, $m = 0, 1, 2, 3$ の場合には,

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

$a = n$ も0以上の整数のとき, これは組み合わせの数と一致する:

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

$f(x) = (1+x)^a$ について,

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

なので,

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

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

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

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

次の形の二項係数を **負の二項係数** と呼ぶことにする:

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

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

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

この公式を **負の二項展開** と呼ぶことにする.

4.27 負の二項分布の定義

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

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

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

に等しいので,

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

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

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

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

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

注意: 以上および以下の計算は k が正の整数でなくても,

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

なので, $k > 0$ ならば確率質量関数

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

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

4.28 負の二項分布の期待値と分散

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

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

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

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

になる(この式の形は二項分布の確率質量関数に近いが少しだけ違っている).

それらの期待値と分散は次のようになる:

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

これらは以下のようにして示される. 二項係数について

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

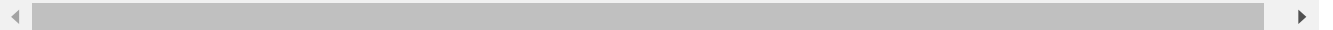
が成立することと負の二項展開を使うと、負の二項分布に従う確率変数 M の期待値と分散は以下のように計算される:

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

上の結果より、 $N = M + k$ の期待値と分散は次のようになる:

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

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



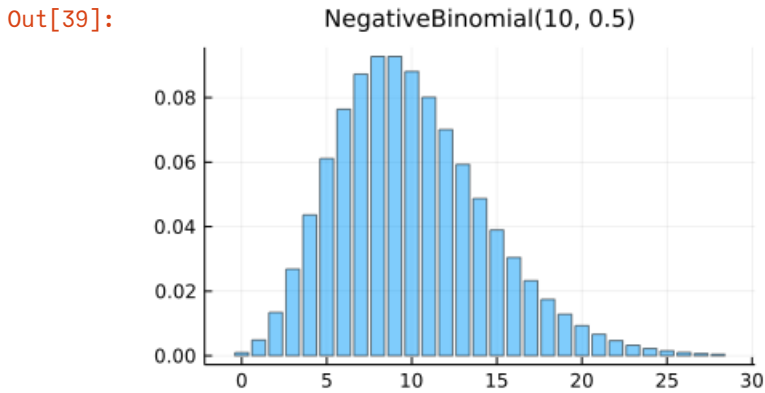
```
In [37]: 1 @vars k p
          2 k*(k+1)*(1-p)^2/p^2 + k*(1-p)/p - k^2*(1-p)^2/p^2 ► simplify
```

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

```
In [38]: 1 k, p = 4, 0.25
          2 negbin = NegativeBinomial(k, p)
          3 @show mean(negbin), k*(1-p)/p
          4 @show mean(negbin) + k, k/p
          5 @show var(negbin), k*(1-p)/p^2;

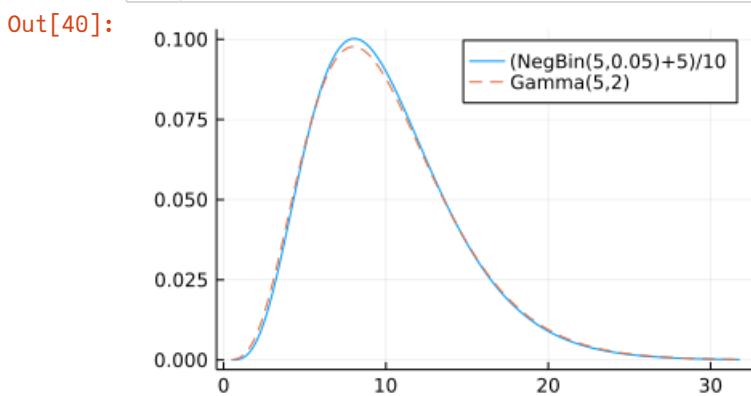
(mean(negbin), (k * (1 - p)) / p) = (12.0, 12.0)
(mean(negbin) + k, k / p) = (16.0, 16.0)
(var(negbin), (k * (1 - p)) / p ^ 2) = (48.0, 48.0)
```

```
In [39]: 1 negbin = NegativeBinomial(10, 0.5)
2 μ, σ = mean(negbin), std(negbin)
3 x = 0:round(Int, μ + 4σ)
4 bar(x, x → pdf(negbin, x); label="",
5       title="NegativeBinomial(10, 0.5)", alpha=0.5)
6 plot!(; xtick=0:5:maximum(x)+5)
```



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

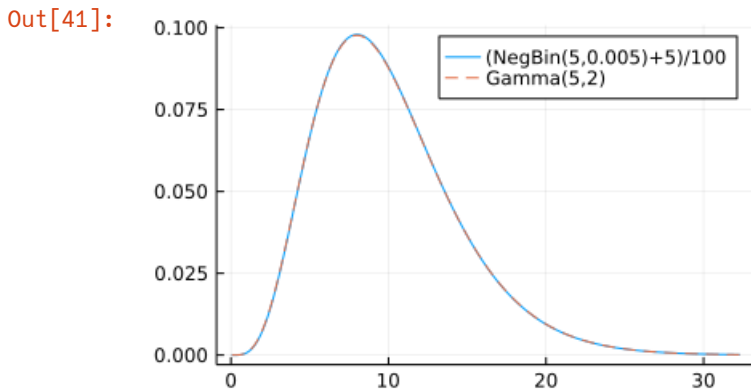
```
In [40]: 1 k, θ = 5, 2
2 L = 10
3 p = 1/(L*θ)
4
5 negbin = NegativeBinomial(k, p)
6 μ, σ = mean(negbin), std(negbin)
7 m = 0:round(Int, μ + 5σ)
8 n = m .+ k
9 t = n ./ L
10 plot(t, pdf.(negbin, m)*L; label="(NegBin($k,$p)+$k)/$L")
11 plot!(t, pdf.(Gamma(k, θ), t); label="Gamma($k,$θ)", ls=:dash)
```



```

In [41]: 1 k, θ = 5, 2
          2 L = 100
          3 p = 1/(L*θ)
          4
          5 negbin = NegativeBinomial(k, p)
          6 μ, σ = mean(negbin), std(negbin)
          7 m = 0:round(Int, μ + 5σ)
          8 n = m .+ k
          9 t = n ./ L
          10 plot(t, pdf.(negbin, m)*L; label="(NegBin($k,$p)+$k)/$L")
          11 plot!(t, pdf.(Gamma(k, θ), t); label="Gamma($k,$θ)", ls=:dash)

```

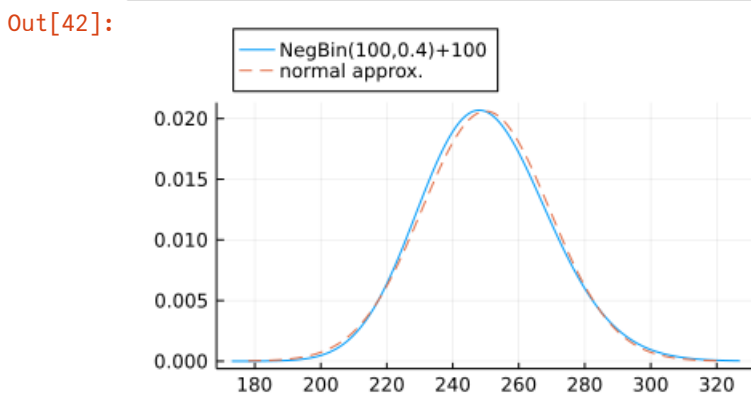


k が大きくなると、負の二項分布 $\text{NegativeBinomial}(k, p)$ は正規分布で近似される。

```

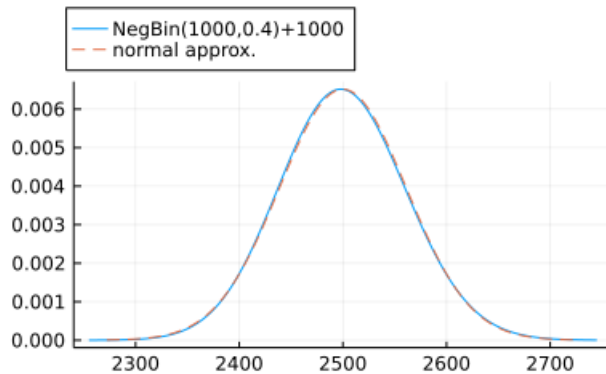
In [42]: 1 k, p = 100, 0.4
          2
          3 negbin = NegativeBinomial(k, p)
          4 μ, σ = mean(negbin), std(negbin)
          5 m = round(Int, μ - 4σ):round(Int, μ + 4σ)
          6 n = m .+ k
          7 plot(n, pdf.(negbin, m); label="NegBin($k,$p)+$k")
          8 plot!(Normal(μ+k, σ); label="normal approx.", ls=:dash)
          9 plot!(); legend=:outertop)

```



```
In [43]: 1 k, p = 1000, 0.4
2
3 negbin = NegativeBinomial(k, p)
4 μ, σ = mean(negbin), std(negbin)
5 m = round(Int, μ - 4σ):round(Int, μ + 4σ)
6 n = m .+ k
7 plot(n, pdf.(negbin, m); label="NegBin($k,$p)+$k")
8 plot!(Normal(μ+k, σ); label="normal approx.", ls=:dash)
9 plot!(); legend=:outertop
```

Out[43]:



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

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

(1) N の期待値と分散と標準偏差を求めよ.

(2) N が150回以上, 200回以上になる確率を求めよ.

確率や標準偏差の値は有効桁3桁以上が合っていれば正解とする.

解答例: $p = 0.05$, $k = 5$ とおく.

(1) N の期待値と分散と標準偏差:

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

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

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

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

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

解答終

```
In [44]: 1 p = 5//100
2 k = 5
3 @show 5/p 5(1 - p)/p^2 √(5(1 - p)/p^2);
```

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

In [45]:

```
1 p = 0.05
2 k = 5
3 for Nmin in (150, 200)
4     @show 1 - sum(binomial(n-1, k-1)*p^k*(1-p)^(n-k) for n in 1:Nmin-1)
5 end
```

```
1 - sum((binomial(n - 1, k - 1) * p ^ k * (1 - p) ^ (n - k) for n = 1:Nmin - 1)) = 0.12922049010314218
1 - sum((binomial(n - 1, k - 1) * p ^ k * (1 - p) ^ (n - k) for n = 1:Nmin - 1)) = 0.02734419238495378
```

In [46]:

```
1 p = 0.05
2 k = 5
3 @eval @show negbin = NegativeBinomial($k, $p)
4 @show mean(negbin) + k
5 @show var(negbin)
6 @show std(negbin)
7 for Nmin in (150, 200)
8     @show ccdf(negbin, (Nmin - 1) - k)
9 end
```

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

Monte Carlo 法による近似計算:

In [47]:

```
1 M = rand(negbin, 10^8) .+ k
2 @show mean(N)
3 @show var(N)
4 @show std(N)
5 @show mean(>=150, N)
6 @show mean(>=200, N);
```

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

4.30 多項分布の定義

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

次のカテゴリカル分布を考える:

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

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

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

になる。この確率質量関数によって定義された有限集合

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

上の多変量離散分布を多項分布と呼び、

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

と表す。 $r = 2$ の場合の多項分布は本質的に二項分布に等しい。

5 連続分布など

5.1 連続分布と確率密度関数

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

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

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

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

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

確率分布を単に分布と呼ぶことがよくある.

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

5.2 一様分布の定義

区間 (a, b) 上の確率密度関数

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

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

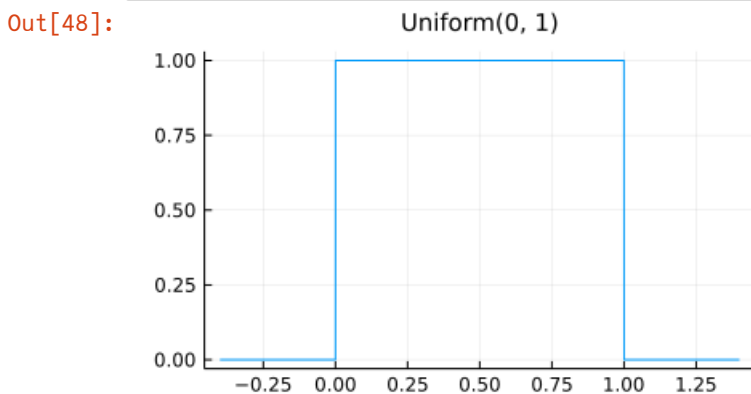
$$\text{Uniform}(a, b).$$

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

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

一様分布 $\text{Uniform}(0, 1)$ の確率密度関数のグラフは以下ようになる.

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



5.3 正規分布の定義

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

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

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

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

と表す. パラメータ σ は標準偏差と呼ばれる.

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

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

と表すことにする.

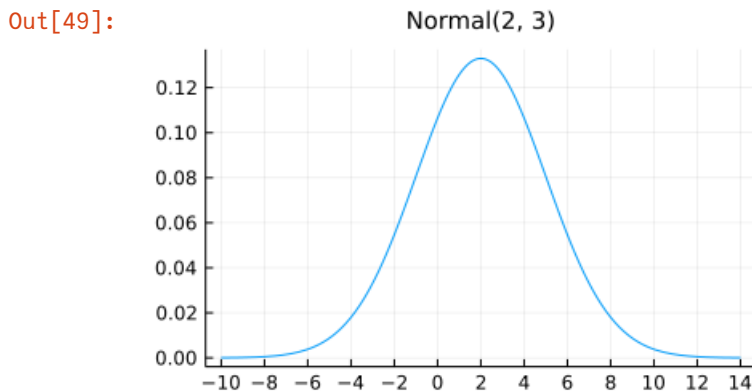
分布のパラメータとしての平均 μ や分散 σ^2 と後で出て来る標本の平均や分散を概念的に厳密に区別する必要がある.

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

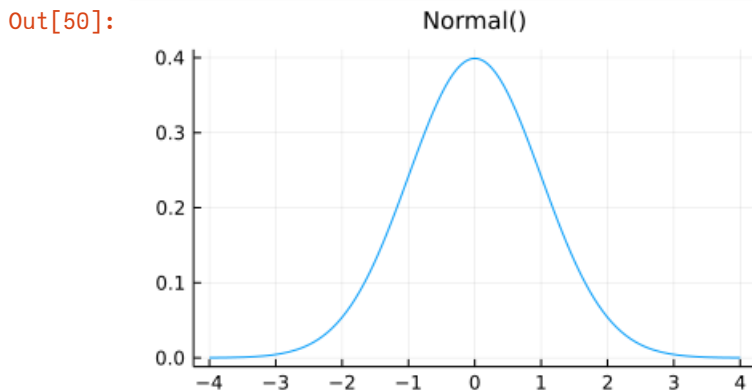
混乱せずに済ますためには「平均」や「分散」という用語を使う場合には必ず「 $\circ\circ$ 分布の平均と分散」や「確率変数 $\circ\circ$ の平均と分散」や「標本 $\circ\circ$ の平均と分散」のように「~の平均と分散**」という言い方をすればよいだろう.**

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

```
In [49]: 1 plot(Normal(2, 3), 2-4*3, 2+4*3; label="", title="Normal(2, 3)", xtick=-100:2:100)
```



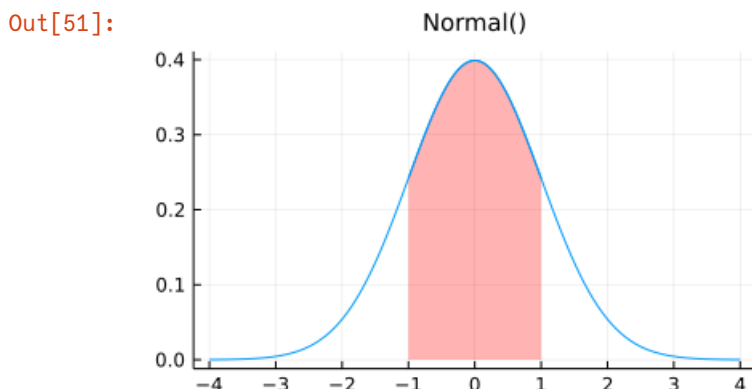
```
In [50]: 1 plot(Normal(), -4, 4; label="", title="Normal()", xtick=-100:100)
```



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

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

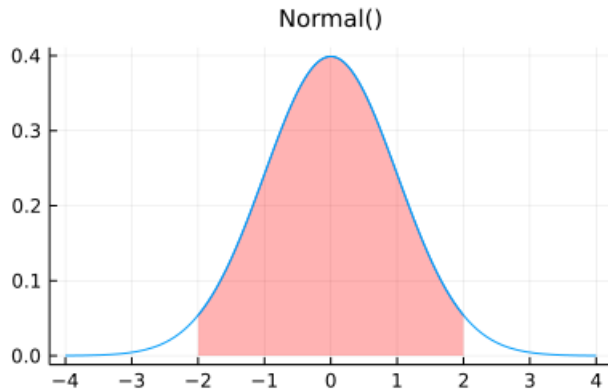
`cdf(Normal(), 1) - cdf(Normal(), -1) = 0.6826894921370861`



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

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

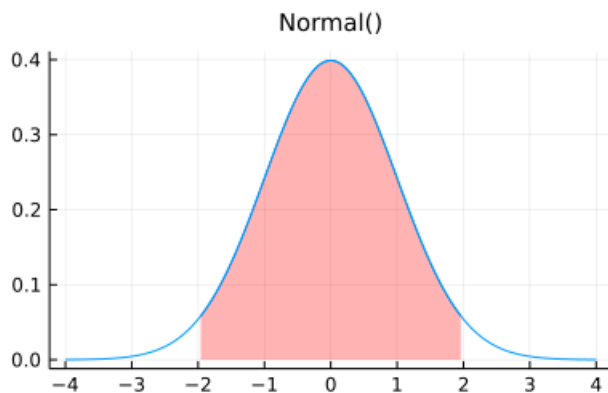
Out[52]:



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

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

Out[53]:



5.4 連続確率変数

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

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

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

$$X \sim p(x)$$

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

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

5.5 連続確率変数に関する確率とその期待値

値が $a \leq c \leq d \leq b$ であると仮定する。

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

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

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

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

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

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

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

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

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

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

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

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

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

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

確率変数 $1_{c \leq f(X) \leq d}(X)$ を

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

と定めると,

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

このように期待値の概念は確率の概念を含んでいると考えてよい.

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

5.6 確率変数の確率密度関数

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

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

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

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

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

ゆえに $Y = X^2$ は確率密度関数

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

を持つ.

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

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

ゆえに $Y = Z^2$ は確率密度関数

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

を持つ. この確率密度関数で定義される連続分布は **自由度1の χ^2 分布** (カイ二乗分布)と呼ばれている.

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

5.7 確率変数の確率質量関数

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

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

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

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

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

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

によって定義する. このとき,

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

ゆえに $X = f(T)$ は確率質量関数

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

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

5.8 連続分布における確率変数の期待値を取る操作の基本性質

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

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

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

5.9 問題: Markovの不等式とChebyshevの不等式

X は確率密度関数または確率質量関数を持つ確率変数であるとする.

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

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

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

解答例: 確率変数 $1_{|Y| \geq t}(X)$ を

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

と定めると,

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

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

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

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

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

が自明に成立する。ゆえに期待値を取る操作の単調性と線形性より、

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

両辺を t^a で割れば示したい不等式が得られる。

解答終

5.10 問題: Jensenの不等式

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

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

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

これらを **Jensenの不等式** (イエンセンの不等式)と呼ぶ。

解答例: $f(x)$ は上に凸な関数であるとし、 $\mu = E[X]$ とおく。ある一次関数 $y = a(x - \mu) + f(\mu)$ が存在して

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

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

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

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

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

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

解答終

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

5.11 問題: 相加相乗平均の不等式

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

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

解答例: $E[f(X)]$ を

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

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

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

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

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

$\log x$ 単調増加関数なので、 $(a_1 \cdots a_n)^{1/n} \leq (a_1 + \cdots + a_n)/n$ が得られる。

解答終

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

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

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

と定める. このとき

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

となることを示せ. この不等式を **Gibbsの情報不等式** と呼ぶ.

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

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

下の凸な関数 $f(x)$ に関するJensenの不等式より,

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

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

解答終

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

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

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

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

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

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

5.13 連続分布における確率変数の分散

連続確率変数 X について, 確率変数 $Y = f(X)$ の期待値を

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

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

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

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

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

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

5.14 問題: 正規分布モデルによる他の分布の最良近似

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

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

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

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

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

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

は次の形になることを示せ:

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

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

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

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

ゆえに μ_0 と σ_0^2 の定義を使うと,

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

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

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

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

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

と増減表を書ける.

解答終

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

5.15 確率変数の一次関数による変換(アフィン変換)の期待値と分散

離散および連続な確率変数 Y と定数 a, b について,

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

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

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

5.16 分布のアフィン変換, スケール変換, 平行移動, 逆数の定義

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

$$aD + b$$

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

さらに, $1/Y$ が従う分布を

$$1/D$$

と書き, D の **逆数** (inverse) と呼ぶことにする.

5.17 確率変数の累積分布関数

確率変数 Y に対して

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

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

Y が連続確率変数で確率密度関数 $p_Y(y)$ を持つとき,

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

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

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

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

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

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

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

5.18 分位点関数 (quantile function)

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

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

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

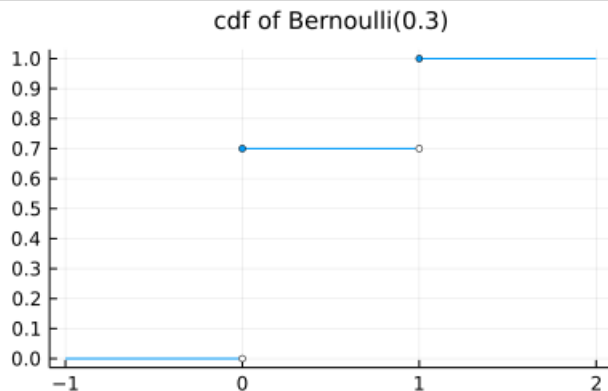
5.19 Bernoulli分布の累積分布関数

Bernoulli分布 $\text{Bernoulli}(p)$ の累積分布関数は次になる:

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

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

Out[54]:



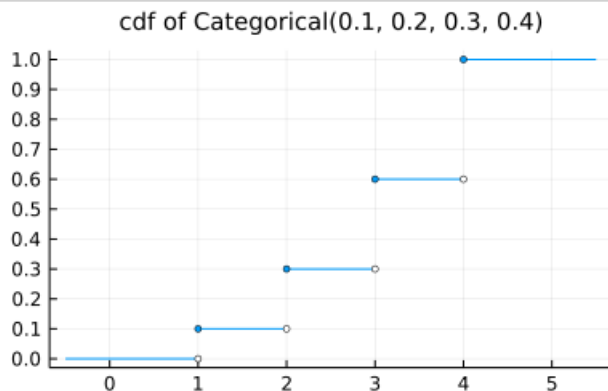
5.20 カテゴリカル分布の累積分布関数

カテゴリカル分布 $\text{Categorical}(0.1, 0.2, 0.3, 0.4)$ の累積分布関数は次になる:

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

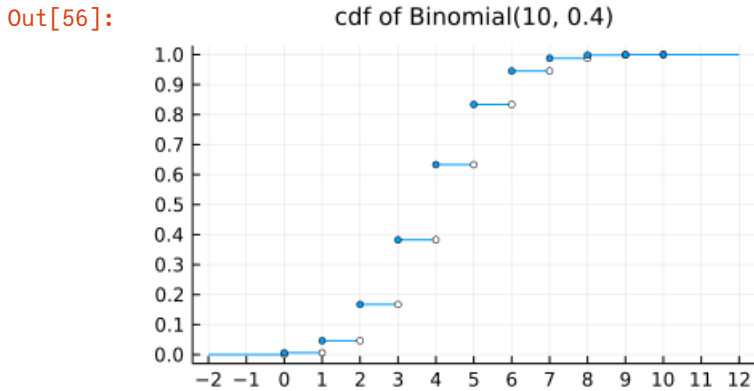
```
In [55]: 1 categ = Categorical(0.1, 0.2, 0.3, 0.4)
2 x = vcat([-0.5:0.02:0.98; NaN], ([k:0.02:k+0.98; NaN] for k in 1:3)..., 4:0.02:5.5)
3 plot(x, x → cdf(categ, x); label="")
4 title!("cdf of Categorical(0.1, 0.2, 0.3, 0.4)")
5 plot!(); xtick = -1:6, ytick = 0:0.1:1
6 scatter!(1:4, x → cdf(categ,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

Out[55]:



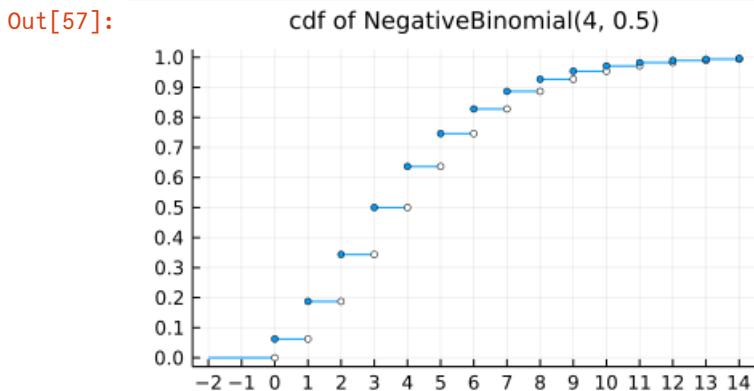
5.21 二項分布の累積分布関数のグラフ

```
In [56]: 1 bin = Binomial(10, 0.4)
2 x = vcat([-2:0.02:-0.02; NaN], ([k:0.02:k+0.98; NaN] for k in 0:9)..., 10:0.02:12)
3 plot(x, x → cdf(bin, x); label="")
4 title!("cdf of Binomial(10, 0.4)")
5 plot!(; xtick = -2:12, ytick = 0:0.1:1)
6 scatter!(0:10, x → cdf(bin, x - 0.02) ; label="", c=:white, ms=3)
7 scatter!(0:10, x → cdf(bin, x) ; label="", c=1, ms=3)
```



5.22 負の二項分布の累積分布関数のグラフ

```
In [57]: 1 negbin = NegativeBinomial(4, 0.5)
2 x = vcat([-2:0.02:-0.02; NaN], ([k:0.02:k+0.98; NaN] for k in 0:13)...)
3 plot(x, x → cdf(negbin, x); label="")
4 title!("cdf of NegativeBinomial(4, 0.5)")
5 plot!(; xtick = -2:14, ytick = 0:0.1:1)
6 scatter!(0:14, x → cdf(negbin, x - 0.02) ; label="", c=:white, ms=3)
7 scatter!(0:14, x → cdf(negbin, x) ; label="", c=1, ms=3)
```



5.23 一様分布の累積分布関数

$a < b$ のとき、一様分布 $\text{Uniform}(a, b)$ の確率密度関数は

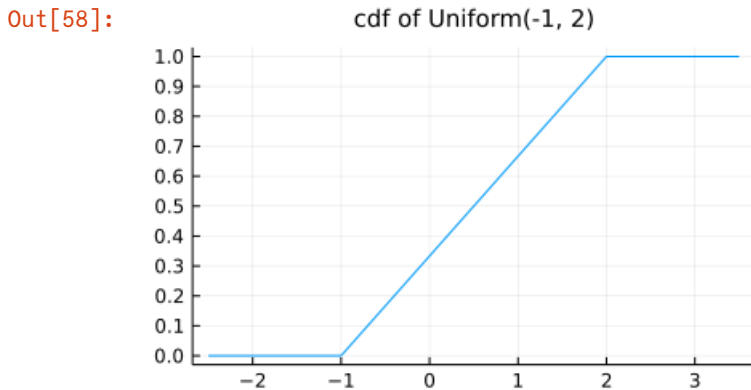
$$p(x) = \begin{cases} 1/(b-a) & (a < x < b) \\ 0 & (\text{otherwise}) \end{cases}$$

である。ゆえに、一様分布 $\text{Uniform}(a, b)$ の累積分布関数は

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

になる。


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



5.24 標準正規分布の累積分布関数と分位点関数

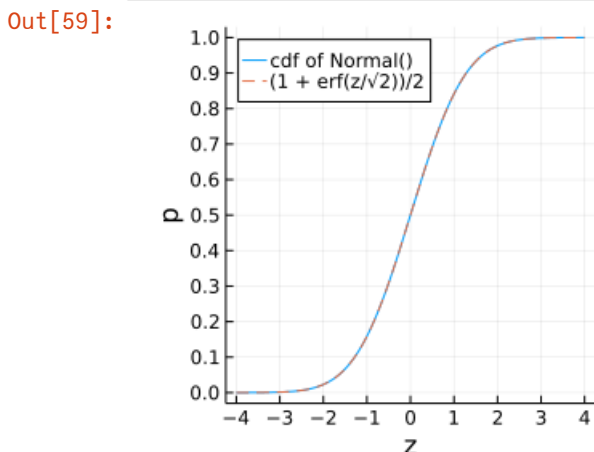
誤差関数が

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

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

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

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



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

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

と書ける. 標準正規分布の分位点関数は統計学で非常によく使われる. 例えば

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

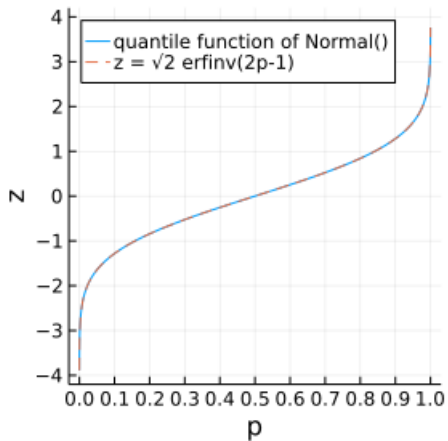
は非常に有名な数値である. これは標準正規分布において値が 1.96 以上になる確率が 2.5% になることを意味している.

```
In [60]: 1 quantile(Normal(), 0.975)
```

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

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

```
Out[61]:
```



5.25 問題: 標準正規分布の四分位数

標準正規分布の累積分布関数と分位点関数はそれぞれ

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

$$z = Q(p) = F^{-1}(p) = \sqrt{2} \operatorname{erfinv}(2p - 1)$$

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

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

```
In [62]: 1 quantile(Normal(), 0.25), quantile(Normal(), 0.75)
```

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

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

- $\sqrt{2} \operatorname{erfinv}(2 \cdot 0.25 - 1)$ (<https://www.wolframalpha.com/input?i=%E2%88%9A2+erfinv%282+0.25+-+1%29>)
- $\sqrt{2} \operatorname{erfinv}(2 \cdot 0.75 - 1)$ (<https://www.wolframalpha.com/input?i=%E2%88%9A2+erfinv%282+0.75+-+1%29>)

WolframAlpha にはこう聞いても答えを教えてくれる:

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

5.26 問題: 連続分布の累積分布関数は一様分布に従う.

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

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

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

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

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

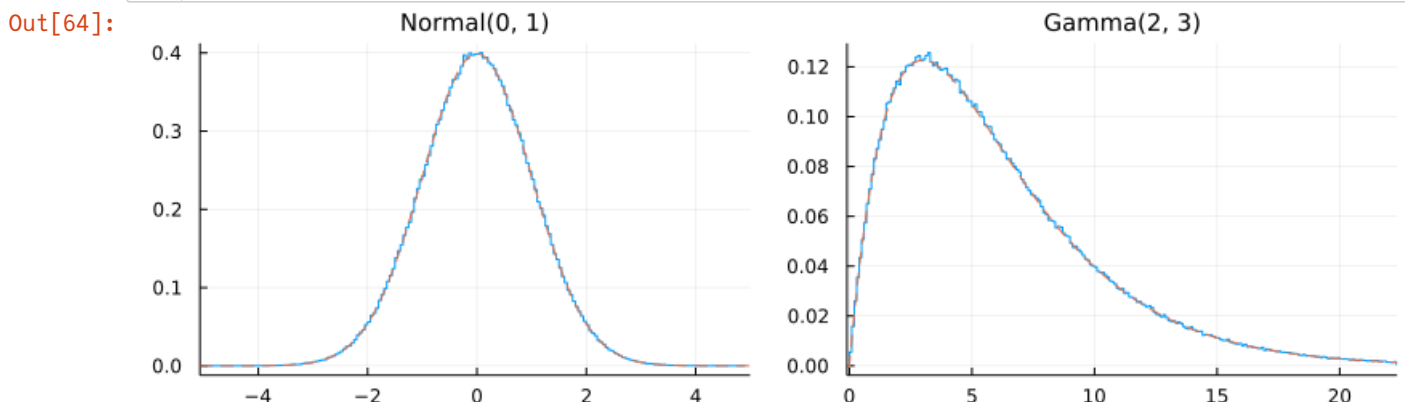
解答終

5.27 分位点函数への一様乱数の代入で元の分布の乱数を作れること

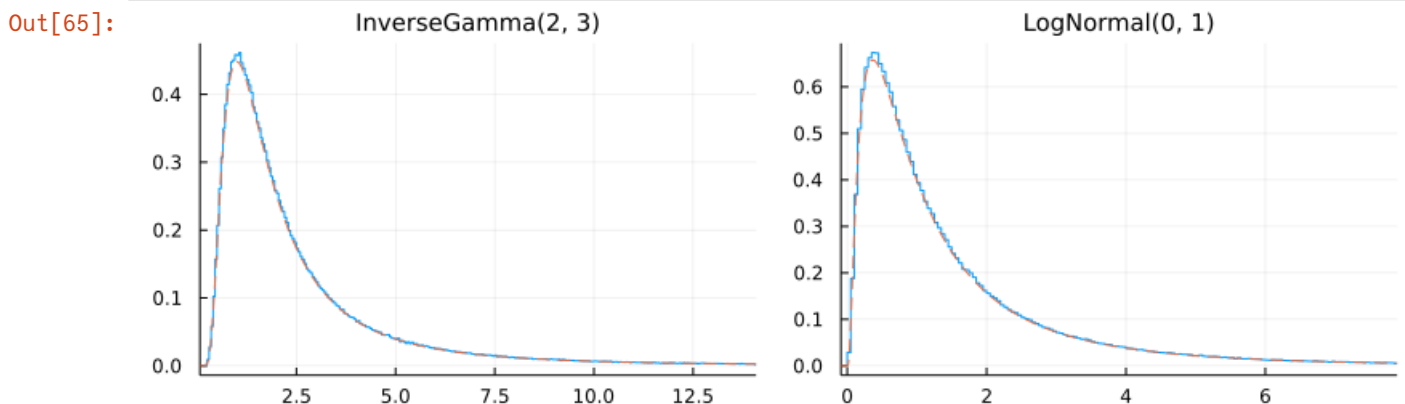
```
In [63]: 1 function plot_quantile_uniform(dist; L=10^6, tlim=(0, 1), title="", kwargs...)
2         T = rand(L) # 一様乱数を大量に生成
3         X = quantile.(dist, T) # 分位点函数で一様乱数を変換
4         xlim = quantile.(Ref(X), tlim) .+ (-0.1, 0.1)
5         X = X[first(xlim) .< X .< last(xlim)]
6         if dist isa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

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

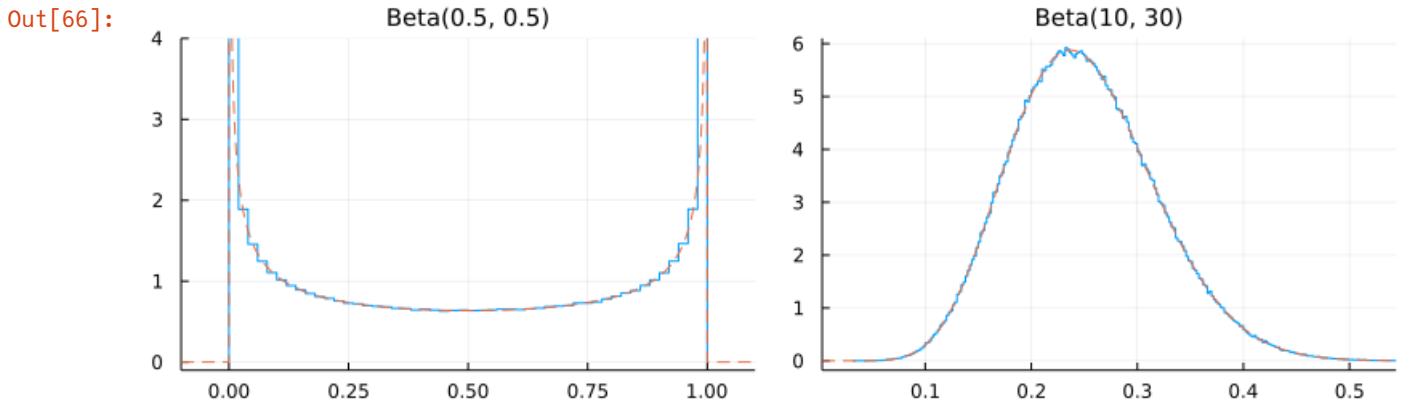
```
In [64]: 1 plot(plot_quantile_uniform(Normal(); title="Normal(0, 1)"),
2           plot_quantile_uniform(Gamma(2, 3); title="Gamma(2, 3)", tlim=(0, 0.995));
3           size=(800, 250))
```



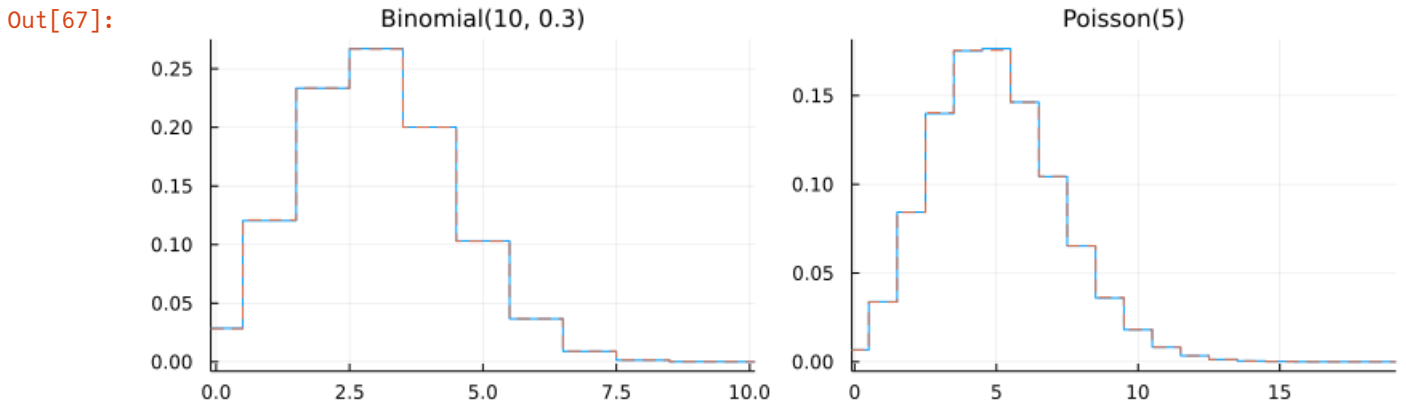
```
In [65]: 1 plot(plot_quantile_uniform(InverseGamma(2, 3); title="InverseGamma(2, 3)", tlim=(0, 0.98)),
2           plot_quantile_uniform(LogNormal(0, 1); title="LogNormal(0, 1)", tlim=(0, 0.98));
3           size=(800, 250))
```



```
In [66]: 1 plot(plot_quantile_uniform(Beta(0.5, 0.5); title="Beta(0.5, 0.5)", xlim=(-0.1, 1.1), ylim=(-0.
2         plot_quantile_uniform(Beta(10, 30); title="Beta(10, 30)", xlim=(0.005, 0.995));
3         size=(800, 250))
```



```
In [67]: 1 plot(plot_quantile_uniform(Binomial(10, 0.3), title="Binomial(10, 0.3)"),
2         plot_quantile_uniform(Poisson(5); title="Poisson(5)");
3         size=(800, 250))
```



5.28 一様分布の期待値と分散

区間 (a, b) 上の確率密度関数

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

によって与えられる一様分布

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

を考え、 X はこれに従う確率変数であるとする。

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

$$\begin{aligned} \mu &= E[X] = \int_a^b x p(x) dx = \frac{1}{b-a} \int_a^b x dx = \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}. \\ E[X^2] &= \int_a^b x^2 p(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{b^3 - a^3}{3(b-a)} = \frac{a^2 + ab + b^2}{3}. \\ \text{var}(X) &= E[X^2] - E[X]^2 = \frac{4(a^2 + ab + b^2) - 3(a+b)^2}{12} = \frac{a^2 - 2ab + b^2}{12} = \frac{(b-a)^2}{12}. \end{aligned}$$

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

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

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

Out[69]: (1//0, 1//12)

```
In [70]: 1 E(x → x, Uniform(0, 1)), E(x → x^2, Uniform(0, 1)) - E(x → x, Uniform(0, 1))^2
```

Out[70]: (0.5, 0.08333333333333331)

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

Out[71]: (5//2, 25//12)

```
In [72]: 1 E(x → x, Uniform(0, 5)), E(x → x^2, Uniform(0, 5)) - E(x → x, Uniform(0, 5))^2
```

Out[72]: (2.5, 2.08333333333333357)

5.29 標準正規分布の期待値と分散

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

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

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

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

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

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

ここで, 5番目の等号では次の公式を使った:

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

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

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

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

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

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

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

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

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

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

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

Out[73]: (0, 1)

```
In [74]: 1 E(x → x, Normal(0, 1)), E(x → x^2, Normal(0, 1)) - E(x → x, Normal(0, 1))^2
```

Out[74]: (4.2013717612707503e-16, 1.0)

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

Out[75]: (2, 9)

```
In [76]: 1 E(x → x, Normal(2, 3)), E(x → x^2, Normal(2, 3)) - E(x → x, Normal(2, 3))^2
```

Out[76]: (1.9999999999999999, 9.0)

標準正規分布の歪度(わいど, skewness):

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

Out[77]: 0

```
In [78]: 1 E(z → z^3, Normal())
```

Out[78]: -5.974359421244178e-16

標準正規分布の尖度(せんど, kurtosis)に 3 を足した値:

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

Out[79]: 3

```
In [80]: 1 E(z → z^4, Normal())
```

Out[80]: 2.9999999999999997

質問: 尖度の定義でどうして 3 を引いておくか?

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

```
In [81]: 1 @syms t κ3 κ4 κ5 μ3 μ4 μ5
2 series(exp(t^2//2 + κ3 * t^3//6 + κ4 * t^4//24 + κ5 * t^5//120), t)
```

Out[81]: $1 + \frac{t^2}{2} + \frac{t^3 \kappa_3}{6} + t^4 \left(\frac{\kappa_4}{24} + \frac{1}{8} \right) + t^5 \left(\frac{\kappa_3}{12} + \frac{\kappa_5}{120} \right) + O(t^6)$

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

Out[82]: $\frac{t^2}{2} + \frac{t^3 \mu_3}{6} + t^4 \left(\frac{\mu_4}{24} - \frac{1}{8} \right) + t^5 \left(-\frac{\mu_3}{12} + \frac{\mu_5}{120} \right) + O(t^6)$

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

```
In [83]: 1 [(prod(1:2:2k-1), E(z → z^(2k), Normal())) for k in 1:7]
```

Out[83]: 7-element Vector{Tuple{Int64, Float64}}:
(1, 1.0)
(3, 2.9999999999999997)
(15, 14.999999999999998)
(105, 105.00000000000004)
(945, 945.00000000000093)
(10395, 10395.0000000000617)
(135135, 135135.0000000145)

5.30 一様乱数の生成の繰り返しの分布

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

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

の積で次のように定める:

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

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

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

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

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

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

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

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

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

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

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

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

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

In []:

1