色々な確率分布

- 黒木玄

- 2022-04-11〜2022-04-12

このノートでは[Julia言語 (https://julialang.org/)](https://julialang.org/)を使用している:

- [Julia言語のインストールの仕方の一例 (https://nbviewer.org/github/genkuroki/msfd28/blob/master/install.ipynb)](https://nbviewer.org/github/genkuroki/msfd28/blob/master/install.ipynb)

自明な誤りを見つけたら, 自分で訂正して読んで欲しい. 大文字と小文字の混同や書き直しが不完全な場合や符号のミスは非常によくある.

このノート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

In [1]:

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

1 確率変数の独立性と無相関性

1.1 同時確率質量函数と同時確率密度函数

複数の確率変数 $X_1, \dots, X_n$ が同時確率質量函数 $P(x_1, \dots, x_n)$ を持つとは, $P(x_1, \dots, x_n) \geq 0$ であり, $P(x_1, \dots, x_n)$ の総和が 1 にあり,

$$E[f(X_1, \dots, X_n)] = \sum_{x_1} \cdots \sum_{x_n} f(x_1, \dots, x_n) P(x_1, \dots, x_n)$$

が成立することである. このとき, X_1 単独の確率質量函数 $P(x_1)$ は

$$P(x_1) = \sum_{x_2} \cdots \sum_{x_n} P(x_1, \dots, x_n)$$

になる. なぜならば,

複数の確率変数 $X_1, \dots, X_n$ が同時確率密度函数 $p(x_1, \dots, x_n)$ を持つとは, $p(x_1, \dots, x_n) \geq 0$ であり, $p(x_1, \dots, x_n)$ の積分が 1 になり,

$$E[f(X_1, \dots, X_n)] = \int \cdots \int f(x_1, \dots, x_n) p(x_1, \dots, x_n)\, dx_1 \cdots dx_n$$

が成立することである. ここで $\int$ は定積分を表す.

X_1

1.2 確率変数の独立性の定義

複数の確率変数 $X_1, \dots, X_n$ が同時に与えられており, それらの函数の期待値 $E[f(X_1, \dots, X_n)]$ が定義されているとする. このとき, $X_1, \dots, X_n$ が **独立** (independent)であるとは,

$$E[f(X_1) \cdots f(X_n)] = E[f(X_1)] \cdots E[f(X_n)]$$

と X_i ごと函数達 $f(X_i)$ の積の期待値が各々の $f(X_i)$ の期待値の積になることだと定める.

注意: 厳密には函数 f_i 達を動かす範囲を確定させる必要があるが, このノートではそのようなことにこだわらずに解説することにする. (厳密な理論の展開のためには測度論的確率論の知識が必要になる.)

上の条件は, 例えば, X_i 達のそれぞれが確率密度函数 $p_i(x_i)$ を持つとき, $(X_1, \dots, X_n)$ の同時確率密度函数 $p(x_1, \dots, x_n)$ が $p(x_1) \cdots p(x_n)$ と積の形になること

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

と同値だと考えてよい. すなわち, $(X_1, \dots, X_n)$ の函数の期待値が次のように書かれることと同値だと思ってよい:

$$E[f(X_1, \dots, X_n)] = \int \cdots \int f(x_1, \dots, x_n) p(x_1) \cdots p(x_n)\, dx_1 \cdots dx_n.$$

ここで $\int \cdots \int$ は適切な範囲での定積分を表す.

X_i 達のそれぞれが確率質量函数 $P_i(x_i)$ を持つ場合には, $X_1, \dots, X_n$ の独立性は

$$E[f(X_1, \dots, X_n)] = \sum_{x_1} \cdots \sum_{x_n} f(x_1, \dots, x_n) P_n(x_1) \cdots P_1(x_n)$$

が成立することと同値である.

確率密度函数と確率質量函数が混ざっている場合も同様である.

大雑把に誤解を恐れずに言うと, 確率が積になるときに独立だという.

実際の計算では, 積の期待値が期待値の積になるという形式で確率変数の独立性を使う.

1.3 独立同分布の定義

確率変数 $X_1, \dots, X_n$ が **独立同分布** (independent and identically distributed, **i.i.d.**, **iid**) であるとは, それらが独立でかつ X_i 達が従う分布が等しいことであると定める.

1.4 互いに異なる任意の2つが独立であっても全体が独立であるとは限らない

確率変数達 X, Y, Z の互いに異なる任意の2つが独立であっても, X, Y, Z の全体が独立でない場合があることを具体的な例を作ることによって証明しよう.

確率変数 X, Y, Z の同時確率質量函数が $P(x, y, z)$ であるとき, X 単独の確率質量函数は $P(x) = \sum_{y,z} P(x, y, z)$ になり, (X, Y) の同時確率質量函数は $P(x, y) = \sum_z P(x, y, z)$ になる. 他の場合も同様である.

注意: 変数名で異なる函数を区別するスタイルを採用したので記号法の運用時に注意すること. このスタイルでは $P(x, y)$ と $P(x, z)$ は異な函数になる. 値を代入する場合には $P(x = 1, y = 1)$ や $P(x = 1, z = 1)$ のように書いて区別できるようにする.

$P(x, y, z)$ で $P(x, y) = P(x)P(y)$, $P(x, z) = P(x)P(z)$, $P(y, z) = P(y)P(z)$ を満たすが $P(x, y, z) \neq P(x)P(y)P(z)$ となるものを具体的に構成すればよい.

例えば以下のような確率質量函数 $P(x, y, z)$ の例を作ることができる:

	$z = 1$		$z = 0$	
	$y = 1$	$y = 0$	$y = 1$	$y = 0$
$x = 1$	$P(1, 1, 1) = 1/4$	$P(1, 0, 1) = 0$	$P(1, 1, 0) = 0$	$P(1, 0, 0) = 1/4$
$x = 0$	$P(0, 1, 1) = 0$	$P(0, 0, 1) = 1/4$	$P(0, 1, 0) = 1/4$	$P(0, 0, 0) = 0$

x, y, z はそれぞれ $1, 0$ を動くとする. このとき, $P(x, y)$, $P(x, z)$, $P(y, z)$ の値はすべて 1/4 になり, $P(x)$, $P(y)$, $P(z)$ の値はすべて 1/2 になることがわかる. たとえば, $P(x = 1, y = 1) = P(1, 1, 1) + P(1, 1, 0) = 1/4 + 0 = 1/4$ であり, $P(x = 1) = \sum_{y,z} P(1, y, z) = 1/4 + 0 + 0 + 1/4 = 1/2$.

だから, $P(x, y) = P(x)P(y)$, $P(x, z) = P(x)P(z)$, $P(y, z) = P(y)P(z)$ が成立するが, $P(x, y, z) \neq 1/8 = P(x)P(y)P(z)$ となっている.

このとき, X, Y, Z が同時確率質量函数 $P(x, y, z)$ を持つとすると, そのうちの任意の2つは独立だが, X, Y, Z の全体は独立ではない.

1.5 確率変数の独立性の現実における解釈に関する重大な注意

上の例は具体的には次のような状況だと解釈可能である.

(1) $P(x) = 1/2$ の解釈: $X = 1$ は薬Aを与えたことを, $X = 0$ は薬Bを与えたことを意味する. 全員に確率 1/2 で薬AまたはBを与えた.

(2) $P(y) = 1/2$ の解釈: $Y = 1$ は薬に効果があったことを, $Y = 0$ は効果が無かったことを意味する. 全体で見たら, 1/2 の確率で薬には効果があった.

(3) $P(z) = 1/2$ の解釈: $Z = 1$ は女性であることをを, $Z = 0$ は男性であることを意味する. 女性である確率と男性である確率は 1/2 だった.

(4) 男女の区別をやめると, 薬Aも薬Bも効果がある確率は半々であり, 薬Aと薬Bのどちらを与えたかと効果があったかどうかは独立である. 男女を合わせた($z = 1, 0$ の場合の和を取って得られる)確率質量函数 $P(x, y)$ の表

	$y = 1$	$y = 0$
$x = 1$	$P(x = 1, y = 1) = 1/4$	$P(x = 1, y = 0) = 1/4$
$x = 0$	$P(x = 0, y = 1) = 1/4$	$P(x = 0, y = 0) = 1/4$

は$x = 1$ の薬Aの場合も $x = 0$ の薬Bの場合も男女全体を見ると半々で効果があり, 男女全体では効果に変わりがないことを意味している.

(5) 薬Aと薬Bのどちらを与えたかと男女のどちらであるかは独立である. そのことは効果の有無を意味する $y = 1, 0$ について和を取って得られる確率質量函数 $P(x, z)$ の表

	$z = 1$	$z = 0$
$x = 1$	$P(x = 1, z = 1) = 1/4$	$P(x = 1, z = 0) = 1/4$
$x = 0$	$P(x = 0, z = 1) = 1/4$	$P(x = 0, z = 0) = 1/4$

からわかる.

(6) 薬Aと薬Bのどちらを与えたかを無視すると, 男女のどちらであるかと薬の効果の有無は独立である. そのことは薬Aと薬Bのどちらを与えたかを意味する $x = 1, 0$ について和を取って得られる確率質量函数 $P(y, z)$ の表

	$z = 1$	$z = 0$
$y = 1$	$P(y = 1, z = 1) = 1/4$	$P(y = 1, z = 0) = 1/4$
$y = 0$	$P(y = 0, z = 1) = 1/4$	$P(y = 0, z = 0) = 1/4$

からわかる.

(7) しかし, 男女を区別すると全然違う結果が見えて来る. 薬Aは女性だけに効果があり, 薬Bは男性だけに効果がある. $z=1$ の女性の場合に制限した確率質量函数の表

	$z=1$	
	$y=1$	$y=0$
$x=1$	$P(1,1,1)=1/4$	$P(1,0,1)=0$
$x=0$	$P(0,1,1)=0$	$P(0,0,1)=1/4$

より, $x=1$ の薬Aの場合には $y=1$ の確率が正で効果があるが, $x=0$ の薬Bの場合には $y=1$ の確率が0で効果がないことがわかる. $z=0$ の男性の場合に制限した確率質量函数の表

	$z=0$	
	$y=1$	$y=0$
$x=1$	$P(1,1,0)=0$	$P(1,0,0)=1/4$
$x=0$	$P(0,1,0)=1/4$	$P(0,0,0)=0$

より, $x=1$ の薬Aの場合には $y=1$ の確率が0で効果がないが, $x=0$ の薬Bの場合には $y=1$ の確率が正で効果があることがわかる.

このように確率変数達が独立か否かは現実において重大な意味を持ち得る. ある重要な条件(上の場合には女性か男性か)を無視して, 「XとYは独立である」と結論すると大変なことになってしまう場合がある. XとYも, XとZも, YとZも独立であっても, XとYとZの全体は独立でないかもしれない.

1.6 共分散と相関係数

確率函数 X, Y の期待値をそれぞれ $\mu_X = E[X]$, $\mu_Y = E[Y]$ と書くことにする.

確率変数 X, Y の **共分散** (covariance) $\sigma_{XY} = \operatorname{cov}(X, Y)$ を次のように定義する:

$$\sigma_{XY} = \operatorname{cov}(X, Y) = E[(X-\mu_X)(Y-\mu_Y)].$$

$X = Y$ のときこれは X の分散になる.

線形代数の言葉を使えば, 共分散 $\sigma_{XY} = \operatorname{cov}(X, Y)$ は内積に対応しており, 分散 $\sigma_X^2 = \operatorname{var}(X)$, $\sigma_Y^2 = \operatorname{var}(Y)$ はノルム(ベクトルの長さ)の二乗に対応しており, 標準偏差 $\sigma_X = \operatorname{std}(X)$, $\sigma_Y = \operatorname{std}(Y)$ はノルム(ベクトルの長さ)に対応している.

さらに確率変数 X, Y の **相関係数** (correlation coefficient) $\rho_{XY} = \operatorname{cov}(X, Y)$ はベクトルのあいだのなす角度 θ に対する $\cos\theta$ の対応物として次のように定義される:

$$\rho_{XY} = \operatorname{cor}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\operatorname{std}(X)\operatorname{std}(Y)} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}.$$

相関係数という言葉を見たら $\cos\theta$ を想像し, 分散, 標準偏差, 共分散という言葉を見たら, ベクトルの長さの2乗, ベクトルの長さ, ベクトルの内積を想像すればよい. $X-\mu_X$ や $Y-\mu_Y$ がベクトルに対応している.

X, Y の共分散が $\operatorname{cov}(X, Y) = 0$ のとき(この条件は相関係数が $\operatorname{cor}(X, Y) = 0$ となることと同値, 直観的には「直交している」と考える), 確率変数達 X, Y は **無相関** であるという.

確率変数達 $X_1, \dots, X_n$ が無相関であるとは, そのうちの互いに異なる任意の2つが無相関であることだと定める. $\mu_i = E[X_i]$, $\sigma_i^2 = \operatorname{var}(X_i)$ のとき, $X_1, \dots, X_n$ が無相関であることは,

$$\operatorname{cov}(X_i, X_j) = E[(X_i-\mu_i)(X_j-\mu_j)] = \sigma_i^2\delta_{ij}$$

と書けることと同値である. ここで δ_{ij} は $i=j$ の場合にのみ 1 になり, それ以外の場合に 0 になるKroneckerのデルタである.

1.7 問題: 独立ならば無相関である (実質1行で解ける)

確率変数達 X, Y が独立ならば確率変数達 X, Y は無相関であることを示せ.

解答例: X, Y は独立なので $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$ となる. ゆえに, $\mu_X = E[X]$, $mu_Y = E[Y]$ とおくと,

$$E[(X-\mu_X)(Y-\mu_Y)] = E[X-\mu_X]E[Y-\mu_Y] = (E[X]-\mu_X)(E[Y]-\mu_Y) = 0\cdot 0 = 0.$$

解答終

1.8 問題: 無相関でも独立とは限らない

確率変数 X, Y で無相関だが独立でないものを具体的に構成せよ.

解答例1: 確率質量関数 $P(x, y)$ を次の表の通りに定める:

	$y = 1$	$y = 2$	$y = 3$	
$x = 1$	$P(1, 1) = 0$	$P(1, 2) = 1/4$	$P(1, 3) = 0$	$P(x = 1) = 1/4$
$x = 2$	$P(2, 1) = 1/4$	$P(2, 2) = 0$	$P(2, 3) = 1/4$	$P(x = 2) = 2/4$
$x = 3$	$P(3, 1) = 0$	$P(3, 2) = 1/4$	$P(3, 3) = 0$	$P(x = 3) = 1/4$
	$P(y = 1) = 1/4$	$P(y = 2) = 2/4$	$P(y = 3) = 1/4$	

確率変数 X, Y は同時確率質量関数 $P(x, y)$ を持つとする. このとき,

$$E[X] = E[Y] = 1 \cdot (1/4) + 2 \cdot (2/4) + 3 \cdot (1/4) = 2.$$

であり, 確率が 0 でない (x, y) については x または y が X と Y の期待値 2 になるので, 共分散は 0 になる:

$$\begin{aligned}\mathrm{cov}(X, Y) = E[(X - 2)(Y - 2)] &= \sum_{x,y} (x - 2)(y - 2)P(x, y) \\ &= (2 - 2)(1 - 2)(1/4) + (1 - 2)(2 - 2)(1/4) \\ &\quad + (3 - 2)(2 - 2)(1/4) + (2 - 2)(3 - 2)(1/4) = 0.\end{aligned}$$

これで X, Y は無相関であることがわかった.

しかし, $P(1, 1) = 0$ と $P(x = 1)P(y = 1) = (1/4)(1/4)$ は等しくないので X, Y は独立ではない.

解答終

解答例2: 確率質量関数 $P(x, y)$ を次の表の通りに定めても上と同様に, X, Y は無相関だが独立ではないことを示せる:

	$y = 1$	$y = 2$	$y = 3$	
$x = 1$	$P(1, 1) = 1/8$	$P(1, 2) = 1/8$	$P(1, 3) = 1/8$	$P(x = 1) = 3/8$
$x = 2$	$P(2, 1) = 1/8$	$P(2, 2) = 0$	$P(2, 3) = 1/8$	$P(x = 2) = 2/8$
$x = 3$	$P(3, 1) = 1/8$	$P(3, 2) = 1/8$	$P(3, 3) = 1/8$	$P(x = 3) = 3/8$
	$P(y = 1) = 3/8$	$P(y = 2) = 2/8$	$P(y = 3) = 3/8$	

解答終

解答例3: $\mathbb{R}^2$ における単位円盤上の一様分布の確率密度関数を次のように定める:

$$p(x, y) = \begin{cases} 1/\pi & (x^2 + y^2 \le 1) \\ 0 & (x^2 + y^2 > 1) \end{cases}$$

これを同時確率密度関数として持つ確率変数 X, Y を考える:

$$E[f(X, Y)] = \iint_{\mathbb{R}^2} f(x, y)p(x, y)\,dx\,dy.$$

単位円盤の対称性から $E[X] = E[Y] = 0$ となることがわかる. (具体的に積分を計算しても易しい.) それらの共分散は

$$\mathrm{cov}(X, Y) = E[XY] = \frac{1}{\pi} \iint_{x^2+y^2 \le 1} xy\ dx\,dy$$

と書けるが, $xy \ge 0$ の部分の積分と $xy \le 0$ の部分の積分が円盤の対称性より互いに打ち消しあって $\mathrm{cov}(X, Y) = 0$ となることがわかる. X 単独の密度函数は

$$p(x) = \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = \frac{2}{\pi}\sqrt{1 - x^2}$$

となり, 同様にして $p(y) = (2/\pi)\sqrt{1 - y^2}$ となるが, $-1 < x, y < 1$ かつ $x^2 + y^2 > 1$ のとき, $p(x, y) = 0$ となるが, $p(x)p(y) \ne 0$ となるので, それらは等しくない. ゆえに X, Y は独立ではない.

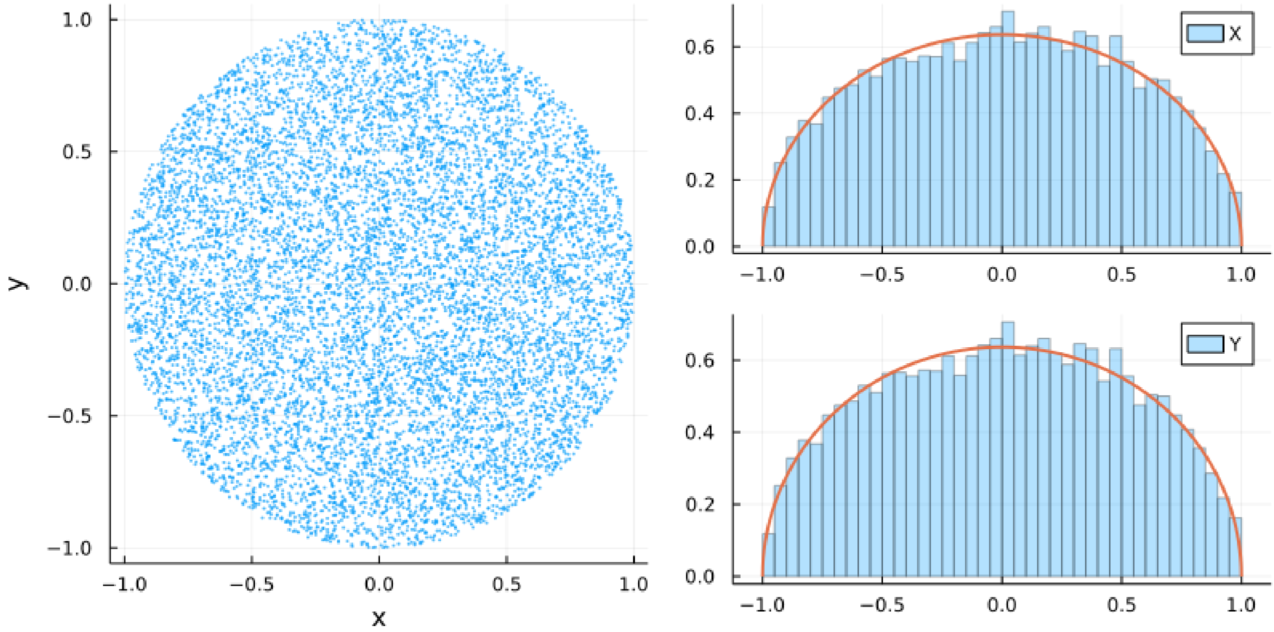
解答終

In [2]:

```julia
n = 10^4
XY = [(r = √rand(); t = 2π*rand(); (r*cos(t), r*sin(t))) for _ in 1:n]
X, Y = first.(XY), last.(XY)
@show cov(X, Y)
P1 = scatter(X, Y; msc=:auto, ms=2, alpha=0.7, label="", xlabel="x", ylabel="y")
P2 = histogram(X; norm=true, alpha=0.3, bin=41, label="X")
plot!(x -> 2/π*√(1 - x^2), -1, 1; label="", lw=2)
P3 = histogram(X; norm=true, alpha=0.3, bin=41, label="Y")
plot!(y -> 2/π*√(1 - y^2), -1, 1; label="", lw=2)
plot(P1, P2, P3; size=(800, 400), layout=@layout [a [b; c]])
```

```
cov(X, Y) = -0.004728646065725972
```

Out[2]:



1.9 問題: 無相関な確率変数の和の分散

$X_1, \dots, X_n$ は独立でも無相関とも限らない確率変数達であるとする. このとき, 期待値を取る操作の線形性より,

$$E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n]$$

となる. 確率変数達の和の期待値は, 独立性や無相関性が成立していなくても, それぞれの期待値の和になる.

無相関性を仮定すると, 分散についても簡明な結果を得ることができる.

$X_1, \dots, X_n$ が **無相関な確率変数達** ならば(特に独立な確率変数ならば), 次が成立することを示せ:

$$\mathrm{var}(X_1 + \cdots + X_n) = \mathrm{var}(X_1) + \cdots + \mathrm{var}(X_n).$$

ヒント: 互いに直交するベクトル達 $v_1, \dots, v_n$ について, 内積を $(\ ,\ )$ と書くとき, $(v_i, v_j) = \|v_i\|^2\delta_{ij}$ が成立することを使えば, $\|v_1 + \cdots + v_n\|^2 = \|v_1\|^2 + \cdots + \|v_n\|^2$ を示せることと本質的に同じ. この結果はPythagorasの定理(平面の場合は三平方の定理)そのものである.

解答例: $\mu_i = E[X_i]$ とおくと,

$$E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = \sum_{i=1}^n \mu_i$$

が成立しており, さらに, $\sigma_i^2 = \mathrm{var}(X_i) = E[(X_i - \mu_i)^2]$ とおくと, $X_1, \dots, X_n$ が無相関であることより,

$$E[(X_i - \mu_i)(X_j - \mu_j)] = \mathrm{cov}(X_i, X_j) = \sigma_i^2\delta_{ij}$$

が成立しているので(δ_{ij} は $i = j$ の場合にのみ 1 でそれ以外のとき 0), 一般に

$$\left(\sum_{i=1}^n a_i\right)^2 = \sum_{i=1}^n a_i \sum_{i=j}^n a_j = \sum_{i,j=1}^n a_i a_j$$

と計算できることを使うと,

$$\begin{aligned}
\operatorname{var}\left(\sum_{i=1}^n X_i\right) &= E\left[\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i\right)^2\right] = E\left[\left(\sum_{i=1}^n (X_i - \mu_i)\right)^2\right] \\
&= E\left[\sum_{i,j=1}^n (X_i - \mu_i)(X_j - \mu_j)\right] = \sum_{i,j=1}^n E\left[(X_i - \mu_i)(X_j - \mu_j)\right] \\
&= \sum_{i,j=1}^n \sigma_i^2 \delta_{ij} = \sum_{i=1}^n \sigma_i^2 = \sum_{i=1}^n \operatorname{var}(X_i).
\end{aligned}$$

解答終

1.10 問題: 二項分布と負の二項分布の平均と分散のBernoulli分布と幾何分布の場合への帰着

Bernoulli分布 $\mathrm{Bernoulli}(p)$ の平均と分散がそれぞれ $p, p(1-p)$ であることと, 幾何分布 $\mathrm{Geometric}(p)$ の平均と分散がそれぞれ $(1-p)/p, (1-p)/p^2$ であることを認めて, 二項分布 $\mathrm{Binomial}(n, p)$ と負の二項分布 $\mathrm{NegativeBinomial}(k, p)$ の平均と分散を平易に求めてみよ.

解答例:

二項分布は試行回数 n の成功確率 p のBernoulli試行で生成された 1 と 0 からなる長さ n の数列中に含まれる 1 の個数の分布であった. Bernoulli試行の確率質量函数は

$$P(x_1, \dots, x_n) = P(x_1) \cdots P(x_n), \quad P(x_i) = p^{x_i}(1-p)^{1-x_i} \quad (x_i = 1, 0)$$

とBernoulli分布の確率質量函数 $P(x_i)$ の積で書けるのであった. この事実はBernoulli分布に従う確率変数 $X_1, \dots, X_n$ が独立であることを意味する. そして, Bernoulli試行で生成された 1 と 0 からなる長さ n の数列中に含まれる 1 の個数を意味する確率変数は $K = \sum_{i=1}^n X_i$ と書ける. このことから,

$$\begin{aligned}
E[K] &= \sum_{i=1}^n E[X_i] = \sum_{i=1}^n p = np, \\
\operatorname{var}(K) &= \sum_{i=1}^n \operatorname{var}(X_i) = \sum_{i=1}^n p(1-p) = np(1-p).
\end{aligned}$$

となることがわかる.

幾何分布は成功確率 p のBernoulli試行を 1 が1つ出るまで続けたときに出た 0 の個数の分布であった. $M_1, \dots, M_k$ はそれぞれが成功確率 p の幾何分布に従う独立な確率変数であるとする. このとき, $M = \sum_{i=1}^k M_i$ はBernoulli試行を 1 が k 回出るまで続けたときに 0 が出た個数に等しい. M_i は $i-1$ 番目の 1 から i 番目の 1 が出るまでに出た 0 の個数を意味する確率変数だと解釈される. このことは, M が負の二項分布 $\mathrm{NegativeBinomial}(k, p)$ に従う確率変数になることを意味する. このことから,

$$\begin{aligned}
E[M] &= \sum_{i=1}^k E[M_i] = \sum_{i=1}^k \frac{1-p}{p} = \frac{k(1-p)}{p}, \\
\operatorname{var}(M) &= \sum_{i=1}^k \operatorname{var}(M_i) = \sum_{i=1}^k \frac{1-p}{p^2} = \frac{k(1-p)}{p^2}.
\end{aligned}$$

となることがわかる.

解答終

注意: 計算が大幅に簡単になった!

2 正規分布

$\mu, \sigma \in \mathbb{R}, \sigma > 0$ と仮定する. 確率密度函数

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

で定義される連続分布を平均 μ, 分散 σ^2 (標準偏差 σ)の **正規分布** (normal distribution)と呼び, $\mathrm{Normal}(\mu, \sigma)$ と書くのであった.

さらに, 平均 0, 分散 1^2 の正規分布を **標準正規分布** (standard normal distribution)と呼び, Normal() と書くのであった. 標準正規分布の密度函数は次の形になる:

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

2.1 正規分布のロケーションスケール変換も正規分布

$a, b \in \mathbb{R}$, $a \neq 0$ と仮定する. 一般に確率変数 X について,

$$E[aX + b] = aE[X] + b, \quad \operatorname{var}(aX + b) = a^2 \operatorname{var}(X)$$

が成立するのであった. X が正規分布に従う確率変数の場合には $aX + b$ も正規分布に従うことを示せる. 実際, $X \sim \operatorname{Normal}(\mu, \sigma)$ のとき, $y = ax + b$ すなわち $x = (y - b)/a$ とおくと,

$$\begin{aligned} E[f(aX+b)] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} f(ax+b) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} f(y) \exp\left(-\frac{((y-b)/a-\mu)^2}{2\sigma^2}\right) \frac{dy}{|a|} \\ &= \frac{1}{\sqrt{2\pi(a^2\sigma)^2}} \int_{-\infty}^{\infty} f(y) \exp\left(-\frac{(y-(a\mu+b))^2}{2a^2\sigma^2}\right) dy \end{aligned}$$

なので, $aX + b$ は平均 $a\mu + b$, 分散 $a^2\sigma^2$ の正規分布に従う: $aX + b \sim \operatorname{Normal}(a\mu + b, |a|\sigma)$.

2.2 問題: 正規分布の平均と分散

分布 $\operatorname{Normal}(\mu, \sigma)$ の平均と分散がそれぞれ μ, σ^2 になることを示せ.

解答例1: すでに標準正規分布 $\operatorname{Normal}(0, 1)$ の平均と分散がそれぞれ 0 と 1^2 になることは示してある. $X \sim \operatorname{Normal}(\mu, \sigma)$ のとき, $Z = (X - \mu)/\sigma$ とおくと $Z \sim \operatorname{Normal}((\mu - \mu)/\sigma, \sigma/\sigma) = \operatorname{Normal}(0, 1)$ となるので, $X = \sigma Z + \mu$ の平均と分散はそれぞれ $\sigma \cdot 0 + \mu = \mu$, $\sigma^2 \cdot 1^2 = \sigma^2$ になる.

解答終

解答例2: 直接的に計算してみよう. $X \sim \operatorname{Normal}(\mu, \sigma)$ のとき, 積分変数を $x = \sigma z + \mu$ と変換すると,

$$E[X] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/(2\sigma^2)}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + \mu) e^{-z^2/2}\, dz = \mu.$$

最後の等号で $\int_{-\infty}^{\infty} z e^{-z^2/2}\, dz = 0$ と $\int_{-\infty}^{\infty} e^{-z^2/2}\, dy = \sqrt{2\pi}$ を使った. この結果を使うと,

$$\operatorname{var}(X) = E[(X-\mu)^2] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-(x-\mu)^2/(2\sigma^2)}\, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sigma^2 z^2 e^{-z^2/2}\, dx = \sigma^2.$$

最後の等号で $\int_{-\infty}^{\infty} z^2 e^{-z^2/2}\, dz = \sqrt{2\pi}$ を使った. この結果は

$$\int_{-\infty}^{\infty} e^{-\alpha z^2}\, dz = \sqrt{\pi}\, \alpha^{-1/2}$$

の両辺を α で微分して -1 倍して $\alpha = 1/2$ とおいても得られるし, ガンマ函数に帰着する方法でも得られる.

解答終

2.3 問題: 正規分布に従う独立な確率変数達の和も正規分布に従う

X, Y はともに正規分布に従う独立な確率変数達であるとき, $X + Y$ も正規分布に従うことを示せ. $X + Y$ の平均と分散は X, Y の平均と分散でどのように表されるか?

解答例: $X \sim \operatorname{Normal}(\mu_X, \sigma_X)$, $Y \sim \operatorname{Normal}(\mu_Y, \sigma_Y)$ でかつ X, Y は独立であると仮定する.

X, Y は独立なので $X + Y$ の平均と分散はそれぞれ $\mu_X + \mu_Y$, $\sigma_X^2 + \sigma_Y^2$ になる.

$X' = X - \mu_X$, $Y' = Y - \mu_Y$ とおくと, $X' \sim \operatorname{Normal}(0, \sigma_X)$, $Y \sim \operatorname{Normal}(0, \sigma_Y)$, $X + Y = (X' + Y') + (\mu_X + \mu_Y)$ となるので, $X' + Y'$ が正規分布に従うことを示せば十分である. ゆえに $\mu_X = 0$, $\mu_Y = 0$ と仮定してよいので, そのように仮定する.

$X + Y$ の確率密度函数を計算して, それが正規分布の密度函数になっていることを示せばよい.

$$E[f(X+Y)] = \frac{1}{\sqrt{(2\pi)^2\sigma_X^2\sigma_Y^2}} \iint_{\mathbb{R}^2} f(x+y)\exp\left(-\frac{1}{2}\left(\frac{x^2}{\sigma_X^2}+\frac{y^2}{\sigma_Y^2}\right)\right)dx\,dy$$

であり, $x = x + y$ すなわち $y = z - x$ とおいて, x について平方完成すると,

$$\frac{x^2}{\sigma_X^2}+\frac{y^2}{\sigma_Y^2} = \frac{x^2}{\sigma_X^2}+\frac{(z-x)^2}{\sigma_Y^2} = \frac{\sigma_X^2+\sigma_Y^2}{\sigma_X^2\sigma_Y^2}\left(x-\frac{\sigma_X^2 z}{\sigma_X^2+\sigma_Y^2}\right)^2 + \frac{z^2}{\sigma_X^2+\sigma_Y^2}.$$

となる(この計算を自分で実行してみること). ゆえに,

$$\begin{aligned}
&E[f(X+Y)]\\
&= \frac{1}{\sqrt{(2\pi)^2\sigma_X^2\sigma_Y^2}}\int_{\mathbb{R}}\left(\int_{\mathbb{R}} f(z)\exp\left(-\frac{\sigma_X^2+\sigma_Y^2}{2\sigma_X^2\sigma_Y^2}\left(x-\frac{\sigma_X^2 z}{\sigma_X^2+\sigma_Y^2}\right)^2-\frac{z^2}{2(\sigma_X^2+\sigma_Y^2)}\right)dx\right)dz\\
&= \frac{1}{\sqrt{(2\pi)^2\sigma_X^2\sigma_Y^2}}\sqrt{\frac{2\pi\sigma_X^2\sigma_Y^2}{\sigma_X^2+\sigma_Y^2}}\int_{\mathbb{R}} f(z)\exp\left(-\frac{z^2}{2(\sigma_X^2+\sigma_Y^2)}\right)dz\\
&= \frac{1}{\sqrt{2\pi(\sigma_X^2+\sigma_Y^2)}}\int_{\mathbb{R}} f(z)\exp\left(-\frac{z^2}{2(\sigma_X^2+\sigma_Y^2)}\right)dz
\end{aligned}$$

以上によって $X + Y \sim \mathrm{Normal}(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$ を示せた.

$z = x + y$, $y = z - x$ とおくと密度函数中の指数函数の中身が z についての二次式になることから, $Z = X + Y$ が正規分布に従うことは計算しなくても明らかだと考えることもできる. 上のように z 以外の変数は積分して消せる. そして, $Z = X + Y$ の期待値は X, Y の期待値の和になり, X, Y が独立という仮定からそれらは無相関になるので $Z = X + Y$ の分散は X, Y の分散の和になることもわかる. これだけで $X + Y \sim \mathrm{Normal}(\mu_X+\mu_Y, \sigma_X^2+\sigma_Y^2)$ を証明できていたともみなせる.

解答終

上の解答中の最も面倒な部分の計算が正しいことはコンピュータで以下のように確認できる.

$a = \sigma_X^2$, $b = \sigma_Y^2$ とおくと,

In [3]:
```
@vars a b x z
(a+b)/(a*b)*(x - a*z/(a+b))^2 + z^2/(a+b) - x^2/a - (z-x)^2/b ▷ simplify
```

Out[3]: 0

以下のように素朴に計算することもできる.

In [4]:
```
@vars a b positive=true
@vars x y z t
expr = x^2/a + y^2/b
```

Out[4]: $\frac{y^2}{b}+\frac{x^2}{a}$

In [5]:
```
expr = expr(y⇒z-x).expand()
```

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

In [6]:
```
A, B = 1/a+1/b, z/b
expr = sympy.poly(expr(x ⇒ t + B/A), t)
```

Out[6]: $\mathrm{Poly}\left(\frac{a+b}{ab}t^2+\frac{z^2}{a+b}, t, domain=\mathbb{Z}(z,a,b)\right)$

2.4 問題: 正規分布における確率がほぼ95%または99%になる区間

$X \sim \mathrm{Normal}(\mu, \sigma)$ であると仮定する. このとき, X が区間 $[\mu - c\sigma, \mu + c\sigma]$ に含まれる確率

$$P(\mu - c\sigma \le X \le \mu + c\sigma)$$

が 95% になる c と 99% になる c を小数点以下第2桁目まで求めよ.

解答例: 標準正規分布の場合の計算に帰着することを考えよう.

$Z = (X - \mu)/\sigma$ とおくと, $Z \sim \mathrm{Normal}(0, 1)$ と Z は標準正規分布に従うのであった. このとき,

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

なので, 標準正規分布に従う確率変数 Z について $P(-c \le Z \le c)$ が 95% となる c と 99% となる c を求めればよい. 正規分布は左右対称なので, $P(-c \le Z \le c) = 1 - \alpha$ となることと, $1 - P(Z \le c) = \alpha/2$ すなわち

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

となることは同値である. (下の方にある図を見よ.)

$F(z) = P(Z \le z)$ は標準正規分布の累積分布函数である. 標準正規分布の累積分布函数はコンピュータでの基本特殊函数ライブラリに含まれている誤差函数

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

を使えば

$$F(z) = \frac{1 + \mathrm{erf}(z/\sqrt{2})}{2}.$$

と書けるのであった. ゆえに誤差函数の逆函数 $\mathrm{erfinv}(y)$ (この函数のコンピュータでの基本特殊函数ライブラリに含まれている)を使えば, 標準正規分布の累積分布函数 $F(z)$ の逆函数(分位点函数)は

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

と書ける. これを使えば $P(Z \le c) = 1 - \alpha/2$ となる c を

$$c = Q(1 - \alpha/2) = \sqrt{2}\,\mathrm{erfinv}(1 - \alpha)$$

と求めることができる.

- $1 - \alpha = 95\%$ のとき, $c = \sqrt{2}\,\mathrm{erfinv}(1 - \alpha) \approx 1.96$
- $1 - \alpha = 99\%$ のとき, $c = \sqrt{2}\,\mathrm{erfinv}(1 - \alpha) \approx 2.58$

解答終

In [7]:

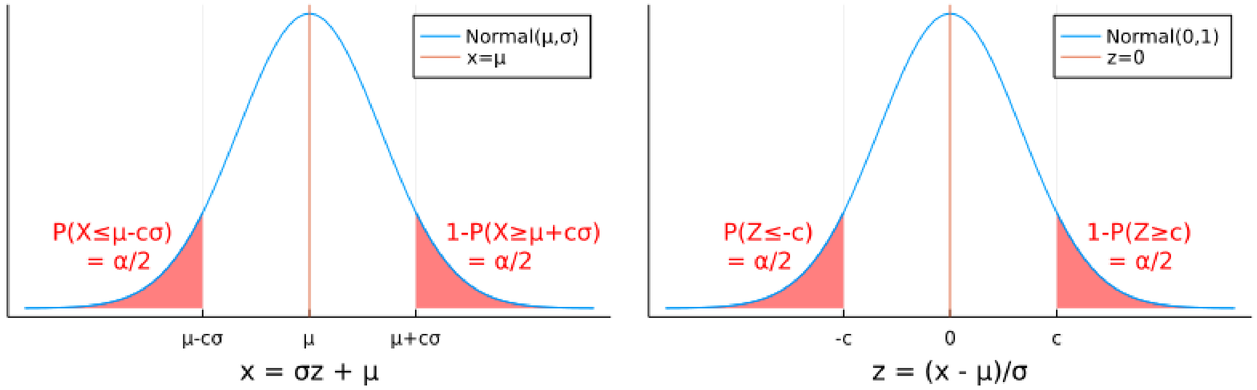
```
μ, σ, c = 2, 3, 1.5
normal = Normal(μ, σ)

P1 = plot(normal, μ - 4σ, μ + 4σ; label="Normal(μ,σ)", xlabel="x = σz + μ")
vline!([μ]; label="x=μ", xtick = ([μ-c*σ, μ, μ+c*σ], ["μ-cσ", "μ", "μ+cσ"]), ytick=false)
plot!(normal, μ - 4σ, μ - c*σ; label="", c=1, frange=0, fc=:red, fa=0.5)
plot!(normal, μ + c*σ, μ + 4σ; label="", c=1, frange=0, fc=:red, fa=0.5)
annotate!(μ - 1.3c*σ, 0.8pdf(normal, μ - c*σ), ("P(X≤μ-cσ)", 10, :red, :right))
annotate!(μ - 1.5c*σ, 0.5pdf(normal, μ - c*σ), ("= α/2", 10, :red, :right))
annotate!(μ + 1.5c*σ, 0.5pdf(normal, μ + c*σ), ("= α/2", 10, :red, :left))
annotate!(μ + 1.3c*σ, 0.8pdf(normal, μ + c*σ), ("1-P(X≥μ+cσ)", 10, :red, :left))

P2 = plot(Normal(), -4, 4; label="Normal(0,1)", xlabel="z = (x - μ)/σ")
vline!([0]; label="z=0", xtick = ([-c, 0, c], ["-c", "0", "c"]), ytick=false)
plot!(Normal(), -4, -c; label="", c=1, frange=0, fc=:red, fa=0.5)
plot!(Normal(),  c,  4; label="", c=1, frange=0, fc=:red, fa=0.5)
annotate!(-1.3c, 0.8pdf(Normal(), -c), ("P(Z≤-c)", 10, :red, :right))
annotate!(-1.5c, 0.5pdf(Normal(), -c), ("= α/2", 10, :red, :right))
annotate!( 1.5c, 0.5pdf(Normal(),  c), ("= α/2", 10, :red, :left))
annotate!( 1.3c, 0.8pdf(Normal(),  c), ("1-P(Z≥c)", 10, :red, :left))

plot(P1, P2; size=(800, 250), bottommargin=4Plots.mm)
```

Out[7]:



In [8]:

```
√2 * erfinv(0.95), quantile(Normal(), 0.975)
```

Out[8]: (1.9599639845400576, 1.9599639845400576)

In [9]:

```
√2 * erfinv(0.99), quantile(Normal(), 0.995)
```

Out[9]: (2.5758293035489053, 2.5758293035489053)

WolframAlpha:

- √2 erfinv(0.95) (https://www.wolframalpha.com/input?i=%E2%88%9A2+erfinv%280.95%29)
- √2 erfinv(0.99) (https://www.wolframalpha.com/input?i=%E2%88%9A2+erfinv%280.99%29)

2.5 問題: 対数正規分布の確率密度函数

3 モーメントとその母函数と特性函数とキュムラント母函数

3.1 モーメントとその母函数と特性函数とキュムラント母函数の定義

確率変数 X と $m = 0, 1, 2, \dots$ について

$$\mu_m = \mu_m(X) = E[X^m]$$

を m 次の **モーメント**(moment, 積率) と呼び,

$$M_X(t) = E[e^{tX}] = \sum_{m=0}^{\infty} \mu_m \frac{t^m}{m!}$$

を **モーメント母函数** (moment generating function, mgf)と呼ぶ.

X が従う確率分布の名前が Dist のとき, これらを **分布** Dist **のモーメントとモーメント母函数** と呼ぶ. 以下も同様である.

モーメント母函数の定義で t を $it = \sqrt{-1}\,t$ で置き換えたもの

$$\varphi_X(t) = E[e^{itX}] = \sum_{m=0}^\infty (i)^m \mu_m \frac{t^m}{m!}$$

を **特性函数** (characteristic function)と呼ぶ.

モーメント母函数の対数

$$K_X(t) = \log M(t) = \log E[e^{tX}] = \sum_{m=1}^\infty \kappa_m \frac{t^m}{m!}$$

を **キュムラント母函数** (cumulant generating function, cgf) と呼び, その展開係数 $\kappa_m = \kappa_m(X)$ を X の m 次のキュムラントと呼ぶ.

3.2 特性函数による期待値の表示

X が確率密度函数 $p(x)$ を持つとき, 函数 $f(x)$ のFourier変換を

$$\hat{f}(t) = \int_{-\infty}^\infty f(x)e^{-itx}\,dx$$

と書くと, $f(t)$ がそう悪くない函数ならば逆Fourier変換によってもとの函数に戻せる:

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^\infty \hat{f}(t)e^{itx}\,dt.$$

Fourier解析の基礎については次のリンク先を参照せよ(逆Fourier変換に関する結果はこのノート内では認めて使ってよい):

- 12 Fourier解析 (https://nbviewer.org/github/genkuroki/Calculus/blob/master/12%20Fourier%20analysis.ipynb)

ゆえに, x に確率変数 X を代入して両辺の期待値を取り, 期待値を取る操作と積分を交換すると,

$$E[f(X)] = \frac{1}{2\pi}\int_{-\infty}^\infty \hat{f}(t)E[e^{itX}]\,dt = \frac{1}{2\pi}\int_{-\infty}^\infty \hat{f}(t)\varphi_X(t)\,dt.$$

ここで $\varphi_X(t) = E[e^{itX}]$ は X の特性函数である.

確率変数 X が従う分布は様々な函数 $f(x)$ に関する期待値 $E[f(X)]$ から決まるので, $E[f(X)]$ が X の特性函数 $\varphi_X(t)$ を用いて表せたということは, X の特性函数 $\varphi_X(t)$ から X が従う分布が唯一つに決まることを意味している. さらに, 特性函数のレベルで近似がうまく行っていれば, 期待値でも近似がうまく行くこともわかる.

3.3 問題: 標準正規分布のモーメント母函数と特性函数とキュムラント母函数

モーメント母函数 $M_X(t) = E[e^{tX}]$ が t の函数として, 定義域を自然に複素数まで拡張できているとき(正確には解析接続できていれば), $\varphi_X(t) = M_X(it)$ が成立する. この事実を用いて, 標準正規分布のモーメント母函数と特性函数を求め, さらにキュムラント母函数を求めよ.

解答例: $Z \sim \mathrm{Normal}(0,1)$ と仮定する. このとき, $z = w + t$ とおくと

$$M_Z(t) = E[e^{tZ}] = \int_{-\infty}^\infty e^{tz}\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-(z-t)^2/2+t^2/2}\,dz = \frac{e^{t^2/2}}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{-w^2/2}\,dw = e^{t^2/2}.$$

ゆえに,

$$\varphi_Z(t) = E[e^{itZ}] = M_Z(it) = e^{-t^2/2}, \quad K_Z(t) = \log M_Z(t) = \frac{t^2}{2}.$$

解答終

注意: 標準正規分布のキュムラント母函数は $t^2/2$ というたったの一項だけになってしまう. 標準でない正規分布のキュムラント母函数は t について1次と2次の項だけになる. 1次の項の係数は分布の期待値で, $t^2/2$ の係数は分散になる. 実際, 平均 μ, 分散 σ^2 を持つ確率変数 X について,

$$
\begin{aligned}
K_X(t) &= \log E[e^{tX}] = \log E[e^{t(X-\mu)+t\mu}] \\
&= t\mu + \log\left(1 + E[X-\mu]t + E[(X-\mu)^2]\frac{t^2}{2} + O(t^3)\right) \\
&= t\mu + \log\left(1 + \sigma^2\frac{t^2}{2} + O(t^3)\right) = t\mu + \sigma^2\frac{t^2}{2} + O(t^3).
\end{aligned}
$$

そして, $\sigma Z + \mu \sim \mathrm{Normal}(\mu, \sigma)$ であり,

$$
M_{\sigma Z+\mu}(t) = E[e^{t(\sigma Z+\mu)}] = e^{\mu t} E[e^{t\sigma Z}] = e^{\mu t} M_Z(\sigma t) = e^{\mu t+\sigma^2 t^2/2}.
$$

ゆえに

$$
K_{\sigma Z+\mu}(t) = \log M_{\sigma Z+\mu}(t) = \mu t + \sigma^2\frac{t^2}{2}.
$$

他の分布のキュムラント母函数を計算したときに出て来る t について3次以上の項はその分布が正規分布からどれだけ離れているかを表している.

3.4 中心極限定理の特性函数を使った証明

中心極限定理: $X_1, X_2, X_3, \dots$ が独立同分布な確率変数の列であるとき, $\mu = E[X_k]$ が定義されていて, $\sigma^2 = \mathrm{var}(X_k) = E[(X_k-\mu)^2] < \infty$ でかつ, $E[|X_k-\mu|^3] < \infty$ で $E[(X_k-\mu)^3]$ が定義されているとき,

$$
\bar{X}_n = \frac{1}{n}\sum_{k=1}^n X_k, \quad Z_n = \frac{\sqrt{n}\,(\bar{X}_n - \mu)}{\sigma}
$$

とおくと, $n \to \infty$ で Z_n の分布は標準正規分布に近付く.

証明: $Y_k = (X_k-\mu)/\sigma$ とおく. $Y_1, Y_2, \dots$ も独立同分布になり, $E[Y_k] = 0, \quad E[Y_k^2] = 1$ が成立している. ゆえに Y_k の特性函数 $\varphi(t)$ は k によらず,

$$
\varphi(t) = E[e^{itY_k}] = 1 + iE[Y_k]t - E[Y_k^2]\frac{t^2}{2} + O(t^3) = 1 - \frac{t^2}{2} + O(t^3).
$$

そして,

$$
\frac{1}{\sqrt{n}}\sum_{k=1}^n Y_k = \frac{\sqrt{n}}{\sigma}\frac{1}{n}\sum_{k=1}^n (X_k - \mu) = \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) = Z_n.
$$

なので, Z_n の特性函数は, $n \to \infty$ で

$$
\begin{aligned}
\varphi_{Z_n}(t) &= E\left[\exp\left(it\frac{1}{\sqrt{n}}\sum_{k=1}^n Y_k\right)\right] = E\left[\prod_{k=1}^n \exp\left(i\frac{t}{\sqrt{n}}Y_k\right)\right] \\
&= \prod_{k=1}^n E\left[\exp\left(i\frac{t}{\sqrt{n}}Y_k\right)\right] = \varphi\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 - \frac{t^2}{2n} + O(n^{-3/2})\right)^n \to e^{-t^2/2}.
\end{aligned}
$$

と標準正規分布の特性函数 $e^{-t^2/2}$ に収束する. (3つめの等号で $Y_1, \dots, Y_n$ の独立性を使った.)

ゆえに Z_n の分布は $n \to \infty$ で標準正規分布に近付く.

証明終

3.5 歪度と中心極限定理の収束の速さ

前節の微論より, $X_1, X_2, \dots$ を独立同分布確率変数列とし,

$$
\mu = E[X_k], \quad \sigma^2 = E[(X-\mu)^2], \quad Y_k = \frac{X_k - \mu}{\sigma}, \quad \varphi(t) = E[e^{itY_k}]
$$

とおいたときの, $n \to \infty$ での $\varphi(t/\sqrt{n})^n \to e^{-t^2/2}$ の収束の速さを調べれば, 中心極限定理による正規分布への収束の速さがわかる.

$$
\mu'_3 = E[Y_k^3] = E\left[\left(\frac{X_k - \mu}{\sigma}\right)^3\right]
$$

とおくと,

$$\varphi\left(\frac{t}{\sqrt{n}}\right) = E[e^{itY_k}] = 1 - \frac{t^2}{2n} - i\mu_3\frac{t^3}{6n\sqrt{n}} + O(n^{-2})$$

なので,

$$\log\varphi\left(\frac{t}{\sqrt{n}}\right)^n = n\log\left(1 - \frac{t^2}{2n} - i\mu_3'\frac{t^3}{6n\sqrt{n}} + O(n^{-2})\right) = -\frac{t^2}{2} - i\mu_3\frac{t^3}{6\sqrt{n}} + O(n^{-1}).$$

これは $n \to \infty$ での $\log\varphi\left(t/\sqrt{n}\right)^n \to -t^2/2$ の収束の速さは, $Y_k = (X_k - \mu)/\sigma$ の3次のモーメント μ_3' の絶対値の大きさであることがわかる. μ_3' の絶対値が小さいほど収束が速く, 大きいほど収束が遅い.

μ_3' は X_k の分布の期待値 μ を中心とする非対称性の σ によって適切に正規化した尺度になっている. μ_3' は $Y_k == (X_k - \mu)/\sigma$ の3次のキュムラントにも一致している:

$$K(t) = \log E[e^{tY_k}] = \left(1 + \frac{t^2}{2} + \mu_3'\frac{t^3}{3!} + O(t^4)\right) = \frac{t^2}{2} + \mu_3'\frac{t^3}{3!} + O(t^4).$$

この μ_3' は X_k が従う分布(同じことだが Y_k が従う分布)の **歪度** (skewness) と呼ばれている.

確率変数 X が従う分布の **歪度** (skewness) $\operatorname{skewness}(X)$ を次のように定める:

$$\operatorname{skewness}(X) = \kappa_3'(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right].$$

ここで $\mu = E[X]$, $\sigma^2 = E[(x-\mu)^2]$. 期待値を中心に左右対称な分布の歪度は0になる. 歪度はロケーションスケール変換 $X \mapsto aX + b$ $(a \neq 0)$ で不変である.

3.6 標準化キュムラントと歪度と尖度

確率変数 X は確率変数であるとし, $\mu = E[X]$, $\sigma = \sqrt{E[(X-\mu)^2]}$ とおく. このとき,

$$Z = \frac{X-\mu}{\sigma}$$

を確率変数の **標準化** (standardization)と呼ぶ. Z の期待値と分散はそれぞれ 0 と 1 になる.

X の標準化のモーメントやキュムラントをそれぞれ **標準化モーメント**, **標準化キュムラント** と呼び, 以下のように表す:

$$\begin{aligned}
\mu_m' &= E\left[\left(\frac{X-\mu}{\sigma}\right)^m\right], \\
E\left[\exp\left(t\frac{X-\mu}{\sigma}\right)\right] &= \sum_{m=0}^\infty \mu_m'(X)\frac{t^m}{m!} = 1 + \frac{t^2}{2} + \mu_3'(X)\frac{t^3}{3!} + \mu_4'(X)\frac{t^4}{4!} + \cdots, \\
\log E\left[\exp\left(t\frac{X-\mu}{\sigma}\right)\right] &= \sum_{m=1}^\infty \kappa_m'(X)\frac{t^m}{m!} = 1 + \frac{t^2}{2} + \kappa_3'(X)\frac{t^3}{3!} + \kappa_4'(X)\frac{t^4}{4!} + \cdots.
\end{aligned}$$

$\kappa_3'(X)$ と $\kappa_4'(X)$ は次のように表される:

$$\kappa_3'(X) = \mu_3'(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right], \quad \kappa_4'(X) = \mu_4'(X) - 3 = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] - 3.$$

X が正規分布に従う確率変数の場合には $\kappa_m' = \delta_{m2}$ (特に $m = 2$ 以外の場合に 0)となる.

κ_3' を X もしくは X が従う分布の **歪度** (わいど, skewness) と呼び, κ_4' を **尖度** (せんど, kurtosis)と呼ぶことにする. 歪度は左右の非対称性の尺度であり, 尖度は分布の尖り方の尺度になっている.

$\kappa_4' = \mu_4' - 3$ ではなく, 3 を引く前の μ_4' を尖度と定義する流儀もあるが, このノートでは **正規分布の尖度を 0 にしたいので, 3 を引いた側の κ_4' を尖度の定義として採用する.**

In [10]:
```
@vars t μ3 μ4 μ5 κ3 κ4 κ5
Mt = 1 + t^2/2 + μ3*t^3/6 + μ4*t^4/24 + μ5*t^5
expr = series(log(Mt), t)
```

Out[10]: $\frac{t^2}{2} + \frac{t^3\mu_3}{6} + t^4\left(\frac{\mu_4}{24} - \frac{1}{8}\right) + t^5\left(-\frac{\mu_3}{12} + \mu_5\right) + O\left(t^6\right)$

3.7 標本平均と不偏分散の定義

3.8 問題: 標本平均の期待値と分散

3.9 問題: 不偏分散の期待値と分散

4 Poisson分布

5 ガンマ分布

6 χ^2分布

7 ベータ分布

8 時間連続極限

In [ ]:
```
1
```