

## 統計と検定・推定の話

2014.4.17

## 内容

☆確率・統計のこと.....	1
★ランダム変数と確率分布.....	1
◇変数変換と分布関数.....	2
★平均と分散.....	3
★正規分布と中心極限定理.....	4
★多数回の測定の平均と分散.....	4
☆検定・推定のこと.....	5
★仮説検定の考え方.....	6
★標準偏差が分かっている場合の検定・推定.....	6
★カイ 2 乗( $\chi^2$ )分布とカイ 2 乗検定.....	7
★ $t$ 検定.....	8
★ $F$ 検定.....	9
★外れ値.....	9
問題.....	10

## ☆確率・統計のこと

確率・統計をめぐるにはさまざまな立場、観点があるが、ここではできるだけ直観的、実用的な面から述べよう。

## ★ランダム変数と確率分布

あるランダム変数  $x$  が  $x$  の近傍  $dx$  の値をとる確率を  $f(x) dx$  で表し、 $f(x)$  を (確率) 分布関数あるいは確率密度と呼ぶ。分布関数  $f(x)$  は 0 か正で  $x$  の全領域について積分したものは 1 になる。

$$f(x) \geq 0 \quad \text{かつ} \quad \int_{-\infty}^{\infty} f(x) dx = 1$$

ここで分布関数  $f(x)$  は密度であり、一般に単位があるということに注意する。ランダム変数が  $x$  以下の値を取る確率

$$F(x) = \int_{-\infty}^x f(x) dx$$

を積算 (確率) 分布関数と呼ぶ\*。  $F(-\infty) = 0$ 、 $F(\infty) = 1$  である。

多変数の場合についても分布関数、積算分布関数は同様に定義され、2 変数  $x$ 、 $y$  に対する分布関数  $f_{x,y}(x, y)$  は次の性質を満たす。

$$f_{x,y}(x, y) \geq 0 \quad \text{かつ} \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{x,y}(x, y) dx dy = 1$$

また一方の変数について全区間積分したものは 1 変数の分布関数になる。

$$\int_{-\infty}^{\infty} f_{x,y}(x, y) dy = f_x(x) \quad \text{あるいは} \quad F_{x,y}(x, \infty) = F_x(x)$$

多変数の分布関数が 1 変数の分布関数の積で表されるとき、それぞれのランダム変数は統計的に独

\* 特に数学の分野では積算 (確率) 分布関数を単に (確率) 分布関数と呼び、(確率) 分布関数を確率密度と呼ぶことが多い。分野により、文脈により、「分布関数」という言葉で、確率密度を指す場合と積算分布を指す場合があるので注意する。

立であると呼ぶ。たとえば2変数  $\mathbf{x}$ ,  $\mathbf{y}$  が統計的に独立なら、次の関係が成立する

$$f_{x,y}(x, y) = f_x(x) f_y(y) \quad \text{あるいは} \quad F_{x,y}(x, y) = F_x(x) F_y(y)$$

ここで「確率」は何らかの事象の起きる頻度に比例するものと考えなくてよいことに注意する。たとえばある砂浜の砂の粒の粒径  $r$  の分布を考えてみよう。砂粒を実際に数えて、粒径  $r$  の砂粒の分布関数  $f_n(r)$  を得るのも一つの考えだが、粒径が  $r_m$  ( $r_1 > r_2 > \dots > r_n$ ) 以上の砂粒を通さないふるいに順次かけて、それぞれの分画の重さを量って、粒径  $r$  の分画の重さの比を分布関数  $f_w(r)$  とみなすこともできる。かりに粒径  $r$  の砂粒の重さ  $w$  について  $w = ar^3$  という関係がなりたっておれば、

$$f_w(r) = \frac{f_n(r)r^3}{\int f_n(r)r^3 dr} \quad \text{あるいは} \quad f_n(r) = \frac{f_w(r)r^{-3}}{\int f_w(r)r^{-3} dr}$$

という関係が成立し、どちらの式も要件を満たしている。出現頻度に注目するなら、出現した事象の重み(測度)によって、分布関数は異なる装いを示すとも言えよう。

#### ◇変数変換と分布関数

ランダム変数  $\mathbf{y}$  がランダム変数  $\mathbf{x}$  の関数として与えられている時、 $\mathbf{x}$  の分布関数  $f_x(x)$  と  $\mathbf{y}$  の分布関数  $f_y(y)$  の関係を考えよう。今簡単のため逆関数が存在するとし、 $\mathbf{y} = h(\mathbf{x})$ 、 $\mathbf{x} = h^{-1}(\mathbf{y})$  であるものとする。この時、 $\mathbf{x}$  が  $x$  の近傍  $dx$  の値をとる確率と  $\mathbf{y}$  が  $h(x)$  の近傍  $dh(x)$  の値をとる確率は等しいので(確率の保存)、次の関係が成立する。

$$f_y(y) = \left| \frac{dx}{dy} \right| f_x(x)$$

ここで  $x = h^{-1}(y)$  である。したがって例えば  $\mathbf{y} = \mathbf{x}/2$  であれば  $f_y(y) = 2 f_x(2y)$  である。

多変数の場合の変数変換もヤコビアンを用いて同様に定式化できるが、ここでは特に2変数の分布関数を1変数に縮約することを考えてみよう。ランダム変数  $\mathbf{x}$  と  $\mathbf{y}$  の分布関数を、 $\mathbf{z} = h(\mathbf{x}, \mathbf{y})$  と  $\mathbf{y}$  の分布関数に変換すると次の関係が成立する ( $\mathbf{x} = h^{-1}(\mathbf{z}, \mathbf{y})$  という逆関数があるとする)。

$$f_{z,y}(z, y) = \left| \frac{dx}{dz} \right| f_{x,y}(x, y)$$

したがって  $\mathbf{z}$  の分布関数は次式で与えられる

$$f_z(z) = \int_{-\infty}^{\infty} \left| \frac{dx}{dz} \right| f_{x,y}(x, y) dy$$

特に  $\mathbf{z} = \mathbf{x} + \mathbf{y}$  という変数の加算について考えてみると  $\mathbf{x} = h^{-1}(\mathbf{z}, \mathbf{y}) = \mathbf{z} - \mathbf{y}$  であるから

$$f_z(z) = \int_{-\infty}^{\infty} f_{x,y}(z - y, y) dy$$

特に  $\mathbf{x}$  と  $\mathbf{y}$  が統計的に独立であるなら次式が成り立つ

$$f_z(z) = \int_{-\infty}^{\infty} f_x(z - y) f_y(y) dy$$

この形の積分を畳み込み convolution と呼び、化学現象を扱う上でしばしば登場する。

\* 砂粒のように不定形の粒子の粒径・粒子サイズをどのように定めるかには Feret (Green) 径、Martin 径、Krummbein 径等、種々の手法があるが、ここでは代表的な長さといった程度の意味で考える。

## ★平均と分散

ランダム変数の分布関数が与えられた時、その特徴をいくつかの数値・特性値に凝縮させて理解することができれば便利である。そうした特性値として平均と分散がある。

ランダム変数  $x$  の分布関数  $f(x)$  で重みを付けて、関数  $q(x)$  を全区間で積分したものを  $\langle q(x) \rangle$  と書き、 $q(x)$  の平均あるいは期待値と呼ぶ\*。

$$\langle q(x) \rangle \equiv \int_{-\infty}^{\infty} q(x)f(x)dx$$

通常  $x$  の平均と呼ぶのは  $\langle x \rangle$  で、簡単のため  $\mu$  (あるいは  $\mu_x$ ) で表わすこともある：

$$\langle x \rangle = \mu_x \equiv \int_{-\infty}^{\infty} xf(x)dx$$

平均は分布の中心位置を示すものと考えてよい。また  $x$  の  $n$  乗の平均  $\langle x^n \rangle$  を  $x$  の  $n$  次のモーメントと呼び  $\mu_n$  と表記する。なお表記の際、ランダム変数であることを明示せず、 $x$  と  $y$  の和の平均を  $\langle x + y \rangle$  と書くなど、かなり便宜的な使い方をすることがあるが、そこは文脈で判断しよう・・・

平均については次のような関係が成立する

$$\langle ax + b \rangle = a\langle x \rangle + b \quad a \text{ と } b \text{ は定数}$$

$$\langle x + y \rangle = \langle x \rangle + \langle y \rangle \quad x \text{ と } y \text{ はランダム変数}$$

特に  $x$  と  $y$  が統計的に独立であるなら、 $x$  と  $y$  の積について次の関係式が成立する

$$\langle xy \rangle = \langle x \rangle \langle y \rangle$$

平均は分布の中心位置に相当するものを与えるのに対し、分散・標準偏差は分布がどの程度の広がりを持っているかを与える。ランダム変数  $x$  の分散はよく  $\sigma^2$  (あるいは  $\sigma_x^2$ ) で表され、ここでは  $\langle\langle x, x \rangle\rangle$  あるいは  $\langle\langle x^2 \rangle\rangle$  という表記も用いる。ここで  $\langle\langle x, y \rangle\rangle$  は次の量に相当する

$$\langle\langle x, y \rangle\rangle = \langle\langle xy \rangle\rangle = \langle xy \rangle - \langle x \rangle \langle y \rangle$$

したがって  $x$  の分散は次式で定義されることになる

$$\langle\langle x, x \rangle\rangle = \langle\langle x^2 \rangle\rangle = \sigma_x^2 \equiv \langle x^2 \rangle - \langle x \rangle^2 = \mu_2 - \mu_1^2$$

標準偏差  $\sigma$  (あるいは  $\sigma_x$ ) は分散の平方根である。

$$\sigma_x \equiv \sqrt{\langle\langle x^2 \rangle\rangle}$$

標準偏差はランダム変数の平均値周りの変動幅の大きさを与えるものと考えられる。先に導入した  $\langle\langle x, y \rangle\rangle$  はランダム変数  $x$  と  $y$  の共分散と呼ばれる。共分散に関わっては次の関係式が成立する ( $x, y, z$  はランダム変数。  $a$  は定数)

$$\langle\langle x, y \rangle\rangle = \langle\langle y, x \rangle\rangle$$

$$\langle\langle x, a \rangle\rangle = 0$$

$$\langle\langle x + y, z \rangle\rangle = \langle\langle x, z \rangle\rangle + \langle\langle y, z \rangle\rangle$$

$$\langle\langle ax, y \rangle\rangle = a\langle\langle x, y \rangle\rangle$$

したがって次式が成立する ( $\mu = \langle x \rangle$ )

$$\langle\langle x, x \rangle\rangle = \langle\langle x - \mu, x - \mu \rangle\rangle = \langle (x - \mu)^2 \rangle \geq 0$$

\* 分布関数のフーリエ変換は  $\langle \exp(i\omega x) \rangle$  と表わされる (特性関数と呼ばれる)。

分散は平均の周りの 2 次のモーメントに相当し、負にはならない。同様に共分散は次式で表現することができる。

$$\langle\langle x, y \rangle\rangle = \langle\langle x - \mu_x, y - \mu_y \rangle\rangle = \langle\langle (x - \mu_x)(y - \mu_y) \rangle\rangle$$

共分散が正であれば、 $y$  は  $x$  の増加にともなって増加し、負であれば減少する傾向にある。また  $x$  と  $y$  が統計的に独立であれば共分散はゼロになる。

$$\langle\langle x, y \rangle\rangle = 0 \quad x \text{ と } y \text{ が独立}$$

共分散がゼロの場合  $x$  と  $y$  は無相関であると呼ぶ。無相関であっても統計的に独立であるとは限らないが、統計的に独立であることの重要な指標となる。また統計的に独立なランダム変数の和  $z = x + y$  の分散は、 $x$  と  $y$  それぞれの分散の和になる

$$\langle\langle z^2 \rangle\rangle = \langle\langle x + y, x + y \rangle\rangle = \langle\langle x, x \rangle\rangle + 2\langle\langle x, y \rangle\rangle + \langle\langle y, y \rangle\rangle = \langle\langle x^2 \rangle\rangle + \langle\langle y^2 \rangle\rangle$$

### ★正規分布と中心極限定理

平均  $\mu$ 、分散  $\sigma^2$  の正規分布は  $N(\mu, \sigma^2)$  と表記され、分布関数は次式で与えられる。

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

$x$  について

$$y = (x - \mu) / \sigma$$

という変換を施すと、 $y$  は平均 0、分散 1 の正規分布  $N(0, 1)$  に従うことになる

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$$

これを標準化正規分布と呼び、標準化正規分布に従うランダム変数を標準化正規変数と呼ぶ。

正規分布の重要な特徴は、正規分布に従う独立なランダム変数の和も正規分布に従うことである。実際 2 つの標準化正規変数  $x$  と  $y$  の和、 $z = x + y$  の分布関数を求めてみると

$$\begin{aligned} f_z(z) &= \int_{-\infty}^{\infty} f(z-y)f(y)dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left[-\frac{z^2}{4} - (y-z/2)^2\right] dy \\ &= \frac{1}{2\sqrt{\pi}} e^{-\frac{z^2}{4}} \end{aligned}$$

となり、平均 0 分散 2 の正規分布になることがわかる。

多数の独立なランダム変数  $x_i$  ( $i = 1, 2, \dots, N$ ) を加えたランダム変数  $y = \sum x_i$  の確率分布はたいいてい平均  $\sum \langle x_i \rangle$ 、分散  $\sum \langle x_i^2 \rangle$  の正規分布に近づくことが知られている (中心極限定理 **central limit theorem**)。測定値のばらつきは多くの独立な不確定要素が絡み合って生じることが多く、分布関数はたいいていの場合、正規分布に従うものと考えてもよい。同じような分布関数に従うランダム変数の場合、少々歪んだ分布を持っていても、典型的には 5 個程度の和を取るとほぼ正規分布と見なせるようになる。このため正規分布、そしてそれを決める要素である平均と分散に大きな注目が集まることにもなる。

### ★多数回の測定の平均と分散

ある集団から、いくつかのサンプル (標本) をランダムに抽出してその特性値 (色、密度、人間

であれば各人の好みでもよい) を調べることで、元の集団(母集団)の中での特性値の分布状況を知ることができる。それぞれ独立に行われる測定を繰り返すことで、計測値がどのような分布をするかを知ることができる。

ある物質の物性値の測定を  $N$  回行い、それぞれ  $x_i$  ( $i = 1, 2, \dots, N$ ) という値を得たものとしよう。この  $N$  回の測定値(標本)と、この物性値の測定結果一般(母集団)の確率分布、特にその平均と分散の関わりについて考えよう。

まず  $N$  回の測定値について、その平均  $\bar{x}$  (標本平均と呼ぶ) を次式で定義する:

$$\bar{x} = (\sum x_i) / N \quad (i = 1, \dots, N)$$

中心極限定理から、標本平均は正規分布に従うと考えてよい。標本平均の平均は母集団の平均  $\mu$  に等しい ( $x_i$  は同じ確率分布に従うそれぞれ独立なランダム変数であることに注意):

$$\langle \bar{x} \rangle = (\sum \langle x_i \rangle) / N = (\sum \mu) / N = \mu$$

標本平均の分散は母集団の分散  $\sigma^2$  の  $1/N$  になる:

$$\langle (\bar{x}, \bar{x}) \rangle = \frac{1}{N^2} \langle (\sum x_i, \sum x_i) \rangle = \frac{1}{N^2} [\sum_i \langle (x_i, x_i) \rangle + \sum_{i \neq j} \langle (x_i, x_j) \rangle] = \frac{1}{N^2} N \sigma^2 = \sigma^2 / N$$

ここで個々の測定が統計的に独立であること ( $\langle (x_i, x_j) \rangle_{i \neq j} = 0$ ) を用いた。これをスローガン風にまとめれば「(標本)平均の平均は(母)平均だが、(標本)平均の分散の平均は(母)分散の  $1/N$  になる」ということになる。つまり 100 回の測定の平均値の分布の変動幅は 1 回の測定の  $1/10$  になると考えられる。

次に母集団の分散を推定することを考えよう。まず取り出してきた  $N$  個のデータについて、残差 2 乗和  $S$  を次式で定義する:

$$S = \sum (x_i - \bar{x})^2 = \sum x_i^2 - N \bar{x}^2$$

さて母集団の平均  $\mu$  からの偏差の 2 乗の和について次の式が成立する:

$$\sum (x_i - \mu)^2 = \sum (x_i - \bar{x} + \bar{x} - \mu)^2 = \sum (x_i - \bar{x})^2 + N(\bar{x} - \mu)^2 = S + N(\bar{x} - \mu)^2$$

ここで  $\langle (x_i - \mu)^2 \rangle = \sigma^2$  また  $\langle (\bar{x} - \mu)^2 \rangle$  は標本平均の分散で  $\sigma^2 / N$  なので、残差 2 乗和の平均は

$$\langle S \rangle = (N - 1) \sigma^2$$

残差 2 乗和の平均は母集団の分散の  $N - 1$  倍になる\*。

従って標本分散  $s^2$  を次式で定義すれば、標本分散の平均は母集団の分散になる:

$$s^2 = \frac{S}{N - 1}$$

今日では標本分散は通常上式で定義され、標本分散の平均は母集団の分散になる ( $\langle s^2 \rangle = \sigma^2$ )。標本標準偏差は標本分散の平方根である。なお標本標準偏差の平均は一般に母集団の標準偏差にならないことに注意する ( $\langle s \rangle \neq \sigma$ )。

## ☆検定・推定のこと

得られたデータから何らかの結果を導くにあたって、結果がどの程度もっともらしいかを評価する

---

\* 直接計算して  $\langle S \rangle = \langle \sum x_i^2 - N \bar{x}^2 \rangle = N \mu_2 - N(\mu^2 + \sigma^2 / N) = N(\mu_2 - \mu^2) - \sigma^2 = (N - 1) \sigma^2$  からも明らかだが、上記の導出の方が 2 次形式の形をよく示しているだろう。

手続き・方法について、あまり抽象的にならない形で簡単にまとめておく。

## ★仮説検定の考え方

現実の世界はさまざまな要素から構成されており、ある実験から得られたデータからただちに何らかの命題を証明することは困難である（たとえば同じ色だからといって、同じ物質とまでは言えない）。したがって多くの場合、「・・・と言えなくはない」といった形の反証の論理 **falsification** が用いられ、仮説検定と言われるものはそれを確率論の基礎の上に構成したものと言える。

通常行われる仮説検定においては、(1)帰無仮説  $H_0$  を立て\*、(2)その仮説が成り立つとして理論的結果（確率分布等）を導き、(3)実験結果が許容範囲内に収まらなければ（棄却域にあれば） $H_0$  を捨てる、という手順を取る。したがって実験結果が許容範囲内に収まったからといって、仮説が正しいとまでは言えない（棄却されない）ことに注意する必要がある。

棄却域の大きさ  $\alpha$  を有意水準 **significance level** と呼ぶ。有意水準は仮説  $H_0$  が正しいにもかかわらず棄却する確率で、この意味で危険率と呼ばれることもある。通常の検定においては  $\alpha = 0.05$  に取ることが多いが、分野や用途によって有意水準の取り方はさまざまである。たとえば平均値  $\mu$  標準偏差  $\sigma$  の正規分布  $N(\mu, \sigma^2)$  では  $\mu \pm 1.96\sigma$  の範囲内であれば 5% の有意水準で棄却されない。この  $\pm 1.96\sigma$  といった範囲を示す値を限界値 **critical value** と呼び、あからさまに有意水準の値ではなく、 $\pm\sigma$  の範囲であれば棄却しないなど、限界値で有意水準を定める場合も多い。

「推定」という作業は、実験結果を満たすように仮説を想定する作業と言える。典型的には実験である物質の物性値  $x$  について  $x_a$  という結果が得られたとき、 $x = x_a$  が満たされるという仮説が有意水準  $\alpha$  で成立する範囲を求め、「その物質の物性値  $x$  は  $x_a \pm \delta$  であった」という形で結論を下す。最小 2 乗法は、こうした物性値（あるいは分布を特徴づけるパラメータ）を推定するのに用いられる代表的な手法である。

仮説から導き出される理論的な結果にはさまざまなものが考えられる。実験の標準偏差があらかじめ分かっている場合には比較的単純であるが、実験の標準偏差があらかじめ分かっている場合には  $t$  分布や  $\chi^2$  分布、 $F$  分布に基づく検定が行われる。古くから各分布に対する数表なども整備されており、今日では Excel などの表計算ソフトでたいの作業はこなせ、さらに進んだ統計の専用パッケージとしては SAS や STATISCA (有料)、R (無料) などが著名である。

## ★標準偏差が分かっている場合の検定・推定

酸化還元滴定でシュウ酸マンガン ( $MnC_2O_4 \cdot 2H_2O$ ) とされる試料 0.2305 g 中に含まれるシュウ酸の質量が  $H_2C_2O_4$  として 0.1151 g であったとする。酸化還元滴定で定めるシュウ酸量の分布が正規分布に従い標準偏差が 0.20 % であるとあらかじめ分かっているものとして、この結果からどのような結論を導けるか考える。

仮説検定の手順に従い、仮説  $H_0$  として「試料は純粋なシュウ酸マンガンである」を立てよう。各物質の式量と与えられた標準偏差から「試料 0.2305 g 中に含まれるシュウ酸含量の実験結果は平均 0.1159 g で標準偏差 0.23 mg の正規分布に従う」という理論的結果が得られる。有意水準 5 %

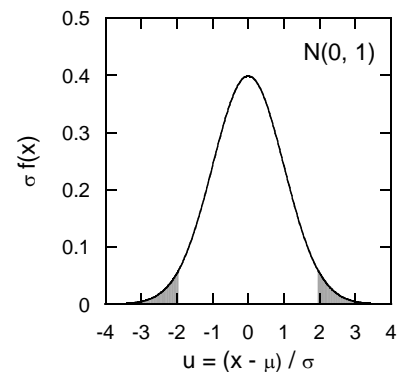


図 正規分布  $N(\mu, \sigma^2)$  を標準化した正規分布の確率密度。図中影を付けた  $|u| > 1.96$  の部分の面積は 0.05。

\*否定（反証）の対象にするのでこのように呼ぶ ( $H_0 = \text{null hypothesis}$ )。あからさまに対立仮説  $H_1$  (alternative hypothesis) を立てて考えることもある。

の棄却域として  $0.1154 \text{ g}$  以下  $0.1164 \text{ g}$  以上を取ろう (正規分布では  $\pm 1.96\sigma$  の範囲内に  $95\%$  が収まる)。実験結果は棄却域にあるので仮説は棄却され、「有意水準  $5\%$  で『試料は純粋なシュウ酸マンガンである』とは言えない」という結論を得る。

「推定」の立場からは、たとえば仮説として「試料は純度  $q$  のシュウ酸マンガンである」を立て、「試料  $0.2305 \text{ g}$  中に含まれるシュウ酸含量の実験結果は平均  $0.1159q \text{ g}$  で標準偏差  $0.23q \text{ mg}$  の正規分布に従う」という理論的結果を得て、実験結果  $0.1151 \text{ g}$  について

$$(0.1159q - 0.00046q) < 0.1151 < (0.1159q + 0.00046q)$$

を満たすように  $q$  の範囲を  $0.989 < q < 0.997$  と定め、「試料のシュウ酸マンガンの純度は  $(99.3 \pm 0.4)\%$  である (有意水準  $5\%$ )」といった結論を与えることになる。化学分析においては、こうした記述がむしろふつうである。なお推定の立場では、たとえば仮説の中の「シュウ酸マンガンである」ことについての判断が背景に押しやられてしまうことがあるので注意が必要になる。

### ★カイ 2 乗 ( $\chi^2$ ) 分布とカイ 2 乗検定

実験的に求められる偏差と母集団の分散 (あるいは標準偏差) とのかかわりを考える上で、カイ 2 乗  $\chi^2$  分布がしばしば登場する。標準正規分布  $N(0, 1)$  に従う  $\phi$  個の標準化正規変数  $u_i$  ( $i = 1, 2, \dots, \phi$ ) を取った時、その 2 乗和

$$x = u_1^2 + u_2^2 + \dots + u_\phi^2$$

の従う確率分布を、自由度  $\phi$  のカイ 2 乗分布と呼ぶ。カイ 2 乗分布の確率密度 (分布関数) は次式で与えられる:

$$f(x) = \begin{cases} C x^{\phi/2-1} \exp(-x/2) & x > 0 \\ 0 & x < 0 \end{cases}$$

ここで  $C$  は全領域にわたる積分が  $1$  になるように決める規格化定数。カイ 2 乗分布の平均は  $\phi$ 、分散は  $2\phi$  になる。中心極限定理が保証するように、カイ 2 乗分布は自由度 (足しこむ 2 乗した標準化正規変数の数) が大きくなると正規分布に接近するが、接近の度合いはあまり早くない。

先に見た残差 2 乗和の表式が、このカイ 2 乗分布の表式と似通っていることはすぐに見て取れるだろう。分散  $\sigma^2$  を与える実験を  $N$  回行い残差 2 乗和が  $S$  であったとすると  $\nu = S/\sigma^2$  は自由度  $N-1$  のカイ 2 乗分布に従うと見なせる。つまり残差 2 乗和の分散は  $2(N-1)\sigma^4$  と考えられ、標本分散  $S/(N-1)$  の標準偏差は  $N$  が大きければおよそ  $\sqrt{2/N}\sigma$  程度になる。したがって  $N$  回の測定から得られる標本標準偏差の相対的なばらつきは  $\sqrt{2/N}$  程度で、 $10$  回程度の測定で得られる標本標準偏差には  $50\%$  程度のばらつきがあることが分かる。

事象の出現頻度を理論 (仮説) と実験で比較し仮説の当否を判定する際にカイ 2 乗  $\chi^2$  検定がよく用いられる。典型的な適合度検定 **goodness of fit test** では  $N$  回の実験の実験結果を  $k$  個の級に分け、それぞれの級のデータの出現回数  $r_i$  と理論 (仮説) から予想される期待回数  $e_i$  から次のような値  $\chi^2$  を求める:

$$\chi^2 = \sum_i \frac{(r_i - e_i)^2}{e_i}$$

$N$  が大きい時、 $\chi^2$  は自由度  $k-1$  のカイ 2 乗分布に従うことが知られており、 $\chi^2$  の値が大きい時仮

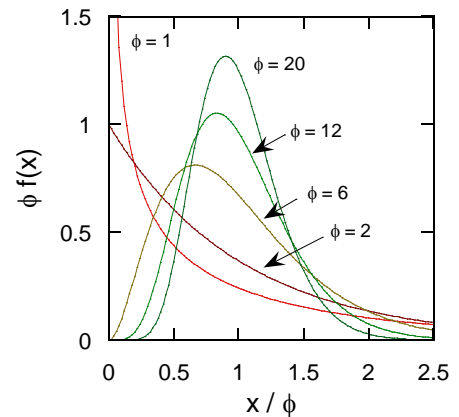


図 種々の自由度  $\phi$  に対するカイ 2 乗分布の確率分布 ( $\phi$  でスケールしてある)。自由度が大きくなるにつれゆっくり正規分布  $N(\phi, 2\phi)$  に漸近する。

説を棄却することになる。もし期待回数  $e_i$  が測定値から割り出される値 (標本平均など) を  $r$  個含む時は、自由度は  $k - r - 1$  になる。

「川沿いに男女が並んで座る時、左右どちらになるかは等確率である」という仮説を検討するため、鴨川沿いの 25 組の男女について調べたところ男が右側に座っていたのは 17 組であったとしよう。 $\chi^2$  を求めると  $\chi^2 = (8 - 12.5)^2/12.5 + (17 - 12.5)^2/12.5 = 3.24$  である。級の数に左右の 2 であるから自由度は 1 であり、自由度 1 で有意水準 5% となる  $\chi^2$  の値\*は 3.84.. なので 5% の有意水準で仮説  $H_0$  は棄却されず「有意水準 5% で、男女が並んで座る時、左右どちらになるかは等確率であるといえなくもない」という結論を得る。

### ★ $t$ 検定

未知の対象について何らかの実験を繰り返し行った際、その平均や分散などについて知られていないことが多い (というか分からないので実験する)。実験で求めた平均値がある仮説を満たすかどうかの判定には  $t$  検定がよく用いられる。

標準正規分布に従う変数  $u$  と自由度  $\phi$  のカイ 2 乗分布に従う変数  $v$  を取った時、 $x = u/\sqrt{v/\phi}$  は  $t$  分布に従う。 $t$  分布の確率密度 (分布関数) は次式で与えられる。

$$f(x) = C(1 + x^2/\phi)^{-(\phi+1)/2}$$

ここで  $C$  は規格化定数。

平均  $\mu$ 、標準偏差  $\sigma$  の正規分布に従う結果が得られると期待される実験を  $N$  回行ってデータ  $x_i (i = 1 \dots N)$  を得て、その標本平均が  $\bar{x}$ 、標本標準偏差が  $s$  であったとする。この時  $z = (\bar{x} - \mu)/(s/\sqrt{N})$  は自由度  $N - 1$  の  $t$  分布に従うことが知られている。自由度  $\phi$  の  $t$  分布の平均は 0、分散は  $\phi/(\phi - 2)$  で与えられ、自由度  $\phi$  が大きい場合には平均 0 分散 1 の標準正規分布に漸近する。

貯金箱にいっぱい入った百円玉から適当に 25 個取り出して重さを量り、その平均が 4.781 g、標本標準偏差が 0.031 g であったとしよう。この貯金箱に入っている百円玉の重さの平均が「通貨の単位及び貨幣の発行等に関する法律施行令」による量目 4.8 g と一致するかどうかを考えよう。帰無仮説として「貯金箱に入っている百円玉の重さの平均が 4.8 g に等しい」を立て、百円玉の重さの分布が正規分布に従うと考える。ここで  $z = (\bar{x} - \mu)/(s/\sqrt{N})$  を求めればこれは  $t$  分布に従うはずである (理論的帰結)。 $t$  分布の数表を参照すると、有意水準 5% の棄却域としては  $z$  が -2.064 以下 2.064 以上を取ればよい†。すると  $(4.781 - 4.8)/(0.031/\sqrt{25}) = -3.06$  であり、有意水準 5% で「貯金箱に入っている百円玉の重さの平均が 4.8 g に等しい」という仮説は棄却される。

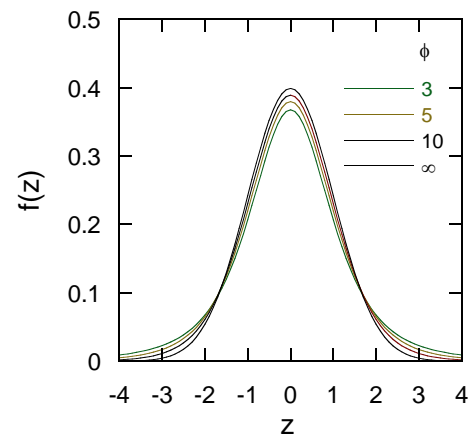


図 種々の自由度に対する  $t$  分布の確率密度。自由度が大きくなるにつれ分布は鋭くなり、正規分布  $N(0, 1)$  に漸近する。

\* Excel では CHINV(0.05,1) で値を得ることができる。Excel2007 以降であれば CHISQ.INV.RT(0.05,1) という形で呼び出す。

† Excel では Excel2007 以降であれば T.INV.2T 関数が用意されており T.INV.2T(0.05,24) という形で呼び出すことになる。Excel2007 以前にはこうした関数が用意されておらず TDIST 関数を利用することになる。TDIST 関数は  $u$  に対応する分布関数を与え TDIST(3.06,24,2) = 0.00538 という形で仮説を棄却する有意水準を与える。この場合「百円玉の重さの平均が 4.8 g」という仮説は 0.5% 程度しか有意性を持たない。



## ★F検定

2つの統計集団の比較にはF検定がよく用いられる。F検定は統計集団の分散に注目した手法で、分散分析 (ANOVA, analysis of variance) の主要なツールでもある。

それぞれの自由度 $\phi_1$ 、 $\phi_2$ のカイ2乗分布に従う変数 $v_1$ 、 $v_2$ を考えると $x = (v_1/\phi_1)/(v_2/\phi_2)$ は自由度( $\phi_1$ ,  $\phi_2$ )のF分布に従う。自由度( $\phi_1$ ,  $\phi_2$ )のF分布の平均は $\phi_1$ によらず $\phi_2/(\phi_2 - 2)$ である。F分布の確率密度 (分布関数) は次式で表される:

$$f(x; \phi_1, \phi_2) = C y^{-1 + \phi_1/2} (1 + y)^{-(\phi_1 + \phi_2)/2}$$

ここでCは規格化定数で $y = (\phi_1/\phi_2)x$ である。

F分布の累積分布がPになるxの値を $F(\phi_1, \phi_2; P)$ と表記してF分布に関する数表が整備されている。 $F(\phi_1, \phi_2; P)$ と $F(\phi_2, \phi_1; P)$ の間には $F(\phi_1, \phi_2; P) F(\phi_2, \phi_1; 1 - P) = 1$ の関係が成立する。

正規分布 $N(\mu_1, \sigma_1^2)$ 、 $N(\mu_2, \sigma_2^2)$ に従う母集団から、大きさ $N_1$ 、 $N_2$ の標本を得て、それぞれの標本分散が $s_1^2$ 、 $s_2^2$ であったとする。この時 $u = (s_1^2/s_2^2)(\sigma_2^2/\sigma_1^2)$ は自由度( $N_1 - 1$ ,  $N_2 - 1$ )のF分布に従う。つまり標本分散の比 $s_1^2/s_2^2$ が自由度( $N_1 - 1$ ,  $N_2 - 1$ )のF分布で極めて低い確率でしか実現されないものなら $\sigma_1^2 = \sigma_2^2$ とは見なせないことになる。

合金中の亜鉛の組成x%を重量分析と容量分析で決定する実験をK大学の優秀な学生たちが行い、重量分析は88件で標本分散 $s_w^2 = 3.453$ 、容量分析は23件で標本分散 $s_v^2 = 2.357$ であったという。この結果から重量分析と容量分析による分析値の分散に差があるかどうかを考える。「重量分析と容量分析の分散は等しい」という仮説が正しいならば、重量分析と容量分析の標本分散の比は自由度(87, 22)のF分布に従う。 $F(87, 22; 0.975) = 2.102$ であり\*、実験で求められた標本分散の比 $3.453/2.357 = 1.465$ はこれより小さく「有意水準5%で重量分析と容量分析の分散は等しいと言えなくはない」ということになる†。

## ★外れ値

測定値の中に他の値からかけ離れたもの (外れ値 outlier と呼ぶ) がある時、それを除外して取り扱う。外れ値を判定する際の明確な基準は難しいが、概ね標本標準偏差の2倍以上外れたものを外れ値とすることが多い。

データ量が多い時には、上位・下位の25%を省くといった手法がとられることがある。データ量があまり多くない時にはスミルノフ Smirnov (-グラフス Grubbs) 検定という手法もあるが、正規分布するという前提が満たされる保証はなく、積極的に採用する理由に乏しい。実際にはもっとも実験を積み重ねるのが最善という場合が多い。

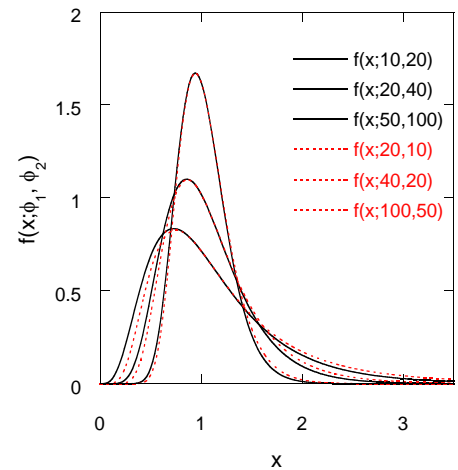


図 種々の自由度におけるF分布の確率密度。分子と分母を入れ替えたものを点線で示してある。

\* ExcelではFINV(0.025,87,22)で値を得ることができる。Excel2007以降であればF.INV(0.975,87,22)という形で呼び出す。

† 丁寧に考えるなら $s_w^2/s_v^2$ に関しxの小さい側に存在する棄却域 $x < F(87, 22; 0.025) = 0.545$ に入るかどうかの判断もする必要があるが $s_w^2/s_v^2 > 1$ なのでこのチェックをあえて行っていない。もし $s_w^2 < s_v^2$ であったなら $s_v^2/s_w^2$ について $F(22, 87; 0.975) = 1.835$ より大きいかどうかで判断する形を取るのが普通。 $F(22, 87; 0.975) = 1/F(87, 22; 0.025)$ に注意。

## 問題

学生番号 \_\_\_\_\_

氏名 \_\_\_\_\_

☆物体  $Q_1$ 、 $Q_2$ 、 $Q_3$  の重さ  $m_1$ 、 $m_2$ 、 $m_3$  を天秤で測定する。天秤の測定値のばらつき  $\delta$  の平均は 0 で標準偏差は  $\sigma$  であるとする。

この時、物体  $Q_1$ 、 $Q_2$ 、 $Q_3$  を 2 つずつ同時にとって、3 回の秤量で下記 3 種類の秤量値  $w_1$ 、 $w_2$ 、 $w_3$  を得たとしよう ( $\delta$  は測定の際のばらつき)。

$$w_1 = m_1 + m_2 + \delta_1$$

$$w_2 = m_2 + m_3 + \delta_2$$

$$w_3 = m_3 + m_1 + \delta_3$$

$w_1$ 、 $w_2$ 、 $w_3$  から  $m_1$ 、 $m_2$ 、 $m_3$  を  $m_1 = (w_1 + w_3 - w_2)/2$  といった関係式で得るとしたとき、得られる  $m_1$  (あるいは  $m_2$  または  $m_3$ ) の測定値の標準偏差はいくらになるか?

-----