確率分布のプロット
——数式処理ソフトの利用

□□□□□ □□□□□ □□□□□ □□□□□ □□□□□ □□□□□　　2010 年 7 月 21 日版

目次

1　一様分布の和

確率変数 $X_1, X_2, \cdots$, は独立で同じ一様分布 $U(0,1)$ に従うとする。$S_n = X_1 + X_2 + \cdots + X_n$, $(n \geq 1)$ とし、分布関数 $U_n(x) = P(S_n \leq x)$, 密度関数 $u_n(x) = \dfrac{dU_n}{dx}(x)$ を考える。確率変数の和を求める式からつぎの関係式が得られる。

定理 1.1　*(i)* 単位区間 $(0,1)$ 上の一様分布の密度関数：

$$u_1(x) = 1,\ if \quad 0 < x < 1$$

(ii) 和の公式による密度関数の再帰関係式と分布関数からの計算

$$u_{n+1}(x) = \int_0^1 u_n(x-y)dy = U_n(x) - U_{n+1}(x-1),\ n = 1, 2, \cdots$$

(iii) 一様分布の分布関数：

$$U_1(x) = x_+ - (x-1)_+, \quad -\infty < x < \infty \text{ ここで } x_+ = \max\{0, x\} = \frac{x+|x|}{2}$$

(iv) 一般のばあいの分布関数：

$$U_n(x) = \frac{1}{n!}\sum_{\nu=0}^{n}(-1)^\nu \binom{n}{\nu}(x-\nu)_+^n,\ n = 1, 2, \cdots$$

(v) 一般のばあいの密度関数：

$$u_n(x) = \frac{1}{(n-1)!}\sum_{\nu=0}^{n}(-1)^\nu \binom{n}{\nu}(x-\nu)_+^{n-1},\ n = 1, 2, \cdots$$

証明は帰納法にもとづく。とくに 2 項係数の加法性：$\binom{n}{\nu} + \binom{n}{\nu-1} = \binom{n+1}{\nu}$ を用いる。ただし密度の範囲が一般のばあい $u_n(x)$ は $0 < x < n$ であるが、関係式から、$(x-n)_+$ では明らかであるが、他の項からの計算も加わるからそう自明なことではない。平均と分散をもとめて、中心極限定理により、n が大きければ正規分布で近似される。とくに $n = 12$ ならば、簡単な計算により、1 つの標準正規分布にしたがう確率変数が得られる。

定理 1.2

$$X_i \sim U(0,1) \quad \Rightarrow \quad "n = 12",\ S_{12} - 6 \sim N(0,1)$$

なぜなら各 $U(0,1)$ の平均は 1/2, 分散は 1/12 であるから、これを加えた 12 個の S_{12} の平均は 6, 分散は 1 となる。したがって、$S_{12} - 6 \sim N(0,1)$ である。

では具体的に分布を計算して、さらに図に表すことをおこなう。参考文献：一様分布の和 Feller Vol .2, p .26

$xp[x] = \max\{0, x\}$ の定義：xp[x_] = $(x + \mathrm{Abs}[x])/2$

１２個の和：１２本の１１次曲線 $([0,1], [1,2], \cdots, [11, 12])$

標準正規分布 $N(0,1)$ の近似式: $\sum_{i=1}^{12} X_i - 6$,

12 個を加えた後の 6 を引いたもの（和の式を平行移動）、平均 0、分散 1 となる.　各区間 $[0,1], [1,2], \cdots, [11.12]$ での 11 次多項式を 12 本描いたグラフである。

```
usma12[x_] = Sum[(−1)∧{k} ∗ Binomial[12, k] ∗ (xp[x − k])∧{12 − 1}, {k, 0, 12}]/(12 − 1)!
Plot[usma12[x + 6], {x, −4, 4}]
```

２つの図 5 と図 4 を比べてみると、見かけのグラフではほとんど差が分からないが、相対誤差の計算 $\frac{\text{近似値} - \text{真値}}{\text{真値}} * 100\%$ を求めてみると以下のようになる。現代のパソコンで簡単に正規分布のシミュレー

(1)　2 個の和：（三角形分布）

```
usma2[x_] =
Sum[(−1){k} ∗ Binomial[2, k]∗
(xp[x − k])^{2−1}, {k, 0, 2}]/(2 − 1)!
Plot[usma2[x],{x,0,2}]
```

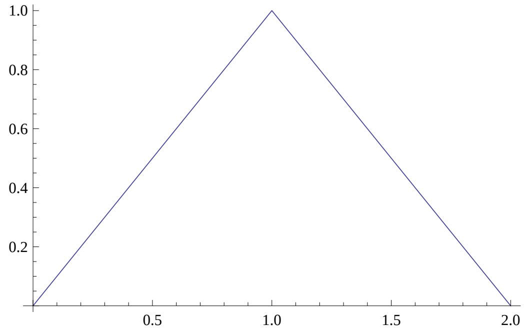


図 1　三角形分布

2 本の一次式：$(0 < x < 1, 1 < x < 2)$

(2)　3 個の和：

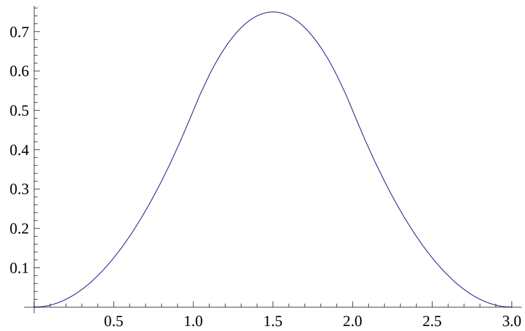


図 2　3 本の 2 次曲線の分布 $([0, 1], [1, 2], [2, 3])$

(3)　5 個の和：

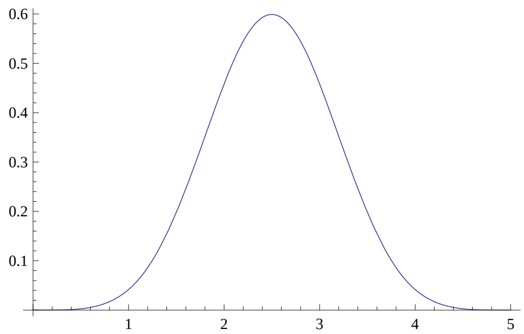


図 3　5 本の 4 次曲線の分布 $([0, 1], [1, 2], \cdots, [4, 5])$

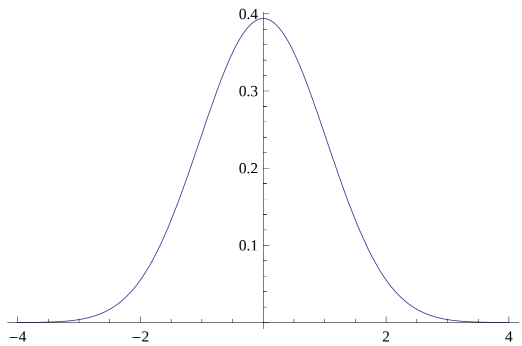


図 4　12 個の和による近似分布

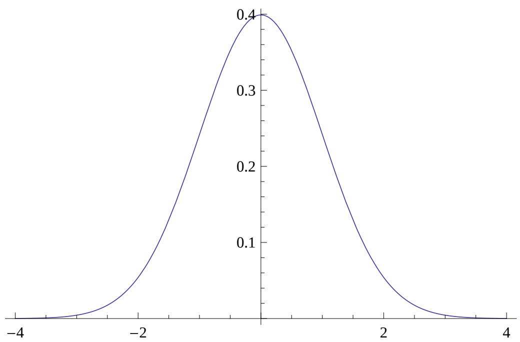


図 5　標準正規分布 $N(0, 1)$

ションを行うことはそう難しいことではないが、至極簡単な一様乱数の 12 個から 1 個の正規乱数をつくることができる。

問 1.1 *2* 個のとき,*3* 個のときのグラフは

$$\frac{1}{2}\{(x-2)_+ - 2(x-1)_+ + x_+\}, \quad \frac{1}{2!}\{-(x-3)_+^2 + 3(x-2)_+^2 - 3(x-1)_+^2 + x_+^2\}$$

となることを示し、*3* 個では滑らかであることを確かめよ。

x	Approx	TRUE
0.0	0.3939	0.3989
0.2	0.3865	0.3910
0.4	0.3650	0.3683
0.6	0.3318	0.3332
0.8	0.2901	0.2897
1.0	0.2440	0.2420
1.2	0.1971	0.1942
1.4	0.1529	0.1497
1.6	0.1137	0.1109
1.8	0.0810	0.0790
2.0	0.0552	0.0540
2.2	0.0359	0.0355
2.4	0.0222	0.0224
2.6	0.0131	0.0136
2.8	0.0073	0.00792
3.0	0.0038	0.0044

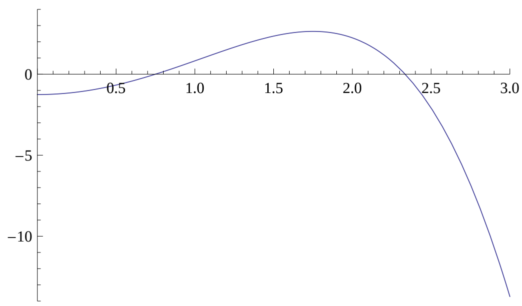


図 6　真値との相対誤差（近似ー真値）／真値%

2　プロット命令

密度関数の関数形を表示させる命令：

PDF[{ 分布名 }[パラメータ], 変数], { 変数、最小値、最大値 }

グラフを描く命令：

Plot[PDF[{ 分布名 }[パラメータ], 変数], { 変数、最小値、最大値 }]

複数のグラフを書くときには PDF[...] の部分を { } で密度関数を並べて括る：

{PDF[{ 分布名 1}[パラメータ], 変数], $\cdots$, PDF[{ 分布名 n}[パラメータ], 変数]}

という形にする。

密度関数の関数形を表示させる命令：	PDF[分布名 [パラメータ], 変数], { 変数、最小値、最大値 }
グラフを描く命令：	Plot[PDF[分布名 [パラメータ], 変数], { 変数、最小値、最大値 }]

3　分布の密度関数

(1) **離散一様分布:** DiscreteUniformDistribution[n_{min},n_{max}]

一様乱数を生成するために用いられる。サイコロのシミュレーションでは $n_{min}=1, n_{max}=6$。同じ確からしさをもつ事象の確率（等しい確率）で起る結果を n_{min} から n_{max} までの整数で表した試行である．

$$\mathrm{U}\{n_{min},\cdots,n_{max}\}\ \sim\ \frac{1}{n_{max}-n_{min}+1},\ x=n_{min},\cdots,n_{max}\quad (\text{一定値})$$

(2) **ベルヌーイ分布:**BernoulliDistribution[p]

成功（値 1 に対応）が p の確率で起り，失敗（値 0 に対応）が $1-p$ の確率で起る 1 回 (単一) 試行の確率分布である． $\binom{n}{k}=\frac{(n)_k}{k!}=\frac{(n)_k=n(n-1)\cdots(n-k+1)}{k!}$ とする。2 項分布 Binom で $n=1$ とした場合。

$$\mathrm{Binom}[1,p]\ \sim\ \binom{1}{x}p^x(1-p)^{1-x}=\begin{cases} p & x=1\\ 1-p & x=0\end{cases}$$

(3) **二項分布:** BinomialDistribution[n,p]

各試行における成功確率を p としたときに，n 回の独立した試行で起る成功を集計した回数値の分布である．

$$\mathrm{Binom}[n,p]\ \sim\ \binom{n}{x}p^x(1-p)^{n-x},x=0,1,2,\cdots,n$$

定理 3.1 *(i)*　n 個の独立なベルヌーイ分布からつくる和の分布は、*2* 項分布にしたがう。

$$X_i\ \sim\ Binom[1,p]\quad\Rightarrow\quad\sum_{i=1}^{n}X_i\ \sim\ Binom[n,p]$$

(ii)　2 個の独立な同じパラメータ p をもつ *2* 項分布からつくる和の分布は、*2* 項分布にしたがう。

$$X\ \sim\ Binom[n,p],\,Y\ \sim\ Binom[m,p]\quad\Rightarrow\quad X+Y\ \sim\ Binom[n+m,p]$$

(4) **幾何分布:** GeometricDistribution[p]

各試行の成功率を p としたときに，最初に成功が起るまでに必要とした試行数の分布である．$q=1-p$ とする。成功／失敗という 2 通りの結果を繰り返したとき、初めて成功が表れるまでの時間間隔を表す。待ち時間の解析によく用いられ、連続型では指数分布と対比される。

$$\mathrm{Geom}(p)\ \sim\ pq^x,\,x=0,1,2,\cdots$$

定理 3.2 n 個の独立な幾何分布からつくる和の分布は、負の *2* 項分布にしたがう。

$$X_i\ \sim\ Geom[p]\quad\Rightarrow\quad\sum_{i=1}^{n}X_i\ \sim\ NegBinom[n,p]$$

(5) **負の二項分布:** NegativeBinomialDistribution[n,p]

正の整数 n に対する負の二項分布とは，各試行における成功確率を p としたときに，n 回の成功するまでに必要なために起った失敗（$q=1-p$ の確率）回数の分布である．はじめて成功が起こるまでの待ち時間が幾何分布であるから、さらに n 回まで成功するまでに要する間隔の回数値分布である。つ

まり幾何分布の和と考えればよい。$n=1$ が幾何分布に他ならない。最終回以前は回数を指定しないから、2 項係数が表れる。また 2 項係数に負の値があり、$\binom{-n}{x} = \frac{(-n)_x}{x!} = \frac{(-n)(-n-1)\cdots(-n-x+1)}{x!} = (-1)^x\frac{(n+x-1)_x}{x!} = (-1)^x\binom{n+x-1}{x}$ となることに注意。

$$\mathrm{NegBinom}[n,p] \ \sim \ \binom{-n}{x}p^n(-q)^x = \binom{n+x-1}{x}p^nq^x, x=0,1,2,\cdots,$$

(6) **ベータ二項分布:** BetaBinomialDistribution[α,β,n]
二項分布とベータ分布を組み合せた混合分布である．この確率変数は，成功確率 p 自体を確率変数として、その分布がベータ分布 BetaDistribution[α,β] に従うとき、この混合分布を BinomialDistribution[n,p] に従うという．$B(\alpha,\beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}dt \quad (\alpha,\beta>0)$ はベータ関数。指数部は $0<\alpha<1, \alpha=1, 1<\alpha$ によって大きく変化する。

$$\mathrm{BetaBinom}[\alpha,\beta,n] \ \sim \ \binom{n}{x}\frac{B(x+\alpha,n+\beta-x)}{B(\alpha,\beta)}, x=0,1,2,\cdots,n$$

(7) **ベータ負の二項分布:** BetaNegativeBinomialDistribution[α,β,n]
ベータ分布と負の二項分布を組み合せた混合分布である．p の値が既知のものとするのではなく、変動するものと考える。このようにパラメータ p をベータ分布とする理由はベータ分布が柔軟で、一様分布を含むさまざまな形状を表すから。

$$\mathrm{BetaNegBinom}(\alpha,\beta,n) \ \sim \ \binom{-1+n+x}{-1+n}\frac{B[\alpha+n,\beta+x]}{B[\alpha,\beta]}, \ x=0,1,2,\cdots,n$$

(8) **超幾何分布:** HypergeometricDistribution[n,n_{succ},n_{tot}]
n 回中のの { 当たり, 外れ (または成功、失敗)} 試行での当たり回数であるが、取り出しの結果によって当たり、外れの確率が変化する。母集団サイズ（総数の個数）は n_{tot} として、最初は n_{tot} 枚数のくじがあり、その内の当たりくじ数 (成功とする回数)n_{succ} 回と外れくじ数（失敗とする回数）$n_{fail}=n_{tot}-n_{succ}$ で構成されている。一度取り出したくじは破棄して 2 度と取り出されることはないものとする。すなわち、元に戻さずにサンプリングする（非復元抽出)。はじめには、総数 n_{tot} のくじが当たり数 n_{succ} と外れ数 $n_{fail}=n_{tot}-n_{succ}$, "tot = total"であり、抽出ごとに何れかが減じていく非復元抽出実験を行う。超幾何分布とは、くじ引きを試みた回数 n 回のうちで当たりくじを引いた回数を表す分布である。このくり返し試行は独立ではないことに注意。二項分布はベルヌーイ分布のくり返し試行で独立であるととに違いがある。当たりくじを引く確率（当たりくじと外れくじの比率）が一定（復元抽出）の場合である。

$$\mathrm{HyperGeom}[n,n_{succ},n_{tot}] \ \sim \ \frac{\binom{n_{succ}}{x}\binom{n_{fail}}{n-x}}{\binom{n_{tot}}{n}} = \binom{n}{x}\frac{(n_{succ})_x\,(n_{fail})_{n-x}}{(n_{tot})_n}, \ x=0,1,2,\cdots,n$$

定理 3.3 超幾何分布のパラメータ $n_{tot}=n_{succ}+n_{fail}, (n_{tot}\to\infty)$ において

$$\frac{n_{succ}}{n_{tot}} \to p, \frac{n_{fail}}{n_{tot}} \to q=1-p \ \Rightarrow \ HyperGeom[n,n_{succ},n_{tot}] \doteqdot Binom[n,p]$$

いいかえると、母集団サイズが大きくなれば、復元と非復元の違いは少なくなる。「石川や浜の真砂は尽くるとも世に盗人の種は尽きまじ」*(石川五右衛門)*

たとえば、スピード三角くじをひく。一度引けばそれは 2 度と引かれることはない。このとき 10 本のくじがあり，そのうち 4 本が当たりで残り 6 本が外れである．いま 3 本くじを引いて，そのうち 2 本が当たる確率は？という問題は超幾何分布である。$n_{tot} = 10, n_{succ} = 4, n_{fail} = 6, n = 3, x = 2$ として計算する。

(9) **ポアソン分布:** PoissonDistribution[λ]
指定された時間内に起る事象数を記述する．ここで $1/\lambda$ は一定時間に対する事象の平均回数 (整数とは限らない) である．
2 項分布 Binom(n,p_n) の極限として得られる、稀な現象を記述するときに用いられ、重要な分布である。稀という意味は 2 項分布での事象が起こる (成功する) 確率 p_n をゼロに近づけるが、試行の繰り返し回数 n を増やすことによる。指数が出てくる理由は $\lim_n(1-p_n)^n = e^{-\lambda}$, ただし、$np_n \to \lambda(n \to \infty)$ で、成功の確率（事象が起こる確率）$p_n \to 0$ としなければばならないから、これが稀ということも表している。

$$Po[\lambda] \;\sim\; e^{-\lambda}\frac{\lambda^x}{x!},\; x = 0, 1, 2, \cdots$$

定理 3.4 *2 項分布 Binom[n,p_n] において，ある定数 $\lambda > 0$ があって $(n \to \infty, p_n \to 0)$ のとき、*

$$np_n \to \lambda \;\Rightarrow\; Binom[n, p] \;\Rightarrow\; Po[\lambda]$$

個数を数えるタイプの現象は近似的にポアソン分布に従うといわれている。たとえば、単位時間にかかってくる電話の回数、活字印刷される本の 1 ページ中の誤植など。有名な例として、ロシアでの馬に蹴られて死んだ兵士の数 (W.Feller vol.I VI 章) がある。

4　パラメータ変化の概形を描く（2 項、幾何、負の 2 項、超幾何）

確率分布にはいくつかのパラメータが含まれている。このパラメータが変化したときには、さまざまな形が表れる場合があり、これらについて調べておくことは重要であり、グラフ描画の機能がよく進化した現代ではより深い洞察も得られるかも知れない。

(1)　幾何分布 Geom(p) の $p = 0.1, 0.3, 0.5, 0.8$

```
ListPlot[Table[{k,
PDF[GeometricDistribution[p], k]},
{p, {0.1, 0.3, 0.5, 0.8}}, {k, 0, 15}],
PlotMarkers → Automatic,
Joined → True,
PlotRange → {{0, 15}, {0, 0.8}}]
```

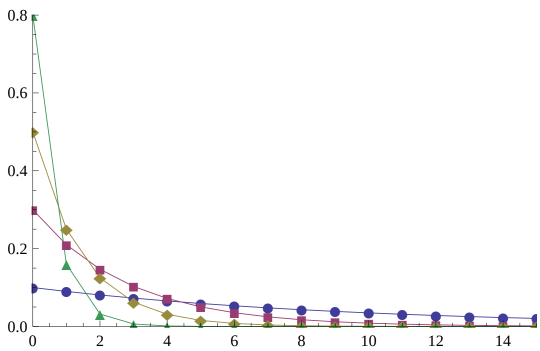


図 7　幾何分布 Geom(p) のパラメータ変化)

(2) 2 項分布 Binom(10,p) の $p = 0.1, 0.3, 0.5, 0.8$

```
ListPlot[Table[{k,
PDF[BinomialDistribution[10, p], k]},
{p, {0.1, 0.3, 0.5, 0.8}}, {k, 0, 10}],
PlotMarkers → Automatic,
Joined → True,
PlotRange → Automatic]
```

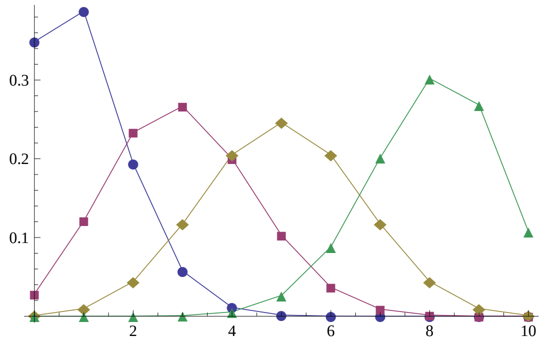


図 8　2 項分布 Binom(10,p) での p 変化

(3) 2 項分布 Binom(n,1/2) の $n = 5, 10, 20$

```
ListPlot[Table[{k,
PDF[BinomialDistribution[n, 1/2], k]},
{n, {5, 10, 20}}, {k, 0, 20}],
PlotMarkers → Automatic,
Joined → True,
PlotRange → Automatic]
```

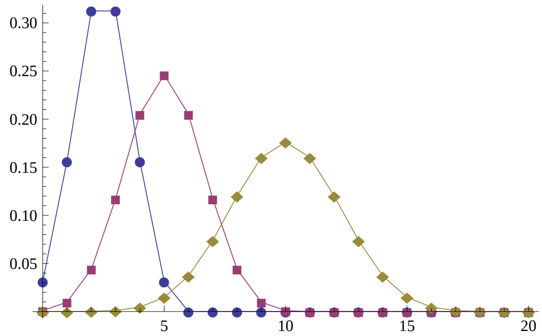


図 9　2 項分布 Binom(n,1/2) での n 変化

(4) 2 項分布 Binom(n,1/4) の $n = 5, 10, 20$

```
ListPlot[Table[{k,
PDF[BinomialDistribution[n, 1/4], k]},
{n, {5, 10, 20}}, {k, 0, 20}],
PlotMarkers → Automatic,
Joined → True,
PlotRange → Automatic]
```

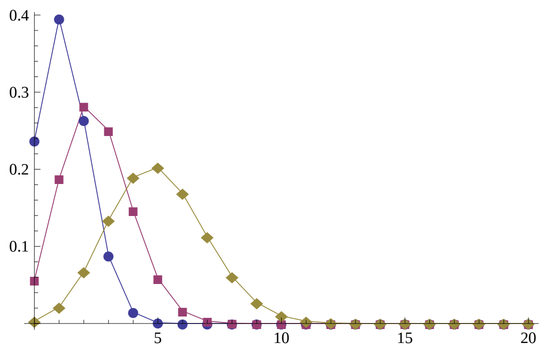


図 10　2 項分布 Binom(n,1/4) での n 変化

(5) 2 項分布 Binom(n,3/4) の $n = 5, 10, 20$

```
ListPlot[Table[{k,
PDF[BinomialDistribution[n, 3/4], k]},
{n, {5, 10, 20}}, {k, 0, 20}],
PlotMarkers → Automatic,
Joined → True,
PlotRange → Automatic]
```

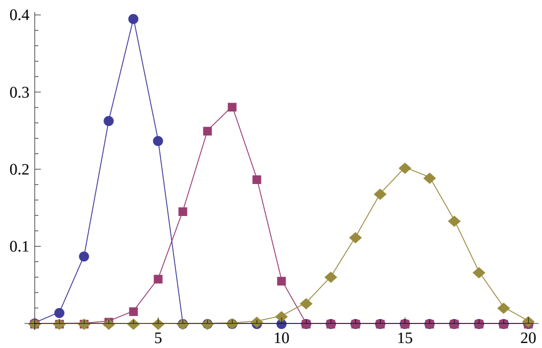


図 11　2 項分布 Binom(n,3/4) での n 変化

(6) 負の 2 項分布 NegBinom(8,p), $p = 0.1, 0.3, 0.5, 0.8$

```
ListPlot[Table[{k,
PDF[NegativeBinomialDistribution[5, p], k]},
{p, {0.1, 0.3, 0.5, 0.8}}, {k, 0, 30}],
PlotMarkers → Automatic,
Joined → True,
PlotRange → {{0, 30}, {0, 0.35}}]
```

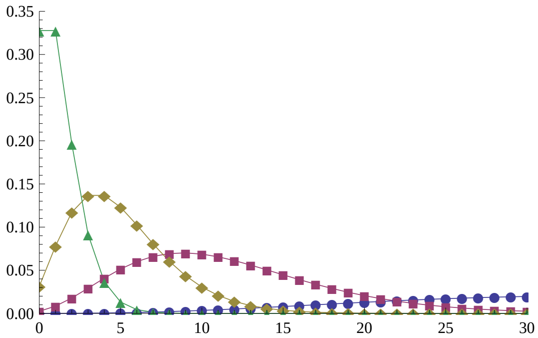


図 12　負の 2 項分布 NegBinom(8,p), p の変化

(7) 負の 2 項分布 NegBinom(n,0.3), $n = 1, 5, 8, 10$

```
ListPlot[Table[{k,
PDF[NegativeBinomialDistribution[n, 0.3], k]},
{n, {1, 5, 8, 10}}, {k, 0, 30}],
PlotMarkers → Automatic,
Joined → True,
PlotRange → {{0, 30}, {0, 0.35}}]
```

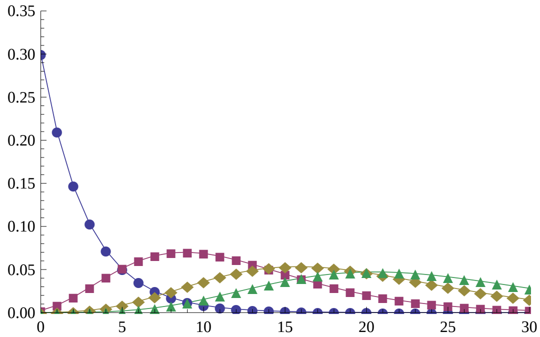


図 13　負の 2 項分布 NegBinom(n,0.3), n の変化

(8) 超幾何分布 HyperGeom(5,n_{succ},n_{tot}) $\dfrac{n_{succ}}{n_{tot}} = \dfrac{2}{10}, \dfrac{5}{10}, \dfrac{7}{10}\ (n < n_{tot})$

```
ListPlot[Table[{k,
PDF[HypergeometricDistribution[5, ns, 10], k]
}, {ns, {2, 5, 7}}, {k, 0, 5}],
PlotMarkers → Automatic,
Joined → True,
PlotRange → Automatic]
```

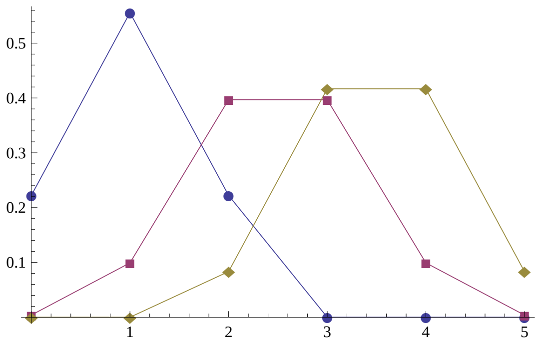


図 14　超幾何分布 HyperGeom(n,n_{succ},n_{tot}), n の変化

5　ベータ分布について

独立に一様分布 U(0, 1) に従う $p + q - 1$ 個の確率変数を大きさの順に並べ替えたとき，小さい方から p 番め（大きい方からは q 番目）の確率変数 X の分布がベータ分布 Beta[p,q] となる。順序統計量を計算することで示される。今日では別の観点で重要視される。ベイズ統計学である。ベータ分布は二項分布の共役事前分布である。少し詳しく述べると、統計学的な（多数のものを扱う）例として、二項分布の母数の事後分布を計算することを考えよう。観察結果が、成功 m 回、失敗 n 回となったとする。具体的には成功・失敗のコイン

(9) 超幾何分布

(HypergeometricDistribution[15,n_{succ},n_{tot}])

$\frac{n_s}{n_t} = \frac{4}{20}, \frac{10}{20}, \frac{14}{20}$

```
ListPlot[Table[{k,
PDF[HypergeometricDistribution[15, ns, 20], k]
}, {ns, {4, 10, 14}}, {k, 0, 15}],
PlotMarkers → Automatic,
Joined → True,
PlotRange → Automatic]
```

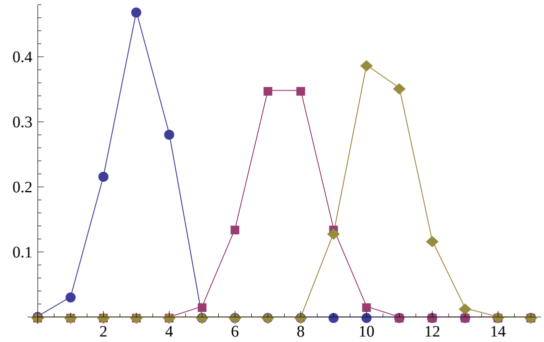


図 15　超幾何分布 Hypergeom[n,n_{succ},n_{tot}], n の変化

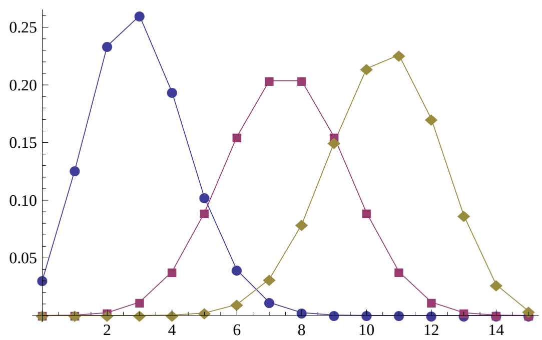


図 16　超幾何分布 Hypergeom[n,n_{succ},n_{tot}], n の変化

(10) 超幾何分布

(HypergeometricDistribution[15,n_{succ},n_{tot}])

$\frac{n_s}{n_t} = \frac{40}{200}, \frac{100}{200}, \frac{140}{200}$

```
ListPlot[Table[{k,
PDF[HypergeometricDistribution[15, ns, 20], k]
}, {ns, {40, 100, 140}}, {k, 0, 15}],
PlotMarkers → Automatic,
Joined → True,
PlotRange → Automatic]
```

(11)　超幾何分布が 2 項分布に近づく ($\frac{n_{succ}}{n_{tot}} \to p$) を比較する。 $n = 15$ での 2 項分布 (BinomialDistribution[15,p,] $p = \frac{n_s}{n_t} = \frac{40}{200}, \frac{100}{200}, \frac{140}{200}$ を描く。

```
ListPlot[Table[{k,
PDF[BinomialDistribution[15, p], k]
}, {ns, {40, 100, 140}}, {k, 0, 15}],
PlotMarkers → Automatic,
Joined → True,
PlotRange → Automatic]
```

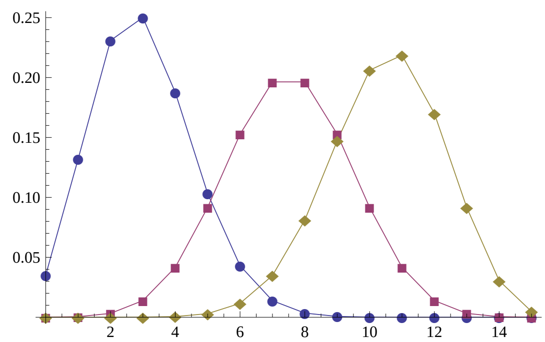


図 17　2 項分布 Binom[n,p

投げとか、提案に対する賛成・反対でもよい。成功確率の母数 p について事前確率で表されるとする。与えられた p の値に対して、全 $m+n$ 回の試行のうち成功が m 回となる確率は、2 項分布 $\binom{n+m}{m}p^m(1-p)^n$, m と n は固定され、a は未知である。ベイズの定理（連続分布の形）から、事後確率の計算をおこなうと、未知の p に対する事前分布が ベータ分布 Beta[m_0,n_0] ならば、ベイズの定理から計算できる事後分布もベータ分布で、Beta[$m+m_0$, $n+n_0$] となる。このベータ分布のように、事後分布が同じタイプの分布になるような事前分布を共役事前分布という。ベイズ推定（ベイズすいてい）とは、ある証拠に基づいて、その原因となった事象を推定するための確率論的方法である。また統計学に応用されてベイズ統計学の代表的な方法となっている。

パラメータの値 α, β は正の実数でよい。極めて広範囲の形状となる分布形。$\alpha=\beta=1$ では連続型一様分布。$0<\alpha<1, \alpha=1, 1<\alpha<\infty$ と変化にするに従い、さまざまな形状となる。β についても同様である。

$$\mathrm{Beta}[\alpha,\beta] \ \sim \ \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}, \ \ 0<x<1$$

(12) 5 本のベータ分布を描く命令とその結果：

```
Show[Plot[{
PDF[BetaDistribution[1/2, 1/2], x],
PDF[BetaDistribution[1, 2], x],
PDF[BetaDistribution[3, 3], x],
PDF[BetaDistribution[5, 2], x],
PDF[BetaDistribution[1, 1], x]},
{x, 0, 1}],
Graphics[{
Text[" ベータ分布の密度関数 ", {0.5, 3.5}],
Text["α=β=1/2", {0.1, 3.2}],
Text["α=1, β=2", {0.25, 1.9}],
Text["α=β=3", {0.5, 2.2}],
Text["α=5,β=3", {0.8, 2.8}],
Text["α=β=1", {0.1, 0.8}]}]]
```

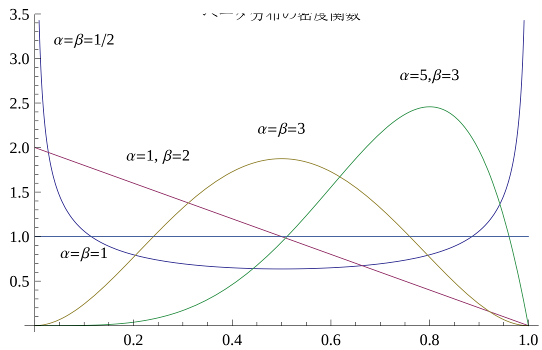


図 18　5 本のベータ分布 Beta[α,β]

6　幾何分布と指数分布

幾何分布と指数分布は、離散型／連続型の違いはありますが、とても仲のよい兄弟分としての性質を示します。(1) メモリーレス、(2) 平均や分散、(3) 条件付き期待値の関係式などが挙げられます。これらの和として、負の 2 項分布、ガンマ分布が対応しています。

幾何分布 Geom(p) のパラメータ変化の状況を調べるために、幾何分布の確率密度を表示する命令 **PDF[GeometricDistribution[p],k]** を入力すると、結果は $(1-p)^k p$ となる。p を $\{0.1, 0.3, 0.5, 0.8\}$ と変化させて描く命令は以下のようになります。

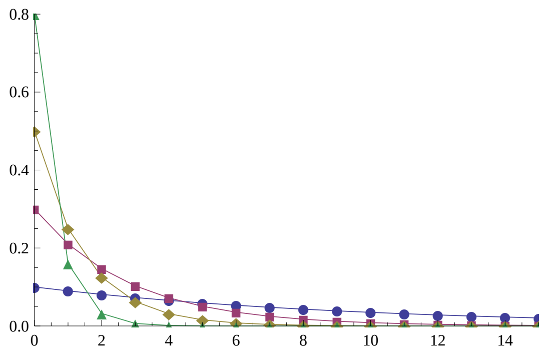


図 19　幾何分布 Geom[p]

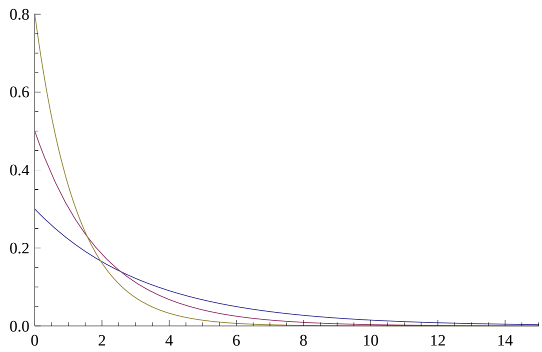


図 20　指数分布 Expo[λ]

```
ListPlot[Table[{k,PDF[GeometricDistribution[p], k]},
{p, {0.1, 0.3, 0.5, 0.8}}, {k, 0, 15}],
PlotMarkers → Automatic, Joined → True, PlotRange → {{0, 15}, {0, 0.8}}]
```

また指数分布のパラメータ変化を調べるには指数分布の確率密度：**PDF[ExponentialDistribution[lm],x]** と入力して $e^{-\mathrm{lm}x}\mathrm{lm}$ を得る。同様に $lm = \{0.3, 0.5, 0.8\}$ の場合を描いてみます。2 つの結果を比べてみてください。ほぼ同じような曲線になっています。

```
Plot[{PDF[ExponentialDistribution[0.3], x],
PDF[ExponentialDistribution[0.5], x],
PDF[ExponentialDistribution[0.8], x]},{x, 0, 15}, PlotRange → {{0, 15}, {0, 0.8}}]
```

7　2 項分布と正規分布

この 2 つの分布を比較することは、正規分布を発見することのきっかけになった、ドモアブル＝ラプラスの定理、さらにはガウスによる中心極限定理と発展していった今日の確率論での最も重要な結果のひとつです。

2 項分布 $Binom[n, p]$ において、n をだんだん大きくして、グラフを描くと、はじめに p の偏りに影響がみられても p が変化しなくても、徐々に少なくなって、平らなひとつ山に近づく。前節でパラメータ変化でみたとおりである。これは一つの中心極限定理の表れであって、ドモアブル＝ラプラスの定理とよばれる。もし p の値が 0 に近づき、積が一定値ならば $(1-p)^n \to 1^\infty = e$ からポアソン分布となる。またスターリングの公式：

$$n! \;\sim\; \sqrt{2\pi n}\, n^n e^{-n}$$

から導かれる 2 項係数の

$$\binom{2n}{n} \;\sim\; \frac{1}{\sqrt{\pi n}} 4^n$$

正規分布の密度の関連した値である。

8　正規分布のとき

標準正規分布 $N(0,1)$ の密度関数は、そう簡単なものではないが、概形は統計学のテキストではかならず概形が描かれる。ベルタイプとよばれ、

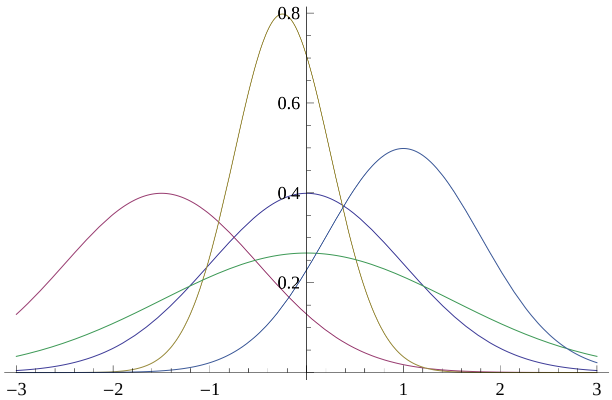
平均 μ と分散 σ^2 をかえた場合の $N(\mu,\sigma^2)$ 密度関数とそのグラフ。実際、密度変数 x の変域は $x_{min}=-\infty$, $x_{max}=\infty$ であって、コンピュータの描画のために設定している。

$$\mathrm{PDF}[\mathrm{NormalDistribution}[\mu,\sigma^2],\{x,x_{min},x_{max}\}]$$

$$\sim\ \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},\ x_{min}<x<x_{max}$$

標準正規分布 $N(0,1)$ の密度関数を表示する: **PDF[NormalDistribution[0, 1], *x*]** を入力して $\dfrac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$ と出力される。平均と分散を変化させた 5 種類のグラフを描く命令は以下の通り。

```
Plot[{
    PDF[NormalDistribution[0, 1], x],
    PDF[NormalDistribution[−1.5, 1], x],
    PDF[NormalDistribution[−1/4, 1/2], x],
    PDF[NormalDistribution[0, 1.5], x],
    PDF[NormalDistribution[1, 0.8], x]}
, {x, −3, 3}]
```



9　カイ 2 乗分布について

統計学で正規分布は重要であるが、これをさらに発展するとカイ 2 乗分布を知っておかなければならない。カイ二乗分布（かいにじょうぶんぷ、カイじじょうぶんぷ)、または χ^2 分布は推測統計学で広く利用される。ヘルメルト (Helmert, F. R. (1875),Zeit Math Phy, 20,300-303) により発見され、K. ピアソンにより命名されたといわれる。カイ二乗検定、適合度検定、独立性の検定など多くの検定法に利用される。

$$\mathrm{PDF}[\mathrm{ChiSquareDistribution}[\nu],x],\{x,x_{min},x_{max}\}]$$

$$\sim\ \frac{2^{-\nu/2}e^{-x/2}x^{-1+\frac{\nu}{2}}}{\mathrm{Gamma}\left[\frac{\nu}{2}\right]},\ x_{min}<x<x_{max}$$

カイ 2 乗分布の密度関数を表示する。ガンマ関数 $\mathrm{Gamma}\,[\nu]=\Gamma(\nu)=\displaystyle\int_0^\infty x^{\nu-1}e^{-x}dx\,(\nu>0)$ をもちいる。階乗の拡張であり、自然数では $\Gamma(n)=(n-1)!$ となり、とくに $\Gamma(1/2)=\sqrt{\pi}$ を覚えておくとよい。

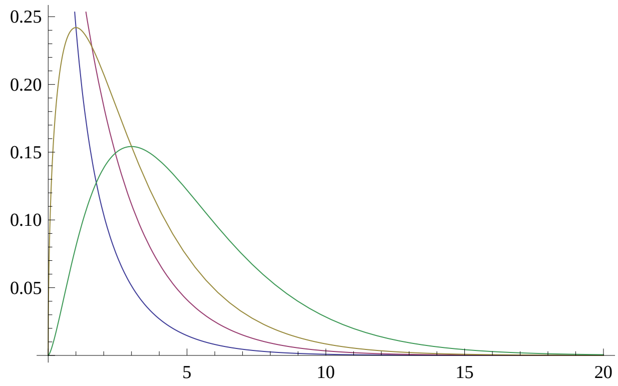
PDF[ChiSquareDistribution[ν], *x*]

$\frac{2^{-\nu/2}e^{-x/2}x^{-1+\frac{\nu}{2}}}{\mathrm{Gamma}\left[\frac{\nu}{2}\right]}$

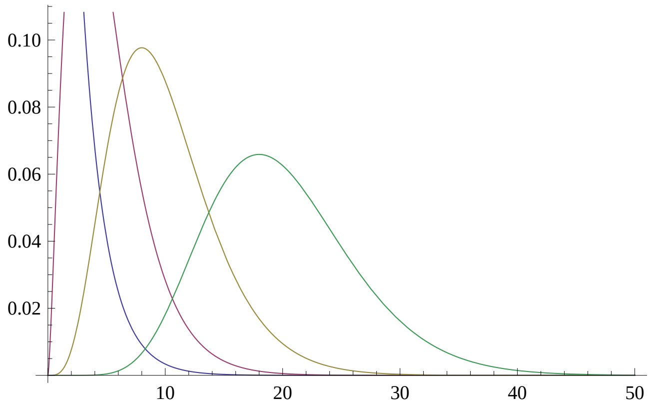
$$\mathrm{Gam}(\nu)\ \sim\ \frac{2^{-\nu/2}e^{-x/2}x^{-1+\dfrac{\nu}{2}}}{\Gamma\left[\frac{\nu}{2}\right]},\ x>0$$

パラメータ $\nu = 1,\ 2,\ 3,\ 5$ の場合をグラフで表示する．

Plot[
{PDF[ChiSquareDistribution[1], x],
PDF[ChiSquareDistribution[2], x],
PDF[ChiSquareDistribution[3], x],
PDF[ChiSquareDistribution[5], x], {x, 0, 20}]



パラメータの値を大きくしてプロットすると
Plot[
{PDF[ChiSquareDistribution[2], x],
PDF[ChiSquareDistribution[5], x],
PDF[ChiSquareDistribution[10], x],
PDF[ChiSquareDistribution[20], x], },
{x, 0, 50}]



10　離散型分布のまとめ

最も広く使われている離散分布であり，その密度，平均，分散，その他の関連特性を計算することができる．分布自体は name[$param_1$,$param_2$,·] という記号形式で表される．統計分布の特性を与える Mean のような関数は，引数として分布の記号表現を取る．

BernoulliDistribution[p]	成功の確率 p の試行を 1 回行う結果、ベルヌーイ (Bernoulli) 分布
BetaBinomialDistribution[α,$beta$,n]	成功確率が BetaDistribution[α,β] 確率変数である二項分布
BetaNegativeBinomialDistribution[α,β,n]	成功確率が BetaDistribution[α,β] 確率変数である負の二項分布
BinomialDistribution[n,p]	各試行における成功率を p としたときに，n 回の試行のうちで起る成功の回数
DiscreteUniformDistribution[$\{i_{min}$,$i_{max}\}$]	i_{min} から i_{max} までの整数値上での離散一様分布
GeometricDistribution[p]	各試行における成功率を p としたときの，最初の成功までの試行回数の幾何分布
HypergeometricDistribution[n,n_{succ},n_{tot}]	成功数 n_{succ} を含む大きさ n_{tot} の母集団から標本数 n を取り出したときの成功数の超幾何分布
LogSeriesDistribution[θ]	パラメータ θ の対数級数分布
NegativeBinomialDistribution[n,p]	パラメータ n および p の負の二項分布
PoissonDistribution[μ]	平均 μ のポアソン (Poisson) 分布

一般的な離散分布のほとんどは，「成功と失敗」のような 2 通りの結果が考えられる試行の繰り返しと考えることで理解できる．

• ベルヌーイ分布 BernoulliDistribution[p] とは，成功（値 1 に対応）が p の確率で起り，失敗（値 0 に対応）が $1-p$ の確率で起る 1 回 (単一) 試行の確率分布である．

• 二項分布 BinomialDistribution[n,p] とは，各試行における成功確率を p としたときに，n 回の独立した試行で起る成功を集計した回数値の分布である．

• 正の整数 n に対する負の二項分布 NegativeBinomialDistribution[n,p] とは，各試行における成功確率を p としたときに，n 回成功するまでに起った失敗回数の分布である．この分布は任意の正の n について定義される．しかし，n が整数でないときは，n が成功数，p が成功確率という解釈は当てはまらない．

• ベータ二項分布 BetaBinomialDistribution[α,β,n] は二項分布とベータ分布を組み合せたものである．BetaBinomialDistribution[α,β,n] の確率変数は，成功確率 p 自体が確率変数でベータ分布 BetaDistribution[α,β] に従うもののとき、BinomialDistribution[n,p] に従う．ベータ負の二項分布 BetaNegativeBinomialDistribution[α,β,n] はベータ分布と負の二項分布を組み合せたものである．p の値が既知のものとするのではなく、変動するものと考える。

• 幾何分布 GeometricDistribution[p] とは，各試行の成功率を p としたときに，最初に成功が起るまでに要した試行数の分布のことである．

• 超幾何分布 HypergeometricDistribution[n, n_{succ}], n_{tot}]] は n 回の試行が，成功可能回数 n_{succ} 回でサイズ n_{tot} の母集団から元に戻さずにサンプリングする（非復元抽出）ことに相当する実験で，二項分布に代り使用されるものである．

• 離散一様分布 DiscreteUniformDistribution[i_{min}, i_{max}] とは，同じ値の確率（等しい確率）で起る結果を i_min から i_{max} までの整数で表した試行を表すものである．

• ポアソン分布 PoissonDistribution[μ] とは，指定された時間内に起る事象数を記述する。ここで μ は一定時間に対する事象の平均回数である。2 項分布の極限として得られる、稀な現象を記述するときに用いられ、

重要な分布である。

• $\theta = 0$ 付近の $\log(1-\theta)$ の級数展開における項は，対数分布 LogSeriesDistribution[θ] に従った離散確率変数の確率に比例する。指定された期間内に購入される製品数の分布が，この分布によってモデル化されることがある。

一般的な密度、分布、平均、分散などを求める命令形：

PDF[dist,x]	分布名 dist の変数 x における確率密度関数 (probabilisty density function)
CDF[dist,x]	x における累積分布関数 (cumulative distribution dunction)、
InverseCDF[dist,q]	CDF[dist,x] が最大で q となるような最大整数 x
Quantile[dist,q]	第 q 分位数
Mean[dist]	平均 (mean)
Variance[dist]	分散 (variance)
StandardDeviation[dist]	標準偏差 (standard deviation)
Skewness[dist]	歪度係数 (skewness)
Kurtosis[dist]	尖度係数 (kurtosis)
CharacteristicFunction[dist,t]	特性関数 $\phi(t) = E[e^{iXt}]$, $X \sim dist$
ExpectedValue[f,dist]	dist の (純) 関数 f の期待値
ExpectedValue[$f[x]$,dist,x]	dist の変数 x に対する 関数 $f[x]$ の期待値
Median[dist]	中央値
Quartiles[dist]	dist に対する 1/4, 1/2, 3/4 番目の四分位数 (quartile)
InterquartileRange[dist]	第 1 四分位数 (first) と第 3 四分位数 (third) との差分
QuartileDeviation[dist]	四分位範囲の半分値
QuartileSkewness[dist]	四分位数ベースの歪度
RandomInteger[dist]	指定された分布の擬似乱数
RandomInteger[dist,dims]	次元が dims で，指定された分布の要素を持つ擬似乱数配列

11 連続型分布のまとめ

ここに記載されている関数は，連続分布で最も一般的に使用されるものであり，連続分布の密度，平均，分散，その他の関連特性を計算することができる．分布は name[$param_1$],$param_2$,...] という記号形式で表現される．統計分布の特性を与える Mean のような関数は，引数として分布の記号表現を取る．「離散分布」は多数の離散分布について記述する．

UniformDistribution[{min, max}]	区間 (min, max) における一様分布
NormalDistribution[μ,σ]	平均 μ，標準偏差 σ (分散 σ^2) の正規（ガウス）分布
HalfNormalDistribution[θ]	尺度がパラメータ θ に反比例する半正規分布
LogNormalDistribution[μ,σ]	平均 μ，標準偏差 σ (分散 σ^2) の正規分布に基づいた対数正規分布
InverseGaussianDistribution[μ,λ]	平均 μ，尺度 λ の逆ガウス分布

• 一様分布 UniformDistribution[min,max] は，矩形分布と呼ばれることもあり，値がどこにでも等しくあるような確率変数を特徴付ける．一様分布の確率変数の例には，min から max までの直線上でランダムに選んだ点の場所がある．

• 対数正規分布 LogNormalDistribution[μ,σ] は正規分布に基づく確率変数の指数関数が従う分布である．

この分布は多数の独立確率変数が乗法的に統合された場合に起る．

半正規分布 HalfNormalDistribution[θ] は領域 $[0, \infty]$) に限られた分布 NormalDistribution$[0, 1/(\theta\sqrt{2/\pi})]$ に比例する

• 逆ガウス分布 InverseGaussianDistribution[μ, λ] はワルド (Wald) 分布とも呼ばれ，正のドリフト付きブラウン運動の初期通過時間の分布である．

ChiSquareDistribution[ν]	自由度 ν の χ^2 分布
InverseChiSquareDistribution[ν]	自由度 ν の逆 χ^2 分布
FRatioDistribution[n,m]	分子の自由度が n，分母の自由度が m である F 分布
StudentTDistribution[ν]	自由度が ν のスチューデントの t 分布
NoncentralChiSquareDistribution[ν,λ]	自由度が ν，非心パラメータが λ の非心 χ^2 分布
NoncentralStudentTDistribution[ν,δ]	自由度が ν，非心パラメータが δ の非心スチューデントの t 分布
NoncentralFRatioDistribution[n,m,λ]	分子の自由度が n，分母の自由度が m，非心パラメータが λ である非心 F 分布

正規分布に従う標本に関連する分布：

• $X_1, \cdots, X_\nu$ が単位分散と零平均を持つ独立した正規確率変数ならば，$\sum_{i=1}^{\nu} X_i^2$ は自由度 ν の χ^2 分布にしたがう．正規変数からその平均を引き標準偏差で割ることにより標準化されると，その数量の平方和はこの分布に従う．χ^2 分布は正規分布する標本の分散を記述する場合に最もよく用いられる．

• もし Y が自由度 ν の χ^2 分布に従うならば，$1/Y$ は逆 χ^2 分布 InverseChiSquareDistribution[ν] に従う．自由度が ν でスケールが ξ のスケールされた逆 χ^2 分布は InverseChiSquareDistribution[ν,ξ] として与えることができる．逆 χ^2 分布は，正規分布に従う標本のベイズ分析における分散の事前分布として広く使われている．

• スチューデントの t 分布になる変数は，正規確率変数の関数として記述することもできる．X と Z が独立した確率変数であるとする．ここで X は標準正規分布，Z は自由度 ν の χ^2 変数である．この場合，$X/\sqrt{Z/\nu}$ は自由度 ν の t 分布を持つ．スチューデントの t 分布は縦軸について対称であり，正規変数のその標準偏差に対する割合を特徴付ける．$\nu = 1$ ならば，t 分布はコーシー分布に一致する．

• F 分布は 2 つの独立した χ^2 変数をそれぞれの自由度で割ったときの比の分布である．これは仮定検定で 2 つの母集団の分散を比較するときに広く使われる．

平均が非零の正規分布から派出される分布は「非心分布」と呼ばれる．

• 非心スチューデントの t 分布: NoncentralStudentTDistribution[ν,δ] は，比 $X/\sqrt{\chi^2_\nu/\nu}$ を記述する．ここで χ^2_ν は自由度 ν の χ^2 確率変数，X は平均 δ, 分散 $\sigma^2 = 1$ の正規分布の独立した確率変数である．

• 非心 F 分布: NoncentralFRatioDistribution[n,m,λ] は $1/m\chi^2_m$ に対する $1/n\chi^2_n$ の比の分布である．ここで $\chi^2_n(\lambda)$ は非心パラメータ λ，自由度 n の非心 χ^2 確率変数であり， $\chi^2_m(\lambda)$ は自由度 m の χ^2 確率変数である．

BetaDistribution[α, β]	形状パラメータが α と β の連続ベータ分布
CauchyDistribution[a,b]	位置パラメータが a，尺度パラメータが b のコーシー分布
ChiDistribution[ν]	自由度 ν の χ 分布
ExponentialDistribution[λ]	尺度がパラメータ λ に反比例する指数分布
ExtremeValueDistribution[α, β]	位置パラメータ α ，尺度パラメータ β の極値 (Fisher-Tippett) 分布
GammaDistribution[α, β]	形状パラメータ α，尺度パラメータ λ のガンマ分布
GumbelDistribution[α, β]	位置パラメータ α，尺度パラメータ β のガンベルの最小極値分布
InverseGammaDistribution[α, β]	形状パラメータ α，尺度パラメータ λ の逆ガンマ分布
LaplaceDistribution[μ,β]	平均 μ，尺度パラメータ β のラプラス（二重指数）分布
LevyDistribution[μ, σ]	位置パラメータ μ，拡散パラメータ σ のレヴィ分布
LogisticDistribution[μ,β]	平均 μ，尺度パラメータ β のロジスティック分布
MaxwellDistribution[σ]	尺度パラメータ σ のマックスウェル（マックスウェル・ボルツマン）分布
ParetoDistribution[k,β]	最小値パラメータ k ，形状パラメータ α のパレート分布
RayleighDistribution[σ]	尺度パラメータ σ のレイリー (Rayleigh) 分布
WeibullDistribution[α,β]	形状パラメータ α ，尺度パラメータ β のワイブル (Weibull) 分布

• X が $[-\pi, \pi]$ 上で一様に分布しているなら，確率変数 $\tan(X)$ は $a = 0$, $b = 1$ コーシー (Cauchy) 分布: CauchyDistribution[a,b] に従う．

• $\alpha = n/2$ で $\lambda = 2$ のとき，ガンマ分布: GammaDistribution[α,λ] は n ユニットの正規確率変数の平方和に分布を記述する．この形式のガンマ分布は自由度 ν の χ^2 分布と呼ばれる．$\alpha = 1$ のとき，ガンマ分布は，イベント間の待ち時間を記述するのによく使われる指数分布 ExponentialDistribution[λ] の形を取る．

• もし確率変数 X がガンマ分布: GammaDistribution[α, β] に従うならば，$1/X$ は逆ガンマ分布: InverseGammaDistribution[α,1/β] に従う．確率変数 X が InverseGammaDistribution[1/2,σ/2] に従うならば，$X + \mu$ はラヴィ分布 LevyDistribution[μ,σ] に従う．

• X_1 と X_2 が，尺度パラメータが等しい別々のガンマ分布であるとき，確率変数 $X_1/(X_1 + X_2)$ はベータ分布: BetaDistribution[α,β] に従う．

• χ 分布: ChiDistribution[ν] は，χ^2 確率変数の平方根が従う分布である．$n = 1$ のときは，χ 分布は $\theta = \sqrt{\pi/2}$ の HalfNormalDistribution[θ] に等しい．$n = 2$ のときは，χ 分布は $\sigma = 1$ のレイリー (Rayleigh) 分布: RayleighDistribution[σ] に等しい．$n = 3$ のときは χ 分布は $\sigma = 1$ のマックスウェル・ボルツマン (Maxwell-Boltzmann) 分布: MaxwellDistribution[σ] に等しい．

• ラプラス (Laplace) 分布: LaplaceDistribution[μ,β] は，同一の指数分布を持つ 2 つの独立した確率変数の差の分布である．ロジスティック分布: LogisticDistribution[μ,β] は裾の長い分布を望む場合に，正規分布の代りに頻繁に使われる．

• パレート分布 (Pareto) 分布: ParetoDistribution[k,α] は収入を表すために使われる．k は最小の収入を

示す．

- ワイブル (Weibull) 分布: WeibullDistribution[α,β] は，物体の寿命を記述するために工学で広く使われている．極値分布: ExtremeValueDistribution[α,β] は，正規分布を含む多様な分布から抽出された大きい標本における最大値に対する極限分布である．このような標本の最小値に対する極限分布は，ガンベル (Gumbel) 分布: GumbelDistribution[α,β] である．最大極値と最小極値は，変数の線形変化により関連しているので，極値分布とガンベル分布は同じ意味で使われることがある．また極値分布した確率変数と，適切にシフト・スケールされたワイブル分布した確率変数との間には対数関係があるので，極値分布は対数ワイブル分布と呼ばれることもある．