

12. 統計分布モデル (2) 連続分布

Models of statistical distribution (2)

Continuous distribution

この章では、連続的な確率変数を持つ統計分布モデルとして、**連続一様分布**と**正規分布**、**ガンマ分布**、その他に実験データや観測データを解析するときに用いられる場合のある統計分布モデル（**カイ自乗分布**、**コーシー分布**、**対数正規分布**、**ロジック・ラムラー分布**、**スチューデント t 分布**、**ロジスティック分布**）について述べます。

12-1 連続一様分布 Continuous uniform distribution

いちよう コンティニューアス ユニフォーム ディストリビューション
連続一様分布 continuous uniform distribution の確率密度関数 $f(x; a, b)$ は、

$$f(x) = \begin{cases} \frac{1}{b-a} & [a \leq x \leq b] \\ 0 & [x < a, b < x] \end{cases} \quad (12.1.1)$$

のような式で表現できます。ただし、 $a < b$ とします。[Fig. 12.1.1](#) に連続一様分布の確率密度関数の例を示します。

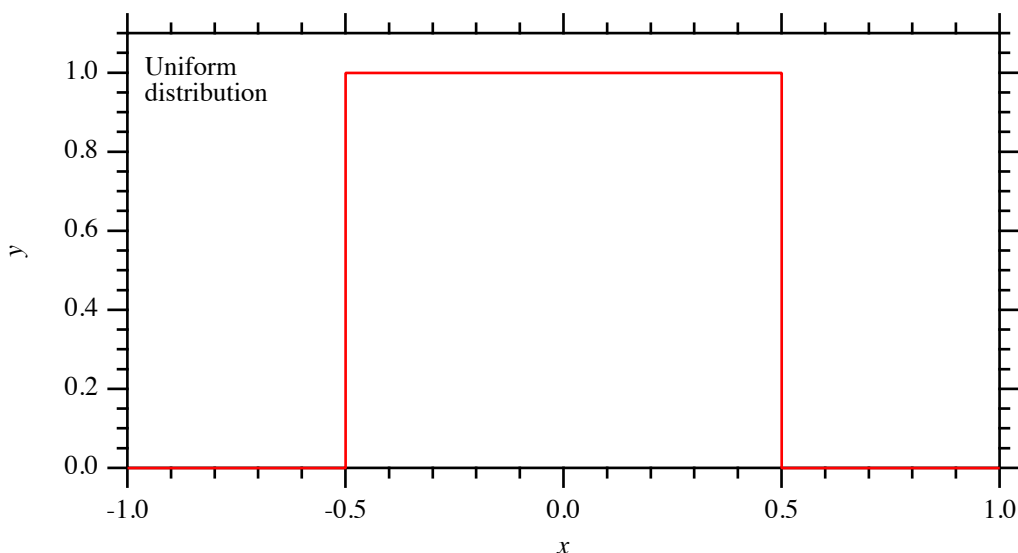


Fig. 12.1.1 連続一様分布の確率密度関数。 $a = -0.5, b = 0.5$ の場合

連続一様分布の**平均** $\langle x \rangle$ は

$$\langle x \rangle = \int_{-\infty}^{\infty} x f(x; a, b) dx = \int_a^b \frac{x dx}{b-a} = \left[\frac{x^2}{2(b-a)} \right]_a^b = \frac{a+b}{2} \quad (12.1.2)$$

となります。

連続一様分布の平方平均 $\langle x^2 \rangle$ は

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 f(x; a, b) dx = \int_a^b \frac{x^2 dx}{b-a} = \left[\frac{x^3}{3(b-a)} \right]_a^b = \frac{a^2 + ab + b^2}{3} \quad (12.1.3)$$

となることと、式 (12.1.2) とから連続一様分布の**分散** σ^2 は

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} = \frac{a^2 - 2ab + b^2}{12} = \frac{(b-a)^2}{12} \quad (12.1.4)$$

となります。

要素数 n 、間隔 d の離散一様分布では分散が $(n^2 - 1)d^2/12$ と表されます。下限を $a = a_1$ 、上限を $b = a_1 + (n-1)d$ とすれば、 $d = (b-a)/(n-1)$ となり、離散一様分布の分散は $\frac{(n+1)(b-a)^2}{12(n-1)}$ となり、 $n \rightarrow \infty$ ($d \rightarrow 0$) のとき、式 (12.1.4) で表される連続一様分布の分散と一致します。

左右対称な形状を持つ確率密度関数の「3階以上の奇数階キュムラント」はすべてゼロになります。連続一様分布をグラフ化した時の形状は左右対称な（平均位置を対称軸とする線対称性を持つ）ので、連続一様分布関数の3階キュムラントも**歪度** skewness もゼロです。

連続一様分布の**尖度** kurtosis (excess kurtosis) は

$$\gamma_2 \equiv \frac{\kappa_4}{\sigma^4} = -\frac{6}{5} \quad (12.1.5)$$

となり、連続一様分布 $f(x; a, b)$ の**4階キュムラント**は、

$$\kappa_4 = \gamma_2 \sigma^4 = -\frac{6}{5} \times \left[\frac{(b-a)^2}{12} \right]^2 = -\frac{(b-a)^4}{120} \quad (12.1.6)$$

と表されます。 ([補足 12.1.A](#))

連続一様分布の**中央値** x_{median} も**ミッドレンジ** $x_{\text{mini-max}}$ も一意に確定し、いずれも平均値と一致します。

連続一様分布の**最頻値** x_{mode} は $a \leq x \leq b$ の範囲の任意の値をとり、一つの数値には決まりません。

12-2 正規分布 Normal distribution

ノーマル ディストリビューション 正規分布 normal distribution ガウシアン ディストリビューション あるいは**ガウス型分布 Gaussian distribution** と呼ばれる確率分布モデルは、実験データを解析する場合に最も頻繁に用いられるものです。独立でランダムな確率変数を足し合わせる操作をしていくと、個々の変数の確率分布が正規分布でなくても、その和として表される変数の確率分布（ひんぱん畳込 convolution）は（多くの場合に）正規分布に近づくことが知られており、このことは**中心極限定理 central limit theorem** セントラル リミット セオレム と呼ばれます（[補足 12.2.A](#)）。

実験で得られる観測値が複数のランダムな誤差要因によって真の値からずれるとき、「誤差の要因が多ければ、その和として表される誤差の分布は正規分布に近くなることが予想される」ことを根拠として、実験データの解析に正規分布の仮定される場合があります。

平均 μ (μ はギリシャ文字のミュー)、標準偏差 σ (σ はギリシャ文字のシグマ) の正規分布の**確率密度関数**は、以下の式：

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right] \quad (12.2.1)$$

で与えられます。

$$\int_{-\infty}^{\infty} f(x; \mu, \sigma) dx = 1 \quad (12.2.2)$$

の関係は常に成立します。正規分布の確率密度関数は左右対称な形状であり、**3階キュムラント** κ_3 と**歪度** γ_1 はゼロになります。

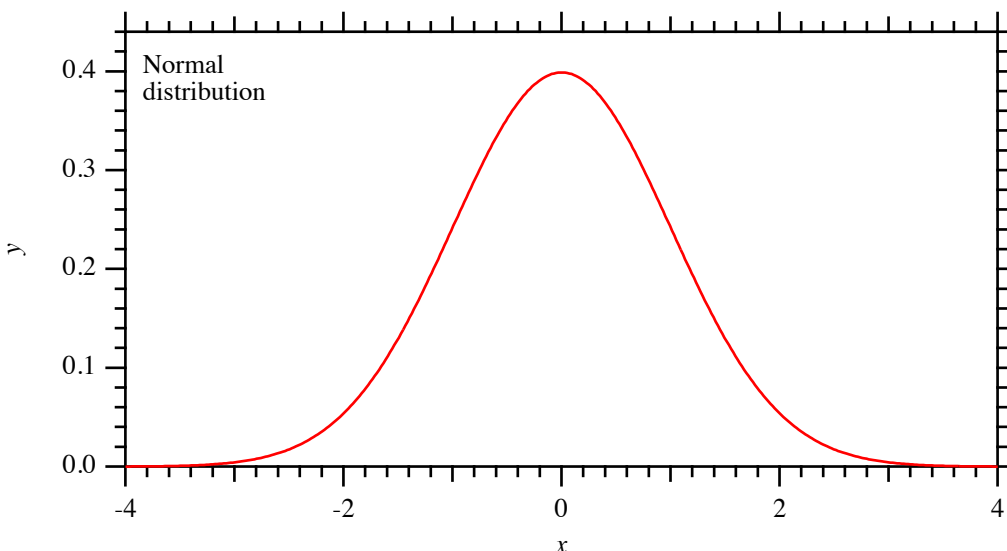


Fig. 12.2.1 平均 0, 標準偏差 1 の正規分布の確率密度関数

平均 0, 標準偏差 1 の正規分布は**標準正規分布 standard normal distribution** スタンダード ノーマル ディストリビューション と呼ばれます（[Fig. 12.2.1](#)）。標準正規分布の確率密度関数 $f_{\text{SND}}(x)$ は、

$$f_{\text{SND}}(x) = f(x; 0, 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (12.2.3)$$

と表されます。 $f_{\text{SND}}(x)$ を使えば一般的な正規分布の確率密度関数を以下のように表せます。

$$f(x; \mu, \sigma) = \frac{1}{\sigma} f_{\text{SND}}\left(\frac{x - \mu}{\sigma}\right) \quad (12.2.4)$$

標準正規分布の4階^{べき}平均 $\langle x^4 \rangle$ は、

$$\begin{aligned} \langle x^4 \rangle &= \int_{-\infty}^{\infty} x^4 \phi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 \exp\left(-\frac{x^2}{2}\right) dx && \text{(部分積分をする→)} \\ &= \frac{1}{\sqrt{2\pi}} \left\{ \left[-x^3 \exp\left(-\frac{x^2}{2}\right) \right]_{-\infty}^{\infty} + 3 \int_{-\infty}^{\infty} x^2 \exp\left(-\frac{x^2}{2}\right) dx \right\} \\ &&& \text{(第1項はゼロ, さらに部分積分→)} \\ &= \frac{3}{\sqrt{2\pi}} \left\{ \left[-x \exp\left(-\frac{x^2}{2}\right) \right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx \right\} \\ &&& \text{(第1項はゼロ, さらに部分積分→)} \\ &= \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx && \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx = 1 \text{ だから→} \right) \\ &= 3 && (12.2.3) \end{aligned}$$

となることから、標準正規分布も一般的な正規分布も**尖度**は

$$\gamma_2 = \langle x^4 \rangle - 3\sigma^4 = 3 - 3 \times 1^2 = 0 \quad (12.2.4)$$

となり、**4階キュムラント**もゼロになります。正規分布では3階以上の階数のキュムラントがすべてゼロになります ([補足 12.2.B](#))。

正規分布の**最頻値**^{さいひんち} mode は平均 μ に一致します。

標準正規分布の**累積分布関数**^{キュミュラティヴ デイトリビューション ファンクション} cumulative distribution function $F_{\text{SND}}(x)$ は

$$F_{\text{SND}}(x) = \int_{-\infty}^x f_{\text{SND}}(t) dt = \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right] \quad (12.2.5)$$

と表されます。ここで関数 $\operatorname{erf}(x)$ は**誤差関数** error function と呼ばれ、以下の式で定義されます。

$$\operatorname{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (12.2.6)$$

誤差関数 $\text{erf}(x)$ は以下の性質を持ちます。

$$\text{erf}(\pm\infty) = \pm 1 \quad (12.2.7)$$

$$\text{erf}(0) = 0 \quad (12.2.8)$$

標準正規分布の累積分布関数 $F_{\text{SND}}(x)$ は

$$F_{\text{SND}}(-\infty) = 0 \quad (12.2.9)$$

$$F_{\text{SND}}(0) = \frac{1}{2} \quad (12.2.10)$$

$$F_{\text{SND}}(\infty) = 1 \quad (12.2.11)$$

となることから、標準正規分布の中央値は 0 となり、一般的な正規分布の**中央値**は平均 μ と一致します。

以下の式で定義される^ほ補誤差関数（^{そうほ}相補誤差関数）^{コンプリメンタリ} complementary error function $\text{erfc}(x)$ の使われる場合があります。

$$\text{erfc}(x) \equiv 1 - \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt \quad (12.2.12)$$

^{かり}仮に観測値の出現確率が平均 μ 、標準偏差 σ の正規分布に従うとすると、観測値が $(\mu \pm \sigma)$ 、 $(\mu \pm 2\sigma)$ 、 $(\mu \pm 3\sigma)$ 、 $(\mu \pm 4\sigma)$ の範囲に収まる確率は

$$F_{\text{SND}}(1) - F_{\text{SND}}(-1) = \text{erf}\left(\frac{1}{\sqrt{2}}\right) = 0.682689\dots \quad (12.2.13)$$

$$F_{\text{SND}}(2) - F_{\text{SND}}(-2) = \text{erf}\left(\frac{2}{\sqrt{2}}\right) = 0.954997\dots \quad (12.2.14)$$

$$F_{\text{SND}}(3) - F_{\text{SND}}(-3) = \text{erf}\left(\frac{3}{\sqrt{2}}\right) = 0.997300\dots \quad (12.2.15)$$

$$F_{\text{SND}}(4) - F_{\text{SND}}(-4) = \text{erf}\left(\frac{4}{\sqrt{2}}\right) = 0.999936\dots \quad (12.2.16)$$

などとなります（[補足 12.2.C](#)）。このことは、表計算ソフトウェアなどには標準的に装備されている誤差関数 $\text{erf}(x)$ を使えば確認することができます。

12-3 ガンマ分布 Gamma distribution

ガンマ分布 Gamma distribution（^{ガンマ} Γ 分布）の**確率密度関数**は以下の式で表されます。

$$f(x; \theta, \alpha) = \frac{1}{\Gamma(\alpha)\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} \exp\left(-\frac{x}{\theta}\right) \quad (12.3.1)$$

ただし、 $\alpha > 0$ 、 $\theta > 0$ とします。ガンマ分布を表現するためには、式 [\(12.3.1\)](#) の表現を使い、「ただし確率変数 x は正の値のみをとるものとする」ことが普通です。

確率変数 x のとりうる範囲に制限を加えるより、ガンマ分布の確率密度関数を

$$f(x; \theta, \alpha) = \begin{cases} \frac{1}{\Gamma(\alpha)\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} \exp\left(-\frac{x}{\theta}\right) & [0 < x] \\ 0 & [x \leq 0] \end{cases} \quad (12.3.2)$$

と表現する方が明確かもしれません。

式 (12.3.2) 中の $\Gamma(\alpha)$ は ガンマ関数 **ガンマ関数** Gamma function と呼ばれる特殊関数で、通常は

$$\Gamma(\alpha) \equiv \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad (12.3.3)$$

という定積分の形式で定義されます。ガンマ関数は、普通の表計算ソフトウェアなどには標準的に装備されます。ガンマ関数は階乗 $n!$ を整数以外に拡張した意味を持つ関数で、 n が自然数 (0 以上の整数) の時には

$$\Gamma(n+1) = n! \quad (12.3.4)$$

の関係があります。例えば $\Gamma(1) = 1$, $\Gamma(2) = 1$, $\Gamma(3) = 2$, $\Gamma(4) = 6$, $\Gamma(5) = 24$, ... となります。

また,

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha) \quad (12.3.5)$$

の関係が成立します。

ガンマ分布の**平均** m は

$$\begin{aligned} m &= \int_0^{\infty} x f(x; \theta, \alpha) dx = \frac{1}{\Gamma(\alpha)\theta} \int_0^{\infty} x \left(\frac{x}{\theta}\right)^{\alpha-1} \exp\left(-\frac{x}{\theta}\right) dx \\ &= \frac{\Gamma(\alpha+1)\theta}{\Gamma(\alpha)} \frac{1}{\Gamma(\alpha+1)\theta} \int_0^{\infty} \left(\frac{x}{\theta}\right)^{\alpha} \exp\left(-\frac{x}{\theta}\right) dx \\ &= \frac{\Gamma(\alpha+1)\theta}{\Gamma(\alpha)} \int_0^{\infty} f(x; \theta, \alpha+1) dx \\ &\quad \left(\text{式 (12.3.5) と } \int_{-\infty}^{\infty} f(x; \alpha) dx = 1 \text{ が任意の } \alpha \text{ に対して成立することから} \rightarrow \right) \\ &= \alpha\theta \end{aligned} \quad (12.3.6)$$

となります。同じようにガンマ分布の平方平均は

$$\langle x^2 \rangle = \alpha(\alpha+1)\theta^2 \quad (12.3.7)$$

となり、**分散** σ^2 は

$$\sigma^2 = \langle x^2 \rangle - m^2 = \alpha\theta^2 \quad (12.3.8)$$

となります。

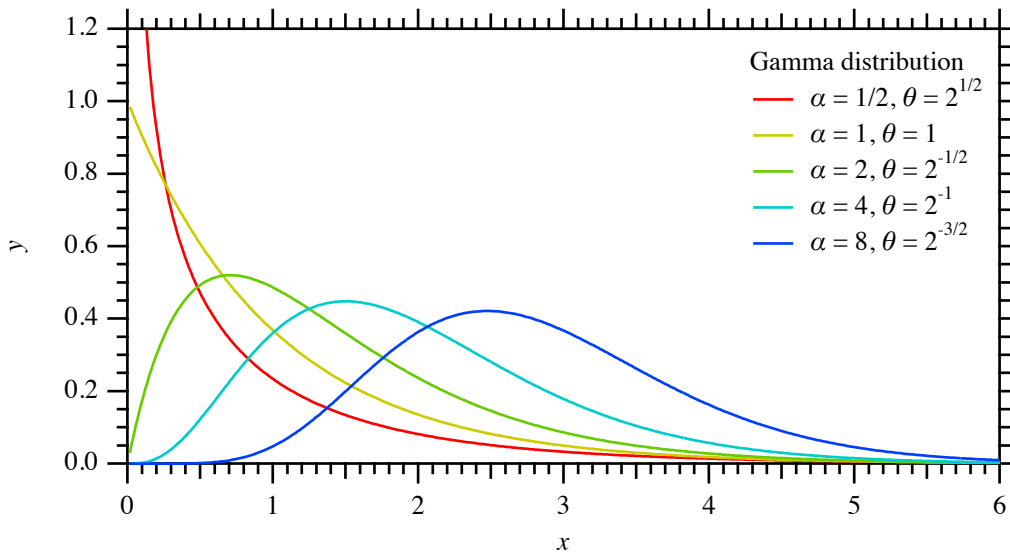


Fig. 12.3.1 標準偏差 1 のガンマ分布の確率密度関数

標準偏差 $\alpha\theta^2 = 1$ の一連のガンマ分布の確率密度関数を [Fig. 12.3.1](#) に示します。

ガンマ分布の立方平均は

$$\langle x^3 \rangle = \alpha(\alpha + 1)(\alpha + 2)\theta^3 \quad (12.3.9)$$

であり、**3階キュムラント** κ_3 と **歪度** γ_1 は、

$$\kappa_3 = \langle x^3 \rangle - 3\langle x^2 \rangle m + 2m^3 = 2\alpha\theta^3 \quad (12.3.10)$$

$$\gamma_1 = \frac{2}{\sqrt{\alpha}} \quad (12.3.11)$$

となります。 α が増えればガンマ分布の歪度はゼロに近づき、左右対称な形状に近づきます。

ガンマ分布の**4階キュムラント** κ_4 と **尖度** γ_2 は、

$$\kappa_4 = 6\alpha\theta^4 \quad (12.3.12)$$

$$\gamma_2 = \frac{6}{\alpha} \quad (12.3.13)$$

となり、 λ が増えれば尖度もゼロに近づきます。

一般的にガンマ分布の確率密度関数の k 階キュムラントは

$$\kappa_k = (k - 1)! \alpha \theta^k \quad (12.3.14)$$

と表されます ([補足 12.3.A](#))。

$$\begin{aligned} \frac{d}{dx} f(x; \theta, \alpha) &= \frac{1}{\Gamma(\alpha)\theta} \left[\frac{\alpha - 1}{\theta} \left(\frac{x}{\theta}\right)^{\alpha-2} \exp\left(-\frac{x}{\theta}\right) - \frac{1}{\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} \exp\left(-\frac{x}{\theta}\right) \right] \\ &= \frac{1}{\Gamma(\alpha)\theta} \frac{(\alpha - 1)\theta - x}{\theta^2} \left(\frac{x}{\theta}\right)^{\alpha-2} \exp\left(-\frac{x}{\theta}\right) \end{aligned} \quad (12.3.15)$$

の関係から、ガンマ分布の^{さいひんち}最頻値 x_{mode} は、

$$x_{\text{mode}} = \begin{cases} 0 & [0 < \alpha \leq 1] \\ (\alpha - 1)\theta & [1 < \alpha] \end{cases} \quad (12.3.16)$$

となり、一つの値には決まりますが、 $\alpha > 1$ の時にしか意味はありません。

ガンマ分布の^{ミディアン}中央値 (中位数, メジアン) x_{median} は、必要であれば

$$x_{\text{median}} = \theta P^{-1}(\alpha, 1/2) \quad (12.3.17)$$

として計算できます。ただし $P^{-1}(\alpha, y)$ は、

$$P(\alpha, x) \equiv \frac{1}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t} dt \quad (12.3.18)$$

として定義されるガンマ分布の^{るい}累積分布関数 (「正則化された不完全ガンマ函数」
レギュラライズド インコンプリート
regularized incomplete Gamma function と呼ばれる) $P(\alpha, x)$ の逆函数とします。

「正則化された不完全ガンマ函数の逆函数」は、装備されている計算システムも装備されていない計算システムもあります。ガンマ分布の中央値を求めること自体にはあまり意味がないかもしれません。

^{カイ スクウェアド ディストリビューション} **カイ自乗分布** ^{カイ} chi-squared distribution (chi はギリシヤ文字 χ の英語表記; 「 χ^2 -分布」と書かれる場合もある) は、ガンマ分布のパラメータ α が半整数値あるいは整数値

($\alpha = 1/2, 1, 3/2, 2, 5/2, \dots$) の特殊な値をとり、 $\theta = 1$ とした場合のもので、 $\alpha = 1/2$ の場合のガンマ分布の確率密度函数として

$$f\left(x; 1, \frac{1}{2}\right) = \begin{cases} \frac{1}{\sqrt{\pi}} x^{-1/2} \exp(-x) & [0 < x] \\ 0 & [x \leq 0] \end{cases} \quad (12.3.19)$$

で表される分布が「^{ディグリー オヴ フリーダム}自由度 (degree of freedom) 1 のカイ自乗分布」と呼ばれ、 $\alpha = 1$ の場合の確率密度函数

$$f(x; 1, 1) = \begin{cases} \exp(-x) & [0 < x] \\ 0 & [x \leq 0] \end{cases} \quad (12.3.20)$$

で表される確率分布が「自由度 2 のカイ自乗分布」, ... などと呼ばれます。自由度 1 のカイ自乗分布は、「標準正規分布に従う確率変数 x の自乗 $y = x^2$ を確率変数とする確率分布」に対応します ([補足 12.3.B](#)) 。

式 (12.3.2) で表されるガンマ分布のうち、 $\alpha = 1$ の場合

$$f(x; \theta, 1) = \begin{cases} \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right) & [0 < x] \\ 0 & [x \leq 0] \end{cases} \quad (12.3.21)$$

が指数分布 エクスポネンシャル ディストリビューション exponential distribution と呼ばれます。また、式 (12.3.21) で表される指数分布について $\theta = 1$ とすれば「自由度 2 のカイ自乗分布」に一致することになります。

12-4 その他の分布 Other distributions

以下、実験データを解析するために用いられる場合のある確率分布として、コーシー分布（ローレンツ型関数）、対数正規分布、ロジック・ラムラー分布（ワイブル分布）、学生 t 分布、ロジスティック分布とをあげます。

12-4-1 コーシー分布 Cauchy distribution

中央値あるいは最頻値 ひん をゼロとするコーシー分布 コーシー ディストリビューション Cauchy distribution（ローレンツ型分布 ローレンツィアン Lorentzian distribution）の確率密度関数は、以下の式で表されます。

$$f(x; w) = \frac{1}{\pi w} \left(1 + \frac{x^2}{w^2} \right)^{-1} \quad (12.4.1.1)$$

単一の緩和過程を持つ（指数函数的に減衰するような）共鳴現象（例えば空気抵抗のみによって振幅が減衰する振り子の運動など）の周波数スペクトルはコーシー分布（ローレンツ型分布）に従います。分光実験ではローレンツ型のピーク形状がよく現れます。

コーシー分布には、平均を含む 1 階以上のキュムラントがすべて不定であるという特徴があります。

コーシー分布の確率密度関数（ローレンツ型関数）の形状を [Fig. 12.4.1.1](#) に示します。

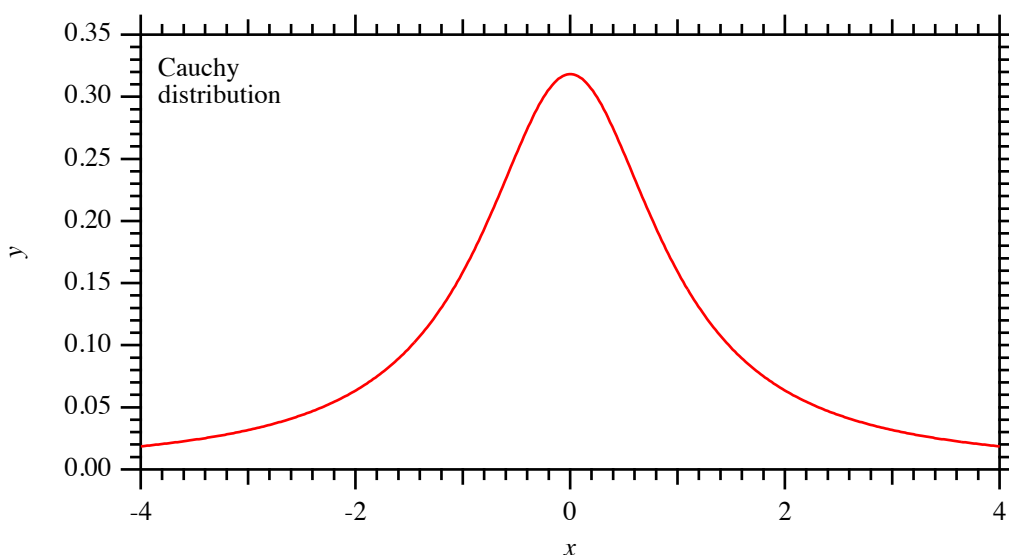


Fig. 12.4.1.1 半値半幅 $w = 1$ のコーシー分布の確率密度関数

コーシー分布の累積分布関数は以下の式で表されます。

$$F(x; w) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{x}{w} \quad (12.4.1.2)$$

1 2-4-2 対数正規分布 Lognormal distribution

対数正規分布 lognormal distribution は、粉体やコロイドの中の粒子の大きさの分布をモデル化する場合などに使われます。対数正規分布の**確率密度関数**は、以下の式で表されます。

$$f(x; m, \omega) = \frac{1}{\sqrt{2\pi}\omega x} \exp \left[-\frac{(\ln x - \ln m)^2}{2\omega^2} \right] \quad (12.4.2.1)$$

正規分布の確率変数 x をその対数 $\ln x$ に置き換えた形式になっており、 $d \ln x = dx/x$ の関係から、係数に $1/x$ のかかる形になっていることに注意してください

([補足 12.4.2.A](#))。

式 ([12.4.2.1](#)) 中の $\ln m$, ω には平均、標準偏差という意味はありません。式 ([12.4.2.1](#)) 中の m は**中央値 median** に相当する値であり、 ω を対数標準偏差 logarithmic standard deviation と呼ぶことがあります。式 ([12.4.2.1](#)) で表される対数正規分布の**平均** $\langle x \rangle$ と平方平均 $\langle x^2 \rangle$, **分散** σ^2 は

$$\langle x \rangle = m \exp \left(\frac{\omega^2}{2} \right) \quad (12.4.2.2)$$

$$\langle x^2 \rangle = m^2 \exp (2\omega^2) \quad (12.4.2.2)$$

$$\sigma^2 = m^2 \exp (\omega^2) [\exp (\omega^2) - 1] \quad (12.4.2.3)$$

という値を取ります。また、対数正規分布の確率変数 x の**冪乗**^{べきじょう}の平均 $\langle x^k \rangle$ は

$$\langle x^k \rangle = m^k \exp \left(\frac{k^2 \omega^2}{2} \right) \quad (12.4.2.4)$$

という値をとります。

対数正規分布の確率密度関数の形状を [Fig. 12.4.2.1](#) に示します。

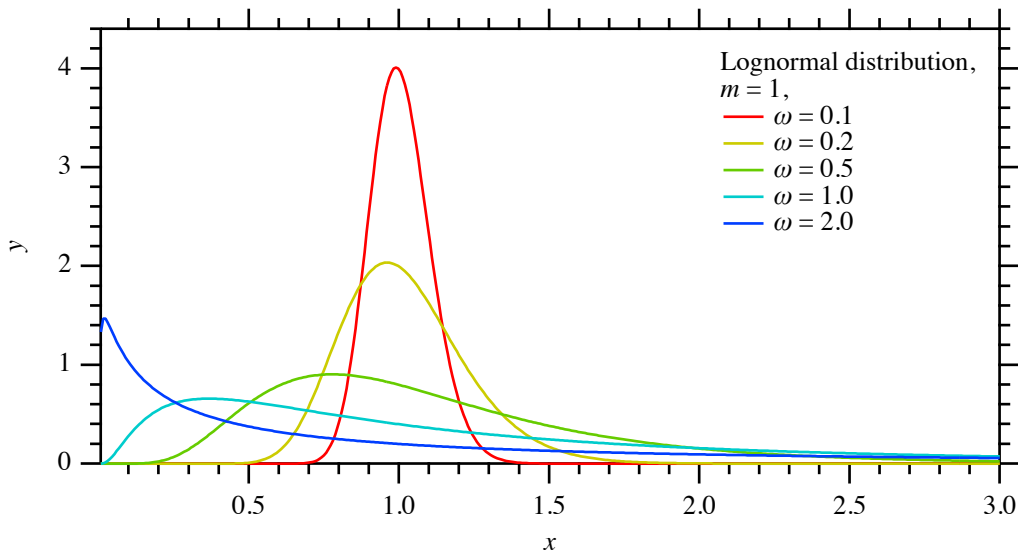


Fig. 12.4.2.1 メジアン $m = 1$ の対数正規分布の確率密度関数

対数正規分布の**累積分布関数**は

$$F(x; m; \omega) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\ln x - \ln m}{\sqrt{2\omega}} \right) \right] \quad (12.4.2.5)$$

と表されます。

対数正規分布は、実際に粉体やコロイドの中の粒子のサイズの分布を良く模擬できるように見える場合が少なくありません。ただし何かの論理的な根拠に基づくわけではなく、主に「極端に非対称に広がった分布でも再現することができて、結果的に合うようにみえる場合が多い」という性格によると思われる。

12-4-3 ロジン・ラムラー分布 Rosin-Rammler distribution (ワイブル分布 Weibull distribution)

1933年にロジン Rosin とラムラー Rammler が篩ふるいにかけた粉体の大きさの分布をこの分布でモデル化した内容の論文 (Rosin & Rammler, [1933](#)) を発表しており、**ロジン・ラムラー分布 Rosin-Rammler distribution** と呼ばれますが、1951年にスウェーデンの機械工学者ワイブル Weibull がこの分布についての論文 (Weibull, [1951](#)) を発表したことが注目され、**ワイブル分布 Weibull distribution** と呼ばれることが多くなりました。最弱リンクモデル ウィークエストリンク (鎖くさりのうち一番弱い輪が壊れると鎖が切れること) に基づくとも言われ、機械部品の劣化・故障などの現象を統計的に扱えるとされる場合もあります。

ただし、ロジン・ラムラーとワイブルのいずれも、合理的な根拠に基づいているわけではなく、「累積分布関数が簡単な数式で表現できるので便利」「なんとなく合っているように見える場合がある」という程度の意味しかないと考えても良いでしょう。

ロジン・ラムラー分布の**確率密度関数**は、

$$f(x; \mu, \beta) = \frac{\beta}{\mu} \left(\frac{x}{\mu}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\mu}\right)^\beta\right] \quad (12.4.3.1)$$

と表され、**累積分布関数**は

$$F(x; \mu, \beta) = 1 - \exp\left[-\left(\frac{x}{\mu}\right)^\beta\right] \quad (12.4.3.2)$$

と表されます。平均 $\langle x \rangle$ は

$$\langle x \rangle = \mu \Gamma\left(1 + \frac{1}{\beta}\right) \quad (12.4.3.3)$$

平方平均 $\langle x^2 \rangle$ は

$$\langle x^2 \rangle = \mu^2 \Gamma\left(1 + \frac{2}{\beta}\right) \quad (12.4.3.4)$$

なので、分散は

$$\sigma^2 = \mu^2 \left[\Gamma\left(1 + \frac{2}{\beta}\right) - \Gamma^2\left(1 + \frac{1}{\beta}\right) \right] \quad (12.3.3.5)$$

となります。

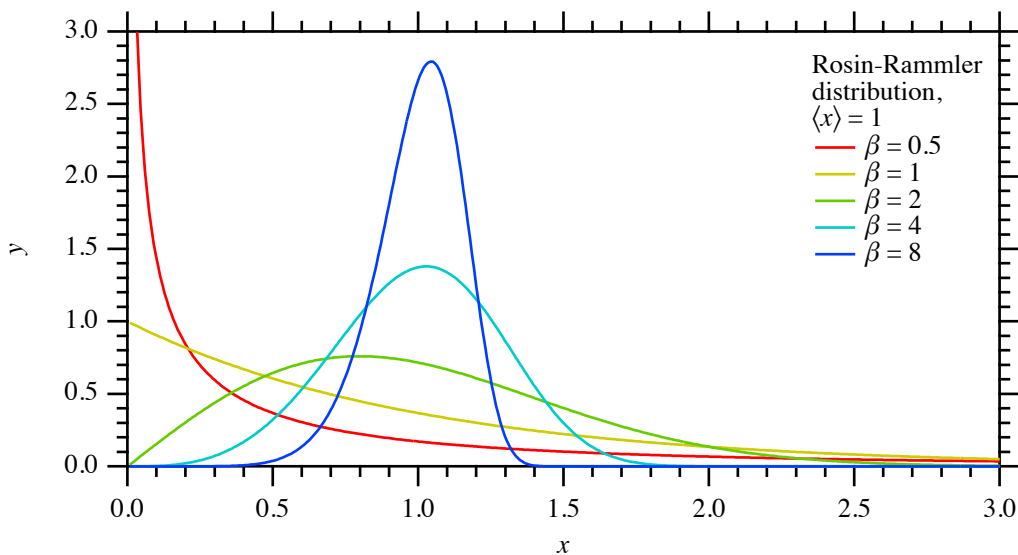


Fig. 12.4.3.1 平均 $\langle x \rangle = 1$ のロジン・ラムラー分布の確率密度関数

12-4-4 スチューデント t 分布 Student's t-distribution

スチューデント t 分布 Student's t-distribution の考え方では、**自由度** degree of freedom ν のとき、**検定統計量** test statistic t の統計分布についての確率密度関数が、

$$f_{\text{Student}}(t; \nu) = \frac{\Gamma(\nu/2 + 1/2)}{\sqrt{\pi\nu} \Gamma(\nu/2)} \left(1 + \frac{t^2}{\nu}\right)^{-\nu/2 - 1/2} \quad (12.4.4.1)$$

$$\Gamma(x) \equiv \int_0^{\infty} t^{x-1} e^{-t} dt \quad (12.4.4.2)$$

と表されるとします。このステューデント t 分布の確率密度関数はピアソンの ^{なな} VII 型関数 Pearson's type VII function としても知られます。

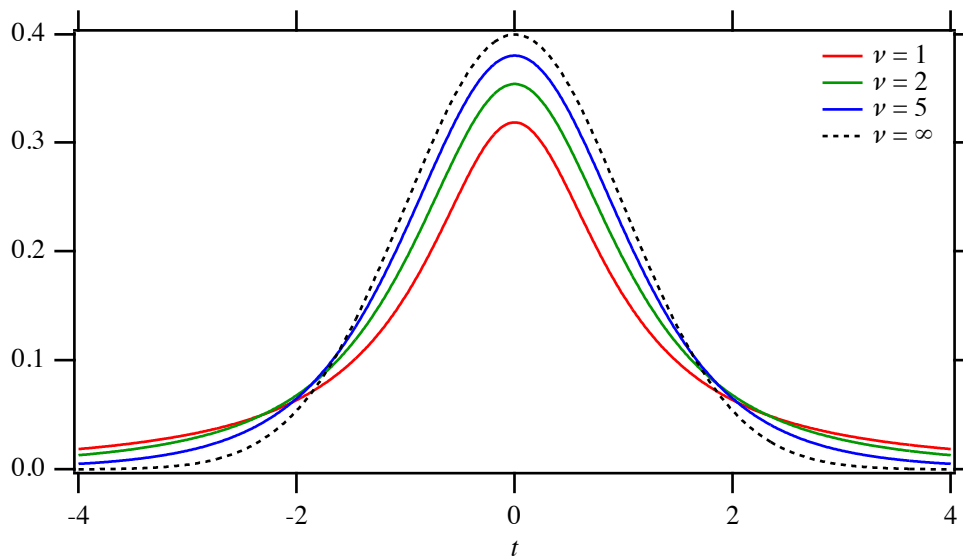


Fig.12.4.4.1 Student t 分布の確率密度関数。自由度を ν とする。

例えば、平均 μ 、標準偏差 σ の正規分布に従う確率変数 X_j について、 n 個の標本値 $\{X_1, \dots, X_n\}$ から、標本平均 \bar{X} を

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j \quad (12.4.4.3)$$

として求め、標本標準偏差 $\hat{\sigma}$ を

$$\hat{\sigma} = \sqrt{\frac{1}{n-1} \sum_j (X_j - \bar{X})^2} \quad (12.4.4.4)$$

として求めたとき、検定統計量 t :

$$t = \frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} \quad (12.4.4.5)$$

の分布は自由度 $\nu = n - 1$ のステューデント t 分布 Student's t-distribution に従い、その確率密度関数が式 (12.4.4.1) で表されます。

ステューデント t 分布の累積分布関数 cumulative distribution function $F_{\text{Student}}(t; \nu)$ は、例えば以下のように表されます。

$$F_{\text{Student}}(t; \nu) = \int_{-\infty}^t f_{\text{Student}}(u; \nu) du = 1 - \frac{1}{2} I\left(\frac{\nu}{t^2 + \nu}; \frac{\nu}{2}, \frac{1}{2}\right) \quad (12.4.4.6)$$

ただし $I(x; a, b)$ は正則不完全ベータ関数 regularized incomplete beta function と呼ばれ、

$$I(x; a, b) \equiv \frac{B(x; a, b)}{B(a, b)} \quad (12.4.4.7)$$

$$B(x; a, b) \equiv \int_0^x t^{a-1}(1-t)^{b-1} dt \quad (12.4.4.8)$$

$$B(a, b) \equiv \int_0^{\infty} t^{a-1}(1-t)^{b-1} dt \quad (12.4.4.9)$$

のように定義されます。函数 $B(x; a, b)$ は**不完全ベータ函数 incomplete beta function**，函数 $B(a, b)$ は，**完全ベータ函数 incomplete beta function** または単にベータ函数と呼ばれます。

自由度 $\nu = 1$ のときステューデント t 分布は**コーシー分布 Cauchy's distribution** と一致し，自由度 $\nu \rightarrow \infty$ の極限では，ステューデント t 分布は**正規分布 normal distribution** と一致します。

ステューデント t 分布は自由度 ν によって何階のキュムラントまで定義できるかが変化します。 $\nu \leq 1$ のときは1階キュムラント（平均）は不定であり， $1 < \nu$ のとき1階キュムラントは $\langle x \rangle = 0$ となります。 $\nu \leq 1$ のとき2階キュムラント（分散）は不定， $1 < \nu \leq 2$ のとき2階キュムラントは ∞ ， $2 < \nu$ のとき2階キュムラントは $\frac{\nu}{\nu-2}$ となります。

12-4-5 ロジスティック分布 logistic distribution

ロジスティック分布 logistic distribution の確率密度函数は，位置パラメータを μ ，尺度パラメータを s として

$$f_{\text{logistic}}(x; \mu, s) = \frac{1}{4s} \operatorname{sech}^2 \frac{x - \mu}{2s} \quad (12.4.5.1)$$

と表されます。ここで $\operatorname{sech} x = \frac{2}{e^x + e^{-x}}$ は，**双曲正割函数 hyperbolic secant function** と呼ばれる**初等函数 elementary function** です。ロジスティック分布の累積分布函数 cumulative distribution function $F_{\text{logistic}}(x; \mu, s)$ は，例えば以下のように表されます。

$$F_{\text{logistic}}(x; \mu, s) = \frac{1}{2} + \frac{1}{2} \tanh \frac{x - \mu}{s} \quad (12.4.5.2)$$

ただし $\tanh x = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ は，**双曲正接函数 hyperbolic tangent function** と呼ばれる**初等函数** です。

式(12.4.5.1)，式(12.4.5.2)のように定義されるロジスティック分布の1階キュムラント（平均）は $\langle x \rangle = \mu$ ，2階キュムラント（分散）は $\sigma^2 = \frac{\pi^2 s^2}{3}$ ，3階キュムラントと歪度は0，尖度は $\frac{6}{5}$ となることが知られています。

ロジスティック分布の累積分布関数は、人工神経網 artificial neural network の中で活性化関数として用いられる場合があります。

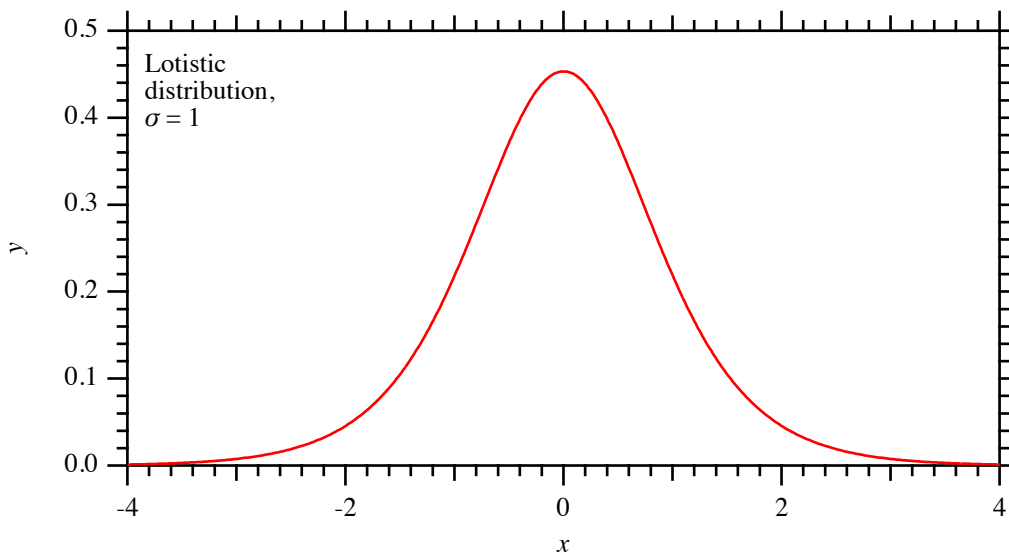


Fig.12.4.5.1 標準偏差 $\sigma = 1$ のロジスティック分布の確率密度関数

(補足 12.1.A) 連続一様分布の尖度 (↔)

連続一様分布の一般形から粗モーメント (冪平均) 計算を経て4階キムラントを求める手順はやや手間がかかるので、ここでは単純化した形式から尖度 kurtosis を求める手順を用います。尖度は平均や分散によらないので、平均がゼロであるような連続一様分布関数 $f(x; -1/2, 1/2)$ の尖度 $\gamma_2 = \kappa_4 / \kappa_2^2$ を求めることにしても一般性は失われません。このときの分散 σ'^2 は平方平均 $\langle x^2 \rangle$ と等しくなり、 $\sigma'^2 = \langle x^2 \rangle$ は式 (12.1.4) から

$$\sigma'^2 = \frac{1}{12} \quad (12.1.A.1)$$

となります。平均がゼロで左右対称な分布の4階キムラントは、一般的に

$$\kappa_4 = \langle x^4 \rangle - 3\langle x^2 \rangle^2 \quad (12.1.A.2)$$

と表されることと、 $f(x; -1/2, 1/2)$ の4階冪平均が

$$\begin{aligned} \langle x^4 \rangle &= \int_{-\infty}^{\infty} x^4 f(x; -1/2, 1/2) dx = \int_{-1/2}^{1/2} x^4 dx = 2 \int_0^{1/2} x^4 dx = 2 \left[\frac{x^5}{5} \right]_0^{1/2} \\ &= \frac{1}{80} \end{aligned} \quad (12.1.A.3)$$

と表されることから、 $f(x; -1/2, 1/2)$ の4階キムラントの値 κ_4' は

$$\kappa_4' = \frac{1}{80} - 3 \times \left(\frac{1}{12} \right)^2 = \frac{3-5}{240} = -\frac{1}{120} \quad (12.1.A.4)$$

となり、連続一様分布の尖度 γ_2 は一般的に

$$\gamma_2 \equiv \frac{\kappa_4'}{\sigma'^4} = -\frac{1}{120} \times 12^2 = -\frac{6}{5} \quad (12.1.A.5)$$

となることがわかります。

連続一様分布 $f(x; a, b)$ の4階キュムラントは,

$$\kappa_4 = \gamma_2 \sigma^4 = -\frac{6}{5} \times \left[\frac{(b-a)^2}{12} \right]^2 = -\frac{(b-a)^4}{120} \quad (12.1.A.6)$$

となります。 (↔)

(補足 12.2.A) 中心極限定理 (↔)

中心極限定理は「独立でランダムな確率変数を足し合わせる操作をしていくと、個々の変数の確率分布が正規分布でなくても、その和として表される変数の確率分布は(多くの場合に)正規分布に近づく」という意味のことですが、このことは双六などのサイコロを使ったゲーム (Fig. 12.2.A.1) をした経験があれば受け入れやすいでしょう。



このプリントはウェブサイトでも無料ダウンロードできます。(c) 幼児の学習素材館 <http://happyillac.net/kisetsu-sozai.html>

Fig. 12.2.A.1 発音練習すごろく【さ行1】 (さ・せ・そ)

サイコロを一つ振って「進む^{マス}目数」を決めるタイプの双六 (「 n 回休み」や「 n 個戻る」「 n 個進む」などのマス目の指示は無視するとします) では、第1ターンで進める数は、1から6のいずれも等しい確率で現れます (Fig. 12.2.A.2)。2ターンで進める数は2から12までですが、可能な場合がすべて同じ確率で現れるのではなく、2マスしか進めない場合も12マス進める場合も確率が低く、7マス進む確率が一番高いこと

に気づくでしょう (Fig. 12.2.A.3)。3 ターンでは 10 マス進む確率と 11 マス進む確率が最も高くなり、最小の 3 マスしか進めない確率と最大の 18 マス進める確率はさらに低くなります (Fig. 12.2.A.4)。

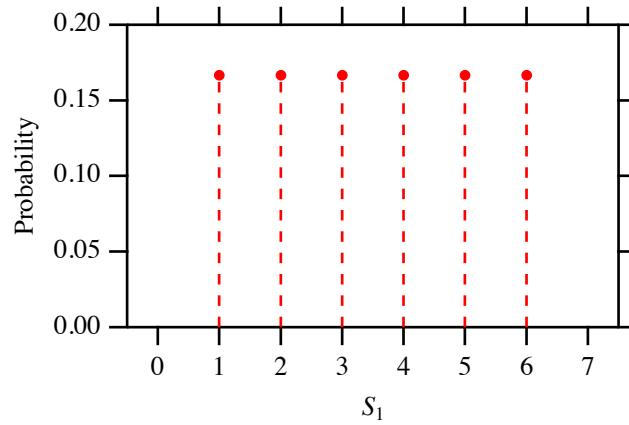


Fig. 12.2.A.2 双六の 1 ターンで進むマス目 (S_1) の出現確率

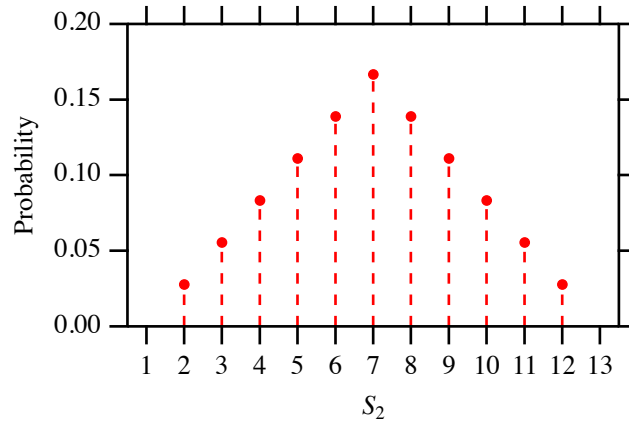


Fig. 12.2.A.3 双六の 2 ターンで進むマス目 (S_2) の出現確率

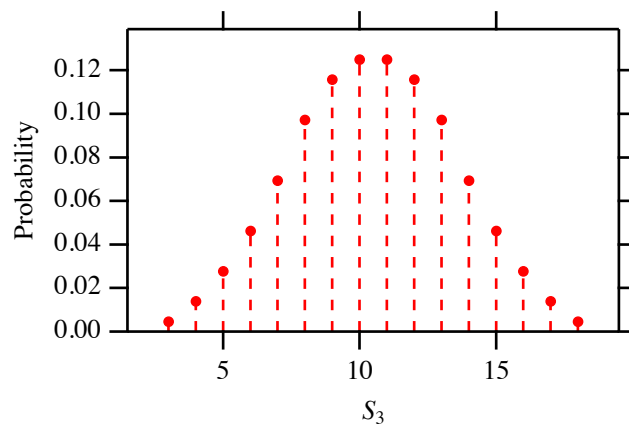


Fig. 12.2.A.4 双六の 3 ターンで進むマス目 (S_3) の出現確率

双六のターン数が増えると、進めるマス目の出現確率は「平均値と分散の等しい正規分布の確率密度関数」に近づきます (Fig. 12.2.A.5)。

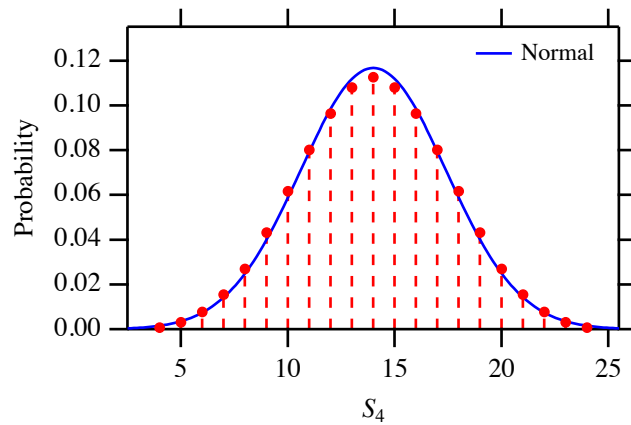


Fig. 12.2.A.5 双六の4ターンで進むマス目 (S_4) の出現確率と正規分布曲線 (青線)

双六の複数ターンで進む^{すごろく} 柁目^{マス}のように「『独立に出現する確率変数の値^{あたひ} (サイコロの目など)』の和で表される値の出現確率」のことを^{たたみこみ} 畳込 convolution^{コンボリューション} と言います。中心極限定理は「畳込を繰り返すと正規分布に近づく」と言い換えることもできます。

任意の階数のキュムラントが定義できる確率分布であれば、双六の例のように、^り 離散一様分布^{さんいちよう} であっても畳込を繰り返せば正規分布に近付きます。逆にコーシー分布 (12-4-1 節) のように、ある階数以上からキュムラントを定義できなくなるような確率分布では、「正規分布に近づく」ような中心極限定理は成立しません。

このことは「畳込」と「キュムラント」の関係から説明することができます (補足 12.2.A.1) (補足 12.2.A.2) (↔)。

(補足 12.2.A.1) 離散分布の畳込とキュムラントの関係, 中心極限定理 (↔)

離散的な確率変数 k, l の出現確率がそれぞれ $P(k), Q(l)$ と表されるとして、 $m = k + l$ で表される値の出現確率 $R(m)$ は、「足して m になる k, l の組み合わせ」のそれぞれの出現確率の積を合わせたものなので、

$$R(m) = \sum_{l=-\infty}^{\infty} P(m-l)Q(l) \quad (12.2.A.1.1)$$

と表されます。この関係は^{クロネッカー} クロネッカーのデルタ **Kronecker delta** と呼ばれる記号「 $\delta_{i,j}$ 」

$$\delta_{i,j} \equiv \begin{cases} 1 & [i = j] \\ 0 & [i \neq j] \end{cases} \quad (12.2.A.1.2)$$

を使って

$$R(m) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \delta_{m,k+l} P(k)Q(l) \quad (12.2.A.1.3)$$

と表すこともできます。

以下、畳込のキュムラントが、成分となる確率分布のキュムラントの和に等しくなることを示します。

離散分布の確率 (質量) 関数 $P(k), Q(l)$ のキュムラント^{ぼかんすう} 母関数が

$$K_P(\theta) = \ln \sum_{k=-\infty}^{\infty} e^{\theta k} P(k) \quad (12.2.A.1.4)$$

$$K_Q(\theta) = \ln \sum_{l=-\infty}^{\infty} e^{\theta l} Q(l) \quad (12.2.A.1.5)$$

と表されるとすると、^{たたみこみ}畳込の^{ぼかんすう}キュムラント母関数 $K_R(\theta)$ は

$$\begin{aligned} K_R(\theta) &= \ln \sum_{m=-\infty}^{\infty} e^{\theta m} R(m) && \text{(式 (12.2.A.1.3) から→)} \\ &= \ln \sum_{m=-\infty}^{\infty} e^{\theta m} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \delta_{m,k+l} P(k) Q(l) && \text{(和記号の順序を入れ替えて→)} \\ &= \ln \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} e^{\theta m} \delta_{m,k+l} P(k) Q(l) && \left(\sum_{m=-\infty}^{\infty} e^{\theta m} \delta_{m,k+l} = e^{\theta(k+l)} \text{ だから→} \right) \\ &= \ln \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} e^{\theta(k+l)} P(k) Q(l) && \left(e^{\theta(k+l)} = e^{\theta k} e^{\theta l}, \text{ 和記号の順序を入れ替えて→} \right) \\ &= \ln \sum_{k=-\infty}^{\infty} e^{\theta k} P(k) \sum_{l=-\infty}^{\infty} e^{\theta l} Q(l) && \text{(} \ln(xy) = \ln x + \ln y \text{ の関係から→)} \\ &= \ln \sum_{k=-\infty}^{\infty} e^{\theta k} P(k) + \ln \sum_{l=-\infty}^{\infty} e^{\theta l} Q(l) \\ &= K_P(\theta) + K_Q(\theta) && (12.2.A.1.6) \end{aligned}$$

となって、成分関数のキュムラント母関数の和に等しくなり、^{たたみこみ}畳込のキュムラントも成分関数のキュムラントの和に等しくなります。この関係を「畳込におけるキュムラントの可加算性」(additivity of cumulants on convolution) と呼びます。

確率質量関数が $P(k)$ 、その ν 次キュムラントが κ_ν と表される場合に、平均 μ と分散 σ^2 の等しい正規分布からどのように食い違うかは、3階以上のキュムラントで決まり、その特徴は $\kappa_\nu / \sigma^\nu = \kappa_\nu / \kappa_2^{\nu/2}$ で表されます。

同じ分布を n 回重ねた畳込のキュムラントは $n\kappa_\nu$ と表されます。3階以上 ($\nu \geq 3$) のキュムラントの特徴は

$$\frac{n\kappa_\nu}{(n\kappa_2)^{\nu/2}} = \frac{1}{n^{\nu/2-1}} \frac{\kappa_\nu}{\kappa_2^{\nu/2}} \quad (12.2.A.1.7)$$

の関係から、 n が大きくなるほど相対的に小さくなり、「3階以上のキュムラントがすべてゼロになる」という正規分布の特徴 ([補足 12.2.D](#)) に近づきます。このことは元の分布が離散分布であっても例外ではありません。 ([↔](#))

(補足 12.2.A.2) 連続分布の畳込とキュムラントの関係、中心極限定理 ([↔](#))

独立でランダムに出現する2つの連続的な確率変数 x, y の確率密度関数がそれぞれ $f(x), g(y)$ と表されるとして、 $z = x + y$ で表される値の出現する統計分布の確率密度関数 $h(z)$ は、「足して z になる x, y の組み合わせ」のそれぞれの確率密度の積を積分したものになり、

$$h(z) = \int_{-\infty}^{\infty} f(z-y)g(y) dy \quad (12.2.A.2.1)$$

と表されます。^{ディラック デルタ ファンクション}ディラックのデルタ関数 Dirac delta function $\delta(x)$ を用いれば、式 ([12.2.A.2.1](#)) の関係を

$$h(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(z - x - y) f(x) g(y) dx dy \quad (12.2.A.2.2)$$

と表すこともできます。

畳込のキュムラントが、成分となる確率分布のキュムラントの和に等しくなることは、離散分布の場合（補足 12.2.A.1）と同様に、連続分布の場合にも成立します。

連続分布の確率密度関数 $f(x)$, $g(x)$ のキュムラント母函数が

$$K_f(\theta) = \ln \int_{-\infty}^{\infty} e^{\theta x} f(x) dx \quad (12.2.A.2.3)$$

$$K_g(\theta) = \ln \int_{-\infty}^{\infty} e^{\theta y} g(y) dy \quad (12.2.A.2.4)$$

と表されるとすると、畳込のキュムラント母函数 $K_h(\theta)$ は

$$\begin{aligned} K_h(\theta) &= \ln \int_{-\infty}^{\infty} e^{\theta z} h(z) dz \\ &= \ln \int_{-\infty}^{\infty} e^{\theta z} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(z - x - y) f(x) f(y) dx dy dz \\ &= \ln \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\theta z} \delta(z - x - y) f(x) f(y) dz dx dy \\ &= \ln \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\theta(x+y)} f(x) f(y) dx dy \\ &= \ln \left[\int_{-\infty}^{\infty} e^{\theta x} f(x) dx \int_{-\infty}^{\infty} e^{\theta y} f(y) dy \right] \\ &= \ln \int_{-\infty}^{\infty} e^{\theta x} f(x) dx + \ln \int_{-\infty}^{\infty} e^{\theta y} f(y) dy \\ &= K_f(\theta) + K_g(\theta) \end{aligned} \quad (12.2.A.2.5)$$

となって、成分函数のキュムラント母函数の和に等しくなります。したがって、畳込のキュムラントは成分函数のキュムラントの和に等しくなります。

連続確率分布の確率密度関数が $f(x)$ で表され、その ν (ギリシャ文字のニュー) 次のキュムラントが κ_ν (カッパ・ニュー) と表される場合に、平均 μ と分散 σ^2 の等しい正規分布からどのように食い違うかは、3階以上のキュムラントで決まり、歪度 skewness や尖度 kurtosis に見られるように、その特徴が $\kappa_\nu / \sigma^\nu = \kappa_\nu / \kappa_2^{\nu/2}$ で表されます。

同じ分布を n 回重ねた畳込の ν 階のキュムラントは $n\kappa_\nu$ と表されます。3階以上 ($\nu \geq 3$) のキュムラントの特徴は

$$\frac{n\kappa_\nu}{(n\kappa_2)^{\nu/2}} = \frac{1}{n^{\nu/2-1}} \frac{\kappa_\nu}{\kappa_2^{\nu/2}} \quad (12.2.A.2.6)$$

の関係から、 n が大きくなるほど相対的に小さくなり、「3階以上の階数のキュムラントがすべてゼロになる」という正規分布の特徴（補足 12.2.D）に近づきます。（↔）

（補足 12.2.B）正規分布のキュムラント（↔）

変数を t とする正規分布のモーメント母函数は

$$\begin{aligned}
M(t) &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
&= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} + tx\right] dx \\
&= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \exp\left[-\frac{(x-\mu)^2 - 2\sigma^2 tx}{2\sigma^2}\right] dx \\
&= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \exp\left[-\frac{(x-\mu-\sigma^2 t)^2 - (\mu+\sigma^2 t)^2 + \mu^2}{2\sigma^2}\right] dx \\
&= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \exp\left[-\frac{(x-\mu-\sigma^2 t)^2 - 2\mu\sigma^2 t - \sigma^4 t^2}{2\sigma^2}\right] dx \\
&= \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \exp\left[-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2} + \mu t + \frac{\sigma^2 t^2}{2}\right] dx \\
&= \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} \exp\left[-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}\right] dx \\
&= \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)
\end{aligned} \tag{12.2.B.1}$$

となり、正規分布のキュムラント母関数は

$$K(t) = \ln M(t) = \mu t + \frac{\sigma^2 t^2}{2} \tag{12.2.B.2}$$

と書けます。正規分布の1階キュムラント（平均）と2階キュムラント（分散），3階キュムラントは

$$\frac{\partial K(t)}{\partial t} = \mu + \sigma^2 t \xrightarrow{t \rightarrow 0} \mu \tag{12.2.B.3}$$

$$\frac{\partial^2 K(t)}{\partial t^2} = \sigma^2 \tag{12.2.B.4}$$

$$\frac{\partial^3 K(t)}{\partial t^3} = 0 \tag{12.2.B.5}$$

となり、3階以上のキュムラントは^{すべて}全てゼロになります。「3階以上のキュムラントが全てゼロになること」が正規分布の特徴であるとも言えます。 (↔)

(補足 12.2.C) 正規分布と偏差値・知能指数 (↔)

日本式の偏差値は、平均を50点、標準偏差を10点として得点を尺度化するものなので、得点の分布が正規分布に従うなら、式(12.2.13)–(12.2.16)から「偏差値40以上60以下は68.3%」「偏差値30以上70以下は95.5%」「偏差値20以上80以下は99.7%」などと、偏差値からおよその位置付けをすることが可能です。

知能指数 (IQ; intelligence quotient) は元々は「精神年齢/身体年齢×100」という意味でしたが、現在主に用いられるウェクスラー式知能検査では、知能テストの得点について、平均を100点、標準偏差を15点として得点を尺度化するものとされています。知能テストの得点が正規分布に従う場合には、式(12.2.13)–(12.2.16)から「IQ85以上115以下は68.3%」「IQ70以上130以下は95.5%」「IQ55以上145以下は99.7%」のようにIQの値からおよその位置付けをすることが可能です。 (↔)

(補足 12.3.A) ガンマ分布のキュムラント (↔)

変数を t とするガンマ分布のモーメント母関数は

$$\begin{aligned} M(t) &= \int_0^{\infty} e^{tx} \frac{1}{\Gamma(\alpha)\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} \exp\left(-\frac{x}{\theta}\right) dx \\ &= \int_0^{\infty} \frac{1}{\Gamma(\alpha)\theta} \left(\frac{x}{\theta}\right)^{\alpha-1} \exp\left[-\left(\frac{1}{\theta}-t\right)x\right] dx \\ &= \int_0^{\infty} \frac{1/\theta-t}{\Gamma(\alpha)} \left(\frac{1/\theta}{1/\theta-t}\right)^{\alpha} \left[\left(\frac{1}{\theta}-t\right)x\right]^{\alpha-1} \exp\left[-\left(\frac{1}{\theta}-t\right)x\right] dx \\ &= \left(\frac{1/\theta}{1/\theta-t}\right)^{\alpha} \int_0^{\infty} \frac{1/\theta-t}{\Gamma(\alpha)} \left[\left(\frac{1}{\theta}-t\right)x\right]^{\alpha-1} \exp\left[-\left(\frac{1}{\theta}-t\right)x\right] dx \\ &= \left(\frac{1/\theta}{1/\theta-t}\right)^{\alpha} = \left(\frac{1}{1-\theta t}\right)^{\alpha} = \frac{1}{(1-\theta t)^{\alpha}} \end{aligned} \tag{12.3.A.1}$$

と表され、キュムラント母関数は

$$K(t) = \ln M(t) = \ln \frac{1}{(1-\theta t)^{\alpha}} = -\alpha \ln(1-\theta t) \tag{12.3.A.2}$$

となるので、ガンマ分布の 1 階から 4 階までのキュムラントの値は

$$\frac{\partial K(t)}{\partial t} = \frac{\alpha\theta}{1-\theta t} = \alpha\theta \left(1 + \theta t + \theta^2 t^2 + \theta^3 t^3 + \theta^4 t^4 + \dots\right) \xrightarrow{t \rightarrow 0} \alpha\theta \tag{12.3.A.3}$$

$$\frac{\partial^2 K(t)}{\partial t^2} = \alpha\theta \left(\theta + 2\theta^2 t + 3\theta^3 t^2 + 4\theta^4 t^3 + \dots\right) \xrightarrow{t \rightarrow 0} \alpha\theta^2 \tag{12.3.A.4}$$

$$\frac{\partial^3 K(t)}{\partial t^3} = \alpha\theta \left(2\theta^2 + 3 \times 2\theta^3 t + 4 \times 3\theta^4 t^2 + \dots\right) \xrightarrow{t \rightarrow 0} 2\alpha\theta^3 \tag{12.3.A.5}$$

$$\frac{\partial^4 K(t)}{\partial t^4} = \alpha\theta \left(3 \times 2\theta^3 + 4 \times 3 \times 2\theta^4 t + \dots\right) \xrightarrow{t \rightarrow 0} 6\alpha\theta^4 \tag{12.3.A.6}$$

などとなり、さらに一般的に k 階キュムラントは

$$\frac{\partial^k K(t)}{\partial t^k} \xrightarrow{t \rightarrow 0} (k-1)! \alpha\theta^k \tag{12.3.A.7}$$

と表されることがわかります。 (↔)

(補足 12.3.B) 標準正規分布とカイ自乗分布の関係 (↔)

スタンダード ノーマル ディストリビューション
標準正規分布 standard normal distribution の確率密度関数

$$f_{\text{SND}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \tag{12.3.B.1}$$

に対して、 $y = x^2$ とすれば、 $dy = 2x dx \Rightarrow dx = \pm y^{-1/2} dy/2$ となることから、 $y = x^2$ の分布は

$$f(y) = \frac{1}{\sqrt{2\pi}} y^{-1/2} \exp\left(-\frac{y}{2}\right) \tag{12.3.B.2}$$

となります。ただし、 $0 < y$ とします。式 (12.3.B.2) によって確率密度関数の表される統計分布が「自由度 1 のカイ自乗分布」と呼ばれます。

標準正規分布の確率密度関数 $f_{\text{SND}}(x)$ を表す式 (12.3.B.1) 中の x^2 を y に置き換えただけでは、カイ自乗分布正しい確率密度関数にはならないことに注意してください。この例では、確率密度関数 $f(y)$ の変数が y から $y + dy$ の値をとる確率が $f(y) dy$ で表され、その時に $f_{\text{SND}}(x)$ の変数 $x = \pm\sqrt{y}$ の値は、「 $-\sqrt{y + dy}$ から $-\sqrt{y}$ 」と、「 \sqrt{y} から $\sqrt{y + dy}$ 」の2通りの範囲の値を取ります。また

$$-\sqrt{y + dy} = -\sqrt{y}(1 + dy/y)^{1/2} \simeq -\sqrt{y}(1 + dy/2y) = -\sqrt{y} - dy/2\sqrt{y} \quad (12.3.B.3)$$

$$\sqrt{y + dy} \simeq \sqrt{y} + dy/2\sqrt{y} \quad (12.3.B.4)$$

の関係から、それぞれの範囲でとりうる x の範囲は $dy/2\sqrt{y}$ の幅を持ち、

$$f(y)dy = f_{\text{SND}}(\sqrt{y}) \times 2 \times \frac{dy}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} y^{-1/2} \exp\left(-\frac{y}{2}\right) dy \quad (12.3.B.5)$$

となることから式 (12.3.B.2) の関係が導かれます。

変数 x の分布に関する確率密度関数が $f(x)$ で表され、微分可能な単調増加関数 $h(x)$ を使い $y = h(x) \Leftrightarrow x = h^{-1}(y)$ として変数変換を施す時には、変数 y の分布の確率密度関数 $g(y)$ は

$$g(y) = \begin{cases} 0 & [y \leq h(-\infty)] \\ f(h^{-1}(y)) \frac{dx}{dy} & [h(-\infty) < y < h(\infty)] \\ 0 & [h(\infty) \leq y] \end{cases} \quad (12.3.B.6)$$

と表されます。

確率密度関数 $f(x)$ の累積分布関数 $F(x)$ を

$$F(x) \equiv \int_{-\infty}^x f(t) dt \quad (12.3.B.7)$$

とすれば、変数を $y = h(x)$ で変換した後の累積分布関数 $G(y)$ については

$$G(y) = \begin{cases} 0 & [y \leq h(-\infty)] \\ F(h^{-1}(y)) & [h(-\infty) < y < h(\infty)] \\ 1 & [h(\infty) \leq y] \end{cases} \quad (12.3.B.8)$$

の関係が成り立ちます。

累積分布関数に関しては、確率密度関数と異なり微分をかける必要がなく、確率変数を置き換えるだけで良いので、確率変数の変換の際に混乱しそうな時は、累積分布関数から考え直す方がわかりやすいかもしれません。 (↔)

(補足 12.4.2.A) 対数正規分布と正規分布の関係 (↔)

平均 m 、分散 ω^2 の正規分布の確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi\omega}} \exp\left[-\frac{(x - m)^2}{2\omega^2}\right] \quad (12.4.2.A.1)$$

に対して、 $x = \ln y \Leftrightarrow y = e^x$ の変数変換を当てはめると、 y の確率密度関数 $g(y)$ は (補足 12.3.B) の関係と $dx/dy = d(\ln y)/dy = 1/y$ の関係から

$$g(y) = \frac{1}{\sqrt{2\pi\omega y}} \exp \left[-\frac{(\ln y - m)^2}{2\omega^2} \right] \quad (12.4.2.A.2)$$

となり、式 (12.4.2.1) の対数正規分布の確率密度関数が導かれます。 (↔)

(補足 12.4.2.B) 確率変数の次元による対数正規分布の変化

対数正規分布は粉末中の粒の大きさの分布をモデル化する場合に使われる場合も多いのですが、例えば「粒子の直径」を変数として対数正規分布でモデル化した時に、その平均や分散にはあまり意味がありません。むしろ粒子の面積や体積の平均や分散に意味があります。

中央値 x_m 、対数標準偏差 ロガリズムック スタンダード デバイエーション logarithmic standard deviation ω の対数正規分布の確率密度関数は

$$f(x; x_m, \omega) = \frac{1}{\sqrt{2\pi\omega x}} \exp \left[-\frac{(\ln x - \ln x_m)^2}{2\omega^2} \right] \quad (12.4.B.2)$$

と書くことができますが、この時

$$\begin{aligned} x^k f(x; x_m, \omega) &= \frac{1}{\sqrt{2\pi\omega x}} \exp \left[-\frac{(\ln x - \ln x_m)^2}{2\omega^2} + k \ln x \right] \\ &= \frac{1}{\sqrt{2\pi\omega x}} \exp \left[-\frac{(\ln x - \ln x_m)^2 - 2k\omega^2 \ln x}{2\omega^2} \right] \\ &= \frac{1}{\sqrt{2\pi\omega x}} \exp \left[-\frac{(\ln x - \ln x_m - k\omega^2)^2 - (\ln x + k\omega^2) + (\ln x_m)^2}{2\omega^2} \right] \\ &= \frac{1}{\sqrt{2\pi\omega x}} \exp \left[-\frac{(\ln x - \ln x_m - k\omega^2)^2 - 2k\omega^2 \ln x_m - k^2\omega^4}{2\omega^2} \right] \\ &= \frac{1}{\sqrt{2\pi\omega x}} \exp \left[-\frac{(\ln x - \ln x_m - k\omega^2)^2}{2\omega^2} + k \ln x_m + \frac{k^2\omega^2}{2} \right] \\ &= x_m^k \exp \left(\frac{k^2\omega^2}{2} \right) \frac{1}{\sqrt{2\pi\omega x}} \exp \left[-\frac{(\ln x - \ln x_m - k\omega^2)^2}{2\omega^2} \right] \end{aligned} \quad (12.4.B.3)$$

となることと、

$$\int_0^\infty \frac{1}{\sqrt{2\pi\omega x}} \exp \left[-\frac{(\ln x - \ln x_m - k\omega^2)^2}{2\omega^2} \right] dx = 1$$

の関係が成立することから、

$$\langle x^k \rangle = \int_0^\infty x^k f(x; x, \omega) dx = x_m^k \exp \left(\frac{k^2\omega^2}{2} \right)$$

となります。

また $y = x^k$ の分布の確率密度関数 $g(y; x_m, \omega)$ は、 $x = y^{1/k}$ 、 $dx/dy = y^{1/k-1}/k$ の関係、(補足 12.3.B) の関係から、

$$g(y; x_m, \omega) = \frac{1}{\sqrt{2\pi\omega y^{1/k}}} \frac{y^{1/k-1}}{k} \exp \left[-\frac{(\ln y - k \ln x_m)^2}{2k^2\omega^2} \right]$$

$$= \frac{1}{\sqrt{2\pi}k\omega y} \exp \left[-\frac{(\ln y - \ln x_m^k)^2}{2k^2\omega^2} \right] \quad (12.4.B.4)$$

となり、「中央値 $y_m \equiv x_m^k$, 対数標準偏差 $k\omega$ の対数正規分布」になることもわかります。

参考文献 12

Rosin, P. & Rammler, E. (1933), “The Laws Governing the Fineness of Powdered Coal”, *Journal of the Institute of Fuel*, **7**, 29-36. ([↔](#))

Weibull, W. (1951), “A statistical distribution function of wide applicability”, *J. Appl. Mech.-Trans. ASME*, **18**(3), 293-297. ([↔](#))