確率と分布

Kosugitti の研究ノート

2004 年ごろに書いたものらしい

1　確率の考え方

日常会話でも，「今の確率は何 % ?」という表現をすることがよくある。「確率」という言葉はそのいかめしい字面のわりに，降水確率など，日常にもなじみのある言葉である。なじみやすいのは，この話がそもそもさいころの目がどれぐらいの確率で出てくるか，といった賭博に端を発しているからで，数学における確率はこの手の話を厳密にしているに過ぎない。そうはいっても，数学的に展開していけば，統計的にこれこれの数値が出る値が何 % であるか，といった話にまで膨らんでいくから不思議なものだ。

ともかく，数学の美しい特性＝厳密性を突き進めていくためには，日常では忘れられがちなものがある。すなわち，「今の確率は・・・」というとき，全体 (= 100%) が何なのか，ということである。確率は，パーセントで表されるのだから，割合のもとになる全体が何なのか，明らかでなければならない。

まずは，「起こりうる可能性，全ての組み合わせ」を数え上げる数学を考えよう。

1.1　順列，組み合わせ

1.1.1　順列

全体の中から，いくつかの要素を取り出す。あらゆる取り出し方を考えるとき，これを**順列**の問題という。

五枚のコイン，A,B,C,D,E があって，その中から最初の一枚を選び取る方法は，5 通りある。二枚目を選ぶときは，一枚減っているから，4 通りある・・・。こうして考えると，$5\times4\times3\times2\times1=120$ 通りあることがわかる。これを特に，5! と書く。

一般に，n 個のなかから，任意に r 個とって一列に並べる並べ方は，

$$ {}_nP_r = n(n-1)(n-2)\cdots(n-r+1) = \frac{n!}{(n-r)!} \tag{1} $$

通りある。P は順列 (permutation) の P ね。これは取り出す順番を考慮したものだが，順番に関わらないのであれば，組み合わせ (combination) を考えることになる。

一般に，n 個の異なるものから，任意に r 個取り出す組み合わせの数は，

$$ {}_nC_r = \frac{{}_nP_r}{r!} = \frac{n(n-1)(n-2)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!} \tag{2} $$

になる。

1.2 二項定理

$(a+b)^2 = a^2 + 2ab + b^2$ という公式があるが，これが 2 乗じゃなくて，3 乗，4 乗となったらどのような展開式になるか。という問いに答えるのが**二項定理**で，これは組み合わせの考えが役に立って，$a^{(n-r)}b^r$ の係数は，n 個の因数 $a+b$ から，a を $n-r$ 個，あるいは b を r 個選ぶ組み合わせの数，${}_nC_r$ に等しい。一般化すると，

$$\begin{aligned}(a+b)^n &= \sum_{r=0}^{n} {}_nC_r a^{n-r}b^r \\ &= {}_nC_0 a^n + {}_nC_1 a^{n-1}b + {}_nC_2 a^{n-2}b^2 + \cdots + {}_nC_r a^{n-r}b^r + \cdots + {}_nC_n b^n \\ &= a^n + na^{n-1}b + \tfrac{n(n-1)}{2}a^{n-2}b^2 + \cdots + \tfrac{n!}{(n-r)!r!}a^{n-r}b^r + \cdots + b^n\end{aligned} \tag{3}$$

あとあと出てくる考え方でもあるので，覚えておいて欲しい。

2 確率変数

2.1 確率の定義

数学的な確率は，以下のように定義される。

ある試行について，起こりうる結果の全体 (標本空間) の大きさが n で，その中でどの (根元) 事象も同程度に確からしく起こるとする。標本空間の中で，ある事象 E をとり，E のおこる場合の数が r であるとき，E の確率 P(E) を

$$P(E) = \frac{r}{n}$$

と定義する。

これは，例えばさいころは出る目の数が 1,2,3,4,5,6 の六つが全てであり ($n=6$)，イカサマさいでない限りは，どの目も均等に出るハズだから，どの目の出る確率も 1/6 である，といっているに過ぎない。

しかし，これはあくまでも数学的な話であって，日常的な感覚で使われる確率とは少し異なる。なぜなら，「雨が降る確率」というようなものは，あくまでも過去の経験から行くと，この空模様なら x% ぐらいの確率で降りますね，という話なのである。全体もわからなければ，事象の確からしさもわからない。そういう場合は，経験的確率の定義を用いる。それは以下のようなものである。

n 回試行を行った結果，ある事象 E が r 回起こったとする。n を大きくしていくとき，r/n が一定の値 p に近づけば，E の確率 P(E) を

$$P(E) = p = \lim_{n\to\infty} \frac{r}{n}$$

とする。

確率の規則は，以下のように定められている (確率の公理)。

標本空間 S の各事象 E に対して，次の 3 つの条件を満たす実数 P(E) が存在するとき，P(E) を事象 E が起こる確率という。

1. $0 \leq P(E) \leq 1$

2. $P(S) = 1, P(\phi) = 0$

3. $E_1, E_2, E_3, \cdots$ が互いに排反な事象の時，

$$P(E_1 \cup E_2 \cup E_3, \cdots) = P(E_1) + P(E_2) + P(E_3) + \cdots$$

これは，そんなに難しい話ではない。最初のものは，確率は 0 から 1 の間の数値ですよ，ということである。「200% 優勝します。」といった表現は，確率論上あり得ない。第二のものは，全体は 100% で，何も起こらない確率は 0% ですよ，という意味だ。第三のものは，それぞれが独立して起こるなら，どれらかが起こる確率はそれを足したものになる，というもの。理想的なさいころがあるとして，1 が出るのは 1/6，2 が出るのは 1/6 の確率であれば，1 か 2 が出るのは 1/6+1/6=2/6 ですよ，というだけの話だ。

数学の公理というのは面白いもので，例えばこちらが確率だと意図していなくても，上の条件を全て満たすような変数を作ったら，確率と見なすこともできる。柔軟な頭の使い方ができるような人間になりたいものですね。

2.2 確率変数

我々がデータを取るとき，例えば五件法で 1 から 5 までの数値で回答を求めたときに，1 に○をしたのが何人，2 に○をしたのが何人・・・と数えていくことができる。そこから，1 を取ったのは全体の何 % だった，という数値を出すこともできる。この相対度数は，0 から 1 までの値をとるし，回答者は 1〜5 以外の数値を書かない。また，1 に○をすることとそれ以外に○をすることの間に関連がないと仮定しよう。そうすると，上の公理から，回答がどのように得られるかは確率の公理に従うと考えることができる。

我々のデータが確率の公理に支配されている，と想定することで，以下の推測統計なり，多変量解析なりが応用できるようになる。検定するというのは，確率で得られた変数から全体を推定することだし，多変量解析では最尤法が，データが生じる確率が最も多いとするとどうなるか，ということを考える (最も尤もらしいことを考える) というものだ。

確率で得られる変数のことを，**確率変数**といい，確率を表現した関数を**確率関数**という。これらを扱うことは，数学的には積分の世界に足を踏み入れることになる。

また，**離散変数**と**連続変数**の考え方も区別しなければならない。離散変数はデジタルな値，連続変数とはアナログな値である。つまり，離散変数が 1,2,3 とあったとして，この数値の間にはなにもない。連続変数は，例えば 1 と 2 の間に 1.5 があり，1 と 1.5 の間にも 1.25 があり，1 と 1.25 の間にも 1.125 があり・・・とどこまで行っても点が取れる数値であることを意味する。

データとして得られるのは，離散変数である。「そう思う」と「ややそう思う」の間に○をつけられたら，極値をとるか欠損値として扱うか，といったことをしてしまうぐらいだ。しかし，一般に七件法以上であれば，数値の連続性を仮定しても差し支えない，といわれている。この連続性の仮定ができるから，正規分布を応用した様々な分布を用い，検定・推定が行われるわけである。

連続的な数値は，微分・積分が可能で，数学的には大変扱いやすいのだが，確率変数や積分の考え方は，生のデータに慣れ親しんだ人間には少し違和感を感じるところのものであるかもしれない。ここが踏ん張りどこ

ろである。

2.2.1 離散的な場合

確率変数 X が離散的で，有限個の値を取るとする。X が $x_i(i = 1, 2, \cdots, n)$ となる確率を

$$P(X = x_i) = p_i$$

と表す。このとき，X のとる値それぞれに対して P(X) が定まるから，P(X) は X の関数で，

$$f(x) = \begin{cases} p_i & (x = x_i \text{ のとき}) \\ 0 & (\text{その他の x}) \end{cases} \tag{4}$$

と書ける。具体的な例を使って説明してみよう。さいころは 1〜6 の目が出るし，全て確率は 1/6 である。1 が出る確率は $f(1) = 1/6$ である。ここまではよい。では下の「その他の x」とは何だろうか。これは例えば，1 が出る確率を知りたいとき，x_2 や x_3 は関係ないですよ，ということだ。x_2 のとき，1 が出る確率は 0 だというわけである。当たり前ですね。

この例だと，

$$\sum_{i-1}^{6} f(x_i) = f(x_1) + f(x_2) + \cdots + f(x_6) = 1$$

である。これもご理解頂けよう。

2.2.2 連続的な場合

連続的な場合は，X がある値を取る，というより X がある範囲に入る，すなわち $x < X \leq x + \Delta x$ にある確率はどれぐらいか？という表現がふさわしい[*1]。この確率の表現は，積分を使って

$$P(x < X \leq x + \Delta x) = \int_x^{x+\Delta x} f(y)dy \tag{5}$$

と書く。

確率変数 X が取りうる範囲を $a \leq X \leq b$ とすると，全ての範囲は 1 になるのが確率の公理だから，

$$\int_a^b f(x)dx = 1 \tag{6}$$

である。それ以外の区間では，確率が 0 になるから，この区間は延長して

$$\int_{-\infty}^{\infty} f(x)dx = 1 \tag{7}$$

でもよい。

2.2.3 期待値

確率変数の平均値は特に，**期待値**と呼ばれ，次の式で表される。

$$\mu = \begin{cases} \sum_{i=1}^{n} x_i f(x_i) = x_1 f(x_1) + x_2 f(x_2) + \cdots + x_n f(x_n) & (\text{離散変数}) \\ \int_{-\infty}^{\infty} x f(x)dx & (\text{連続変数}) \end{cases} \tag{8}$$

[*1] ここで Δx とは x の変化量という意味。$x + \alpha$ でもなんでもいいんだが。

例 1 (さいころの期待値)**.**

$$\mu = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{21}{6}$$

例 2 (サマージャンボ宝くじの期待値)**.** サマージャンボ宝くじは，以下のようになっている *(2004* 年度調べ*)*。確率と数値を積和していくのが期待値なので，

表 1　1 ユニット，1000 万通りあたりの当選金

等級	当選金	本数	当選確率
1 等	2 億円	1	0.00001 %
前後賞	5 千万円	2	0.00002 %
組違い	10 万円	99	0.00099 %
2 等	1 億円	1	0.00001 %
3 等	100 万円	10	0.00010 %
4 等	10 万円	100	0.00100 %
5 等	3000 円	100000	1.00000 %
6 等	300 円	1000000	10.00000 %
夏ラッキー賞	1 万円	40000	0.40000 %

$$\mu = 200000000 \times 0.0000001 + 50000000 \times 0.0000002 + \cdots + 10000 \times 0.4 = 143.0$$

となる。つまり，一枚 *300* 円なので，一枚当たり *157* 円の損[*2][*3]。

2.2.4　分散

確率変数でもなんでも，重要なのは分散の方で，次の式で定義される。

$$\sigma^2 = \begin{cases} \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) = (x_1 - \mu)^2 f(x_1) + (x_2 - \mu)^2 f(x_2) + \cdots + (x_n - \mu)^2 f(x_n) & \text{(離散変数)} \\ \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx & \text{(連続変数)} \end{cases} \tag{9}$$

3　分布いろいろ

3.1　二項分布

分布の最も基本的なものは，二項分布，あるいはベルヌーイ分布と呼ばれる。これは，「表か裏か」といったふたつの状態のうち，一方が出る確率を求める変数である。例えばコイントスのようなものであれば，表も裏も五分五分だということは経験的に知っているが，6:4 で表が出るコインとか，1:9 で裏が出るコインのようなイカサマコインを投げたときに，どのような結果になるのか，を示すのが二項分布である。

普通のコインの例から始めよう。普通のコインは裏も表も，どちらの出る確率も 0.5 である。このコインを 5 回投げて，表が出る回数を数えることを考える。表の出る回数を確率変数 X として捉えるわけだ。このときの $f(x)$ はどうなるだろうか？

[*2] バラで買った場合。

[*3] 主観的には，毎回 1 等が当たるような気がしてならない。

まず，表が 1 回だけ出る場合を考えよう。$X = 1$ である。これは，異なる 5 回のトライアルから，1 回だけ特殊なことが起こりうる場合を取り出す組み合わせの数だから，

$$ {}_5C_1 = \frac{5!}{1!4!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{1 \times 4 \times 3 \times 2 \times 1} = 5 $$

の，5 回ありえる。表が 1 回だけ出るというのは，最初の 1 回が表で残りの 4 回が裏だから，$\left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4$ となり，この確率に従うのが 5 回あるのだから，

$$ 5 \times \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4 = 5 \times \frac{1}{2} \times \frac{1}{16} = \frac{5}{32} $$

となる。

同様に，一回も表が出ない場合は $f(0) = {}_5C_0 \times \left(\frac{1}{2}\right)^0 \times \left(\frac{1}{2}\right)^5 = 1 \times 1 \times \frac{1}{32}$ である。これを続けると表 2 のような結果が得られ，それをグラフにしてみたのが図 1 である。

表 2　コインの表が出る確率

確率関数	確率
f(0)	$\frac{1}{32} = 3.13\%$
f(1)	$\frac{5}{32} = 15.63\%$
f(2)	$\frac{10}{32} = 31.25\%$
f(3)	$\frac{10}{32} = 31.25\%$
f(4)	$\frac{5}{32} = 15.63\%$
f(5)	$\frac{1}{32} = 3.13\%$

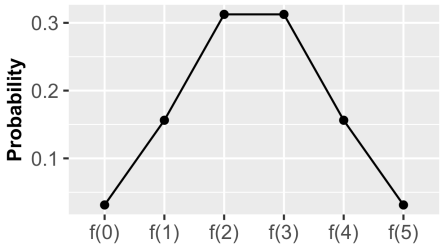

図 1　普通のコインの確率関数は対称型

もしこれがいびつなコインで，表が出る確率が 1/5 であったとすれば，この図はもっと歪んで，図 2 のようになる。この図からわかるように，表が出る回数が 1/5 のコインで，表が 5 回連続で出ることは滅多にない (0.032%)。

ともかく，このような分布を求めるのが二項分布の考え方だ。一般に，ある事象 A の起こる確率 $P(A)$ があって，n 回試行するときに A が x 回起こる確率は，

$$ f(x) = {}_nC_x p^x (1-p)^{n-x} \tag{10} $$

であり，これを二項分布という。これが分布の基本中の基本だと思っておいて欲しい。

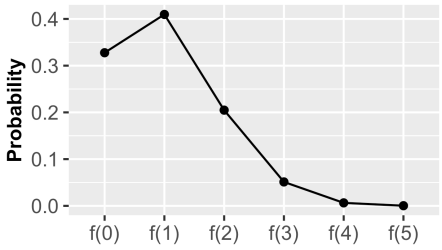

図 2　五回に一回しか表が出ないコインでは

ところで，この表が出る確率 1/5 のいびつコインでも，試行を繰り返せば分布は対称系に近づいていく (図 3)。どちらか一方が出続ける可能性は，等しく減っていき，どこかで (この例だと表が 10 回というところで) バランスが取れるようになるわけだ。

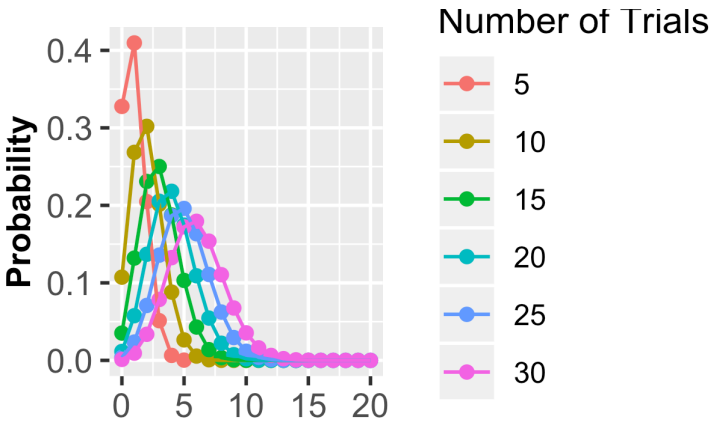

図 3　いびつコインでも反復すると

また，コインの表面が出る確率を変えていくと，ちょうど半々のところで分布が対称形になることもわかる (図 4)。これらの対称形はどこかで見たことがありませんか。そう，正規分布です。正規分布は，二項分布の n を極限にまで大きくしたものとして得られる。正規分布の話をする前に，まず二項分布の平均と分散を考えなければならない。

3.2　二項分布の平均と分散

二項分布は式 10 で与えられたが，$1 - p = q$ とすると

$$f(x) = {}_nC_x p^x q^{n-x} \tag{11}$$

となり，式 3 と同じ形になることに気づいて頂けるだろう。二項分布の平均 $\mu = \sum_{x=0}^{n} x f(x)$ は，式 11 に x を書けた形で得られるので，式 11 を p で微分すれば $\sum_{x=0}^{n} x {}_nC_x p^x q^{n-x}$ の形が出てくる。また，式 11 は二項定理

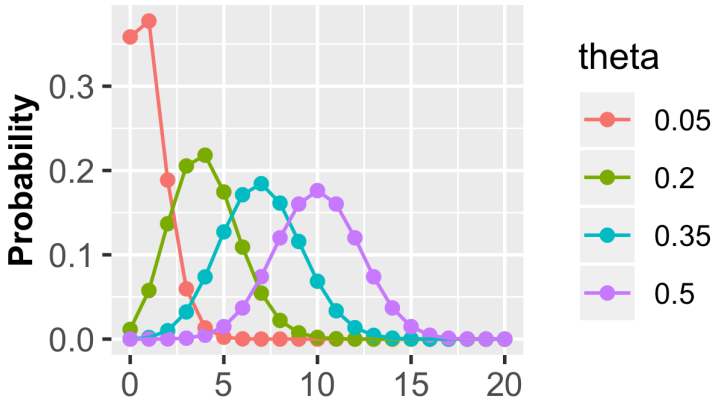

図 4　様々なコインを試してみると

の形なのだから，$(p+q)^n$ に等しくもある。そうすると，

$$\sum_{x=0}^{n} {}_nC_x p^x q^{n-x} = (p+q)^n$$

両辺を微分して

$$\sum_{x=0}^{n} x {}_nC_x p^{x-1} q^{n-x} = n(p+q)^{n-1}$$

p をかけて左辺を元に戻す。また，$p+q-1$ を応用する。

$$\sum_{x=0}^{n} x {}_nC_x p^x q^{n-x} = n \cdot 1^{n-1} \cdot p$$

まとめると

$$\sum_{x=0}^{n} x {}_nC_x p^x q^{n-x} = np = \mu$$

となる。二項分布の平均は，あることが生じる確率に試行回数をかけたもの，というわけだ。これは直感的にもわかりやすい。

　さて，分散は 9 式にもあるように，

$$\begin{aligned} \sigma^2 &= \sum_{x=0}^{n} (x-\mu)^2 f(x) \\ &= \sum_{x=0}^{n} x^2 f(x) - \mu^2 \\ &= \sum_{x=0}^{n} x^2 {}_nC_x p^x q^{n-x} - \mu^2 \end{aligned} \tag{12}$$

である。

これはさっきと同じように，

$$\sum_{x=0}^{n} {}_nC_x p^x q^{n-x} = (p+q)^n$$

両辺を微分して

$$\sum_{x=0}^{n} x {}_nC_x p^{x-1} q^{n-x} = n(p+q)^{n-1}$$

更に微分して

$$\sum_{x=0}^{n} x(x-1) {}_nC_x p^{x-2} q^{n-x} = n(n-1)(p+q)^{n-2}$$

両辺に p^2 をかけて

$$\sum_{x=0}^{n} (x^2 - x) {}_nC_x p^x q^{n-x} = n(n-1)p^2$$

$$\sum_{x=0}^{n} x^2 {}_nC_x p^x q^{n-x} - \sum_{x=0}^{n} x {}_nC_x p^x q^{n-x} = n(n-1)p^2$$

左辺第二項は平均値だから，

$$\sum_{x=0}^{n} x^2 {}_nC_x p^x q^{n-x} - np = n(n-1)p^2$$

である。

ここから，分散は

$$\sigma^2 = n(n-1)p^2 + np - (np)^2 = np(1-p)$$

となる。

3.3 正規分布

二項分布の平均と分散がわかれば，これを標準化した変数 Z を考えることができる。

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

これの分布関数が**正規分布**で，これは

$$g(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \tag{13}$$

に従う。13 式は平均 0，分散 1 であるので，**標準正規分布**と呼ばれ，$N(0,1)$ と書く。平均 μ，分散 σ^2 の正規分布は

$$g(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \tag{14}$$

で表され，$N(\mu, \sigma^2)$ と表す。

正規分布関数で用いられている e は，自然対数の底と呼ばれるもので，微分の中では不思議な性質を持つものである。具体的な説明省略するが，関数の形を目で見て理解しておこう。$f(x) = e^x$ は，$f(x) = exp(x)$ とも表記されるが，図 5 のような関数である。普通の指数関数ですね。で，x の変わりに $-x$ を入れてやると，図 6 のようになる。左右逆転するのもよくわかる。正規分布のグラフでは，指数を 2 乗して 2 で割ってある。まず 2 で割る方を図 7 に示す。これで 2 乗させると，あら不思議 (図 8)。釣り鐘型のできあがり。

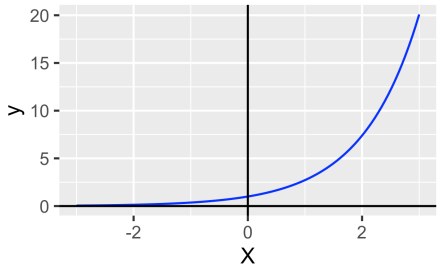

図 5　$f(x) = exp(x)$

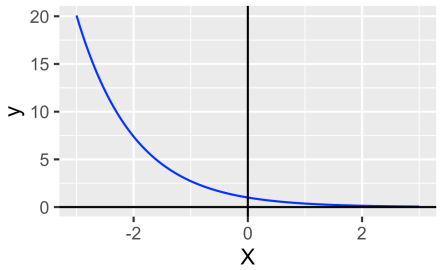

図 6　$f(x) = exp(-x)$

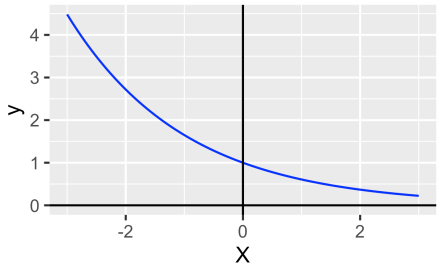

図 7　$f(x) = exp(-x/2)$

さて，13 式や 14 式は，(z) を連続変数として考えているが，離散的な値を用いた二項定理からどうして連続変数になるのだろうか。

まず，

$$z = \frac{x-\mu}{\sigma} = \frac{x-np}{\sqrt{np(1-p)}}$$

が原型であるが，これは離散的な x の値を使っている。x は $0, 1, 2\cdots$ と，間隔 1 のステップであるし，それに対する z の増分は

$$\frac{(x+1)-np}{\sqrt{np(1-p)}}$$

だから

$$x+1-x = \frac{x-np}{\sqrt{np(1-p)}} + \frac{1}{\sqrt{np(1-p)}} - \frac{x-np}{\sqrt{np(1-p)}}$$

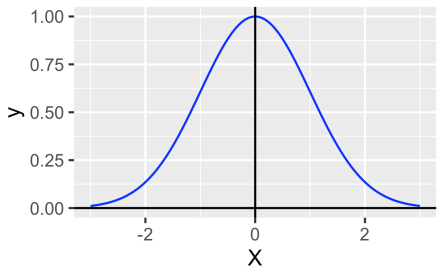

図 8　$f(x) = exp(-x^2/2)$

より，

$$\Delta z = \frac{1}{\sqrt{np(1-p)}}$$

である。さて，z を Δz を用いて表現した式，

$$z = \frac{x - np}{\sqrt{np(1-p)}} = \Delta z(x - np)$$

から，

$$z = \Delta z x - \Delta z np$$

よって

$$x = np + \frac{z}{\Delta z}$$

が得られる。区間幅 Δz は，n を大きくしていくとどんどん縮んでいき，$n \to \infty$ にまで飛ばせば (極限を取れば)，z を連続変数と考えても良いことになろう，というわけである。

さて，

$$f(x) = {}_nC_x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

で，x が $x+1$ に増加すると，

$$f(x+1) = {}_nC_{x+1} p^{x+1} (1-p)^{n-x-1} = \frac{n!}{(x+1)!(n-x-1)!} p^{x+1} (1-p)^{n-x-1}$$

であるから，

$$\frac{f(x+1)}{f(x)} = \frac{n! p^{x+1} (1-p)^{x-n-1}}{(x+1)!(n-x-1)!} \times \frac{x!(n-x)!}{n! p^x (1-p)^{x-n}} = \frac{(n-x)p}{(x+1)(1-p)}$$

である[*4]。

ところで，我々が知りたいのは Z が従う分布 g(z) であるが，二項分布は x の関数で得られている。x の関数を z の関数に変えたいとき，関数の形はどうあれ，$P(x \le X \le x + \Delta x)$ の確率と $P(z \le Z \le z + \Delta z)$ は同じであるようにしなければならない。x の変化 $\Delta x = 1$ に対する z の変化は Δz であるから，x から z に変数変換しても確率が変わらないという条件は

$$g(z) = \frac{\Delta x}{\Delta z} f(x) = \frac{1}{\Delta z} f(x)$$

[*4] 難しい展開ではないので，自分で検算するように。テキストには書ききれないほど煩雑ではあるが，階乗が互いに消し合う部分がうまく働いて，こんなにすっきりした形になります。階乗の計算は面白い。

である。従って，

$$\frac{g(z+\Delta z)}{g(z)} = \frac{f(x+1)}{f(x)} = \frac{(n-x)p}{(x+1)(1-p)}$$

である。さて，$x = np + z/\Delta z$ だったから，

$$\frac{g(z+\Delta z)}{g(z)} = \frac{(n-np-z/\Delta z)p}{(np+z/\Delta z+1)(1-p)} = \frac{np(1-p)-pz/\Delta z}{np(1-p)+(1-p)(z/\Delta z+1)}$$

である。

いつぞや，$\Delta z = 1/\sqrt{np(1-p)}$ であることを示したが，両辺を 2 乗して変形すると，$np(1-p) = 1/(\Delta z)^2$ であることがわかる。これをこの式に入れると

$$\frac{g(z+\Delta z)}{g(z)} = \frac{\frac{1}{(\Delta z)^2} - p\frac{z}{\Delta z}}{\frac{1}{(\Delta z)^2} + (1-p)(\frac{z}{\Delta z}+1)} = \frac{1-pz\Delta z}{1+(1-p)\{z\Delta z+(\Delta z)^2\}}$$

さて，この式を使って z が Δz だけ増えたときの，増加量を計算する[*5]。

$$\frac{g(z+\Delta z)}{g(z)} = \frac{g(z)}{\Delta z}\left(\frac{g(z+\Delta z)-g(z)}{g(z)}\right)$$

括弧内は

$$= \frac{1-pz\Delta z-1-(1-p)\{z\Delta z+(\Delta z)^2\}}{1+(1-p)\{z\Delta z+(\Delta z)^2\}}$$

これの分子は特に

$$\begin{aligned} &= 1-pz\Delta z-1-(1-p)\{z\Delta z+(\Delta z)^2\} \\ &= -pz\Delta z-(z\Delta z+(\Delta z)^2-pz\Delta z-p(\Delta z)^2) \\ &= -z\Delta z-(\Delta z)^2+p(\Delta z)^2 \\ &= -z\Delta z-(1-p)(\Delta z)^2 \end{aligned}$$

であり，結局

$$\frac{g(z+\Delta z)-g(z)}{\Delta z} = \frac{-z-(1-p)\Delta z}{1+(1-p)\{z\Delta z+(\Delta z)^2\}}g(z)$$

が，得られる。このとき，$\Delta z \to 0$ の極限を取ると，

$$\frac{dg(z)}{dz} = -zg(z)$$

となり，

$$\frac{d}{dz}\log g(z) = -z$$

となる。積分すると，

$$\log g(z) = -\frac{1}{2}z^2 + C'$$

となり (C' は積分定数)，

$$g(z) = Ce^{-z^2/2}$$

が得られる。

[*5] これ大変よ。頑張ってついてきてね。

積分定数 C は，確率密度の積分が 1 である。また，$\int_{-\infty}^{\infty} e^{-x^2/2}dx = \sqrt{2\pi}$ であることから，

$$1 = \int_{-\infty}^{\infty} g(z)dz = C\int_{-\infty}^{\infty} e^{-z^2/2}dz = C\sqrt{2\pi}$$

となり，$C = 1/\sqrt{2\pi}$ であることが示される。

こうして得られる正規分布は，平均 $\mu \pm \sigma$ の範囲に変数が 68.26% の確率で存在し，$\mu \pm 2\sigma$ の範囲内には，95.44% が，$\mu \pm 3\sigma$ の範囲には 99.73% が存在する (図 9)。

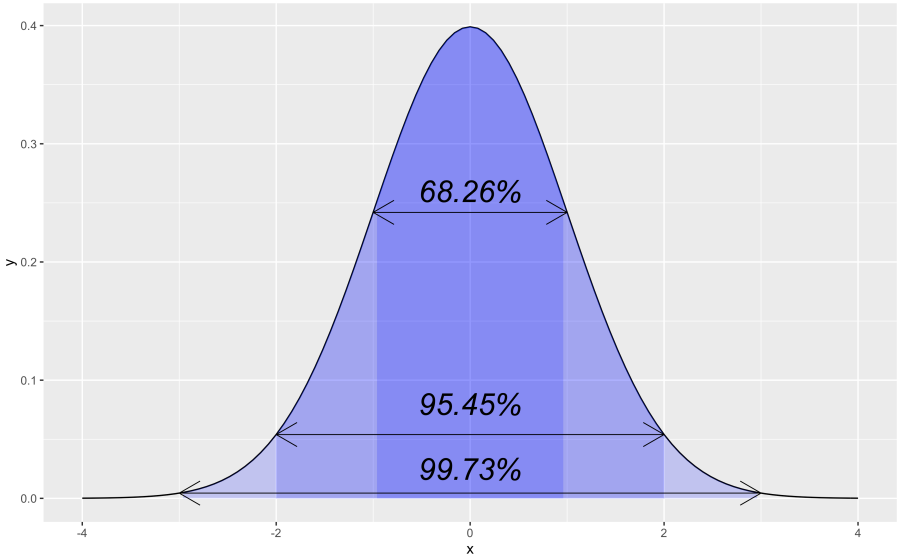

図 9　正規分布の性質

3.4　χ^2 分布

さて，この章の締めくくりとして，いくつかの実用的な分布について紹介しておこう。数学的な厳密さは他書に譲るとして，以下ではイメージで分布を捉えておいてもらいたい。

まず最初は χ^2 分布である。標準正規分布 $N(0,1)$ に従う確率変数 X があったとして，$Z = X^2$ の従う分布 $T_1(z)$ はどうなるだろうか。これは式で言うと

$$T_1(z) = \begin{cases} \frac{1}{\sqrt{2\pi}} z^{-1/2} e^{-z/2} & (z > 0) \\ 0 & (z \leq 0) \end{cases}$$

であり，図にすると図 10 のようになる。

さて，今度は確率変数が二つあったとしよう。それぞれ X_1 と X_2 で，いずれも標準正規分布に従うものとする。このとき，$Z = X_1^2 + X_2^2$ が従う分布は図 11 のようになるし，同様に三つの確率変数を足し合わせた変数，$Z = X_1^2 + X_2^2 + X_3^2$ の従う分布は図 12 のようになる。

急に何をし出すか，と思われた読者も多いだろうから，そろそろ説明を。我々がサンプルデータを使って統計的にものをいうとき，得られたデータはサンプル毎の確率変数が取った値だと考える。自然界のデータは一般に正規分布を仮定できるから，それぞれは標準正規分布 (あるいは平均 μ，分散 σ^2 の正規分布) に従うもの

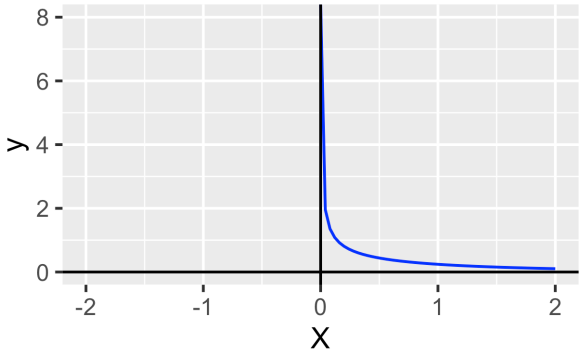

図 10　標準正規分布を二乗した確率変数の分布

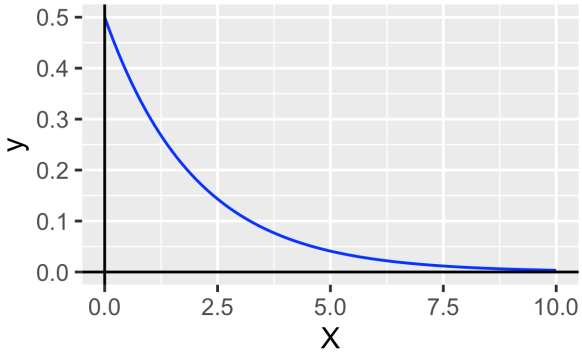

図 11　二つの二乗した標準正規分布を足した分布

とすると，表れてくる統計的指標はこれらの確率変数を足して得られる分布に従うと考えられる。もちろんいくつの変数を足すかによって，分布関数は多少変わってくるが，三つ以上の場合は基本的に図 12 の形をしている。n 個の標準正規分布の二乗和が作る分布は，**自由度 n の χ^2 分布**と呼ばれる。

一般的な言い方をすると，

$N(0,1)$ に従う正規母集団から，大きさ n の標本 $X_1, X_2, \cdots, X_n$ 個を無作為抽出したとき，

$$X = X_1^2 + X_2^2 + X_3^2 + \cdots + X_n^2$$

は自由度 n の χ^2 分布に従う。

といえる。もちろん標準化していない場合でも，

$N(\mu, \sigma^2)$ に従う正規母集団から，大きさ n の標本 $X_1, X_2, \cdots, X_n$ 個を無作為抽出したとき，

$$Z = \frac{1}{\sigma^2}(X_1 - \mu)^2 + (X_2 - \mu)^2 + (X_3 - \mu)^2 + \cdots + (X_n - \mu)^2$$

は自由度 n の χ^2 分布に従う。

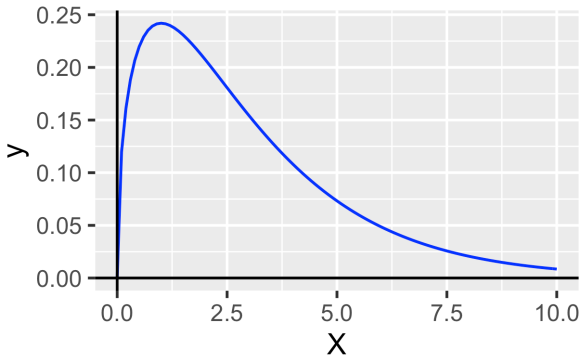

図 12　三つの二乗した標準正規分布を足した分布

と言い換えられる。

χ^2 分布は確率変数の二乗和の分布だから，これは確率変数の分散の分布を表しているといっても良い。

3.5　F 分布

次に，分散の分布，χ^2 分布が二つあったとき，その比の分布を考える[*6]。

X_1 と X_2 が互いに独立で，それぞれ自由度 m, n の χ^2 分布に従うものとする。これを**自由度 (m,n) の F 分布**といい，$F(m,n)$ **で表す**。以下にいくつかの F 分布を示しておこう。

まずは $F(1,2)$ の分布を図 13 に示した。これは χ^2 の自由度 1 の分布とよく似ている。

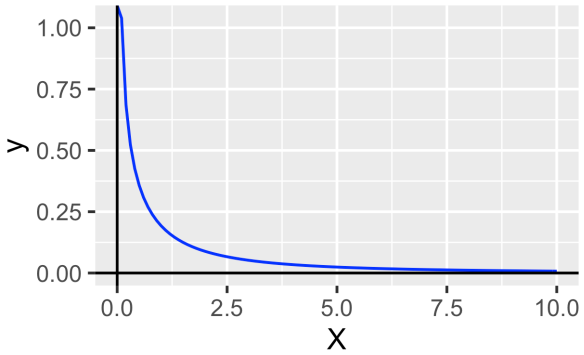

図 13　$F(2)$ の分布

同じく $F(2,2)$ の分布は図 14 のようになる。

分子の自由度が $3 \geq m$ になると，形が少し変わる。図 15 がそれである。

図 16 は，$F(5,10)$ の分布である。

F 分布は，χ^2 分布の比の分布，つまり分散の比の分布だと考えればよい。これらの形が理論的に求まっているから，実際に手にしたデータが出現する確率はどれくらい何だろうか，といった推測が可能になってくるの

[*6] なんでそんな面倒なことを，と思うだろうが，これは先の分散分析などで使うものであるので，我慢して下さい。

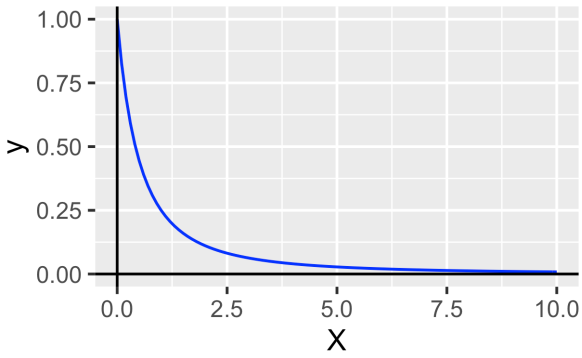

図 14　$F(2,2)$ の分布

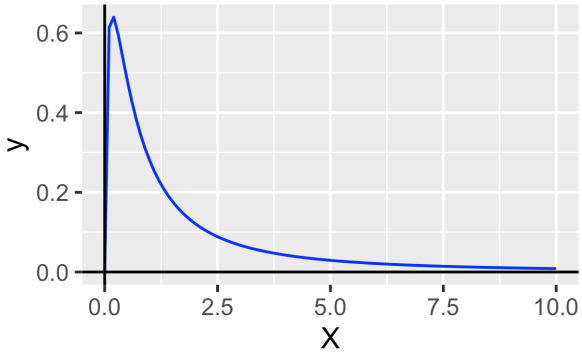

図 15　$F(3,2)$ の分布

である。

3.6　t 分布

最後にひとつ，$F(1,n)$ の分布に関連した t 分布を示しておこう。一方の自由度が 1 だとわかっている場合に使える分布で，本質的には $F(1,n)$ と大きく変わらないが，いくつかの便利な特徴があるので実際の統計処理でよく使われる。すなわち，

- t 分布の自由度が ∞，すなわち $t(\infty)$ の分布は標準正規分布に一致する。
- $F(1,n) = t(n)^2$

という性質がある。分布の形は標準正規分布とそっくりである (図 17)。これらは以下の章で使うものなので，とりあえず形とイメージを把握されたし。

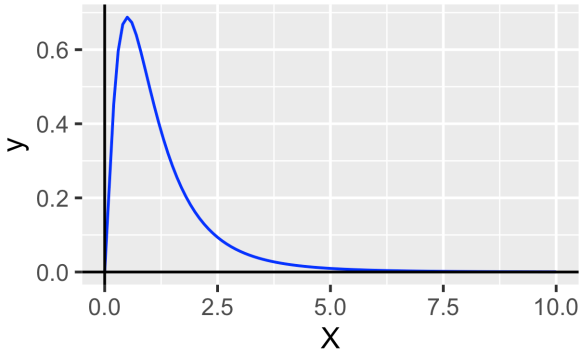

図 16　$F(5, 10)$ の分布

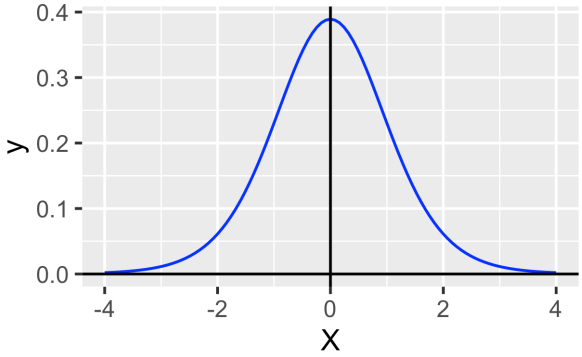

図 17　自由度 10 の t 分布

4　確率分布の近似計算

これらの形は理論的には求まっているものの，実際の推定，検定などの際はある点より上の面積 (上側確率) やパーセント点 (上から $x\%$ の面積を占める点はどれぐらいか) が必要になってくる。

統計の本を見れば，たいてい後ろに付表として載っているが，あれをどうやって計算したんだろう？と気になる人もいるに違いない。これらは近似的にしか計算できないが，少なくともその計算方法は記しておく価値があるだろう[*7]。

4.1　正規分布の上側確率

標準正規分布 $N(0, 1)$ の上側確率，$Q_N(u) = \int_u^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$ の計算は，以下の式で近似できる。

$$Q_N(u) = (1 + d_1 u + d_2 u^2 + d_3 u^3 + d_4 u^4 + d_5 u^5 + d_6 u^6)^{-16}/2 \tag{15}$$

[*7] 注：以下で述べる計算方法は，有効数字三桁を目標にして作られたものである。

このとき，$u > 0$ であることに注意すること。$u < 0$ の場合は，$Q(u) = 1 - Q(|u|)$ で求まる。d_1〜d_6 はそれぞれ，

$$d_1 = 0.0498673470, d_2 = 0.0211410061$$

$$d_3 = 0.0032776263, d_4 = 0.0000380036$$

$$d_5 = 0.0000488906, d_6 = 0.0000053830$$

である。

例 3 (偏差値 63 はどの当たりのポジションか?)**.** z 得点で言うと，$(63 - 50)/10 = 1.3$ なので，

$$Q(1.3) = (1 + 0.049863470 \times 1.3 + 0.0211410061 \times 1.69 + \cdots + 0.0000053830 \times 4.826809)^{-16}/2$$

より，9.680 が求まる[*8]。まだ自分の上に *9.68%* いるわけです。

4.2 正規分布のパーセント点

今度は逆に，上から α% の面積を占める点を調べたいとしよう。上側確率 $\alpha(0 < \alpha < 0.5)$ から標準正規分布 $N(0, 1)$ の上側 100α パーセント点，u_α は

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

ということである。これは以下の式で求める。

$$u_\alpha = \{y(b_0 + b_1 y + b_2 y^2 + \cdots + b_8 y^8\}^{1/2} \tag{16}$$

このとき，

$$y = -\log_e\{4\alpha(1 - \alpha)\}$$

また，b_n はそれぞれ，

$$b_0 = 0.1570796288 \times 10, b_1 = 0.3706987906 \times 10^{-1}$$

$$b_2 = -0.8364353589 \times 10^{-3}. b_3 = -0.2250947176 \times 10^{-3}$$

$$b_4 = 0.6841218299 \times 10^{-5}, b_5 = 0.5824238515 \times 10^{-5}$$

$$b_6 = -0.1045274970 \times 10^{-5}. b_7 = 0.8360937017 \times 10^{-7}$$

$$b_8 = -0.3231081277 \times 10^{-8}$$

である。$0.5 < \alpha < 1$ のときは，$u_\alpha = -u_{1-\alpha}$ で求まる。

例 4 (正規分布で上から 10% のパーセント点は?)**.** $y = 1.021651247532...$ だから，表 *3* より，$u_\alpha = 1.64237439950817^{0.5} = 1.281...$ である。偏差値 *62.8* を越えれば，上位 *10%* に入ったことになる。

以下の分布はいずれも，この標準正規分布の近似を元に求められるといっても過言ではない。

[*8] −16 乗は Excel で ⁻16 と入れれば出てくる。16 乗根を求める！と意気込まなくても良い。便利。

表 3　y の累乗と b_n，およびその積

y	1.021651247532	b1	0.037069879060000000	0.037872488187508700
y^2	1.043771271584	b2	-0.000836435358900000	-0.000873047198156583
y^3	1.066370221751	b3	-0.000225094717600000	-0.000240034303922199
y^4	1.089458467383	b4	0.000006841218299000	0.000007453223203064
y^5	1.113046602336	b5	0.000005824238515000	0.000006482648890318
y^6	1.137145449838	b6	-0.000001045274970000	-0.000001188629675965
y^7	1.161766067453	b7	0.000000083609370170	0.000000097134529185
y^8	1.186919752153	b8	-0.000000003231081277	-0.000000003835034188

4.3　χ^2 分布の場合

4.3.1　上側確率

自由度 v の χ^2 分布の上側確率は，以下の式で定義される。

$$Q_{\chi^2}(\chi^2, v) = \int_{\chi^2}^{\infty} \frac{1}{\frac{v}{2}\Gamma(\frac{v}{2})} x^{v/2-1} e^{-x/2} dx$$

これの近似値を求めるとき，自由度が偶数か奇数かによって算出方法が異なる。

まず偶数の時。

$$Q_{\chi^2}(\chi^2, v) = e^{-\chi^2/2} \left\{ 1 + \frac{\chi^2}{2} + \frac{\chi^4}{2 \cdot 4} + \cdots + \frac{\chi^{v-2}}{2 \cdot 4 \cdots (v-2)} \right\} \tag{17}$$

ついで奇数の時。

$$2Q_N(\chi) + \sqrt{\frac{2}{\pi}} e^{-\chi^2/2} \left\{ \frac{\chi}{1} + \frac{\chi^3}{1 \cdot 3} + \cdots + \frac{\chi^{v-2}}{1 \cdot 3 \cdot 5 \cdots (v+2)} \right\} \tag{18}$$

しかし，自由度 v が十分に大きければ (v40)，以下の近似式で求められる。

$$Q_{\chi^2}(\chi^2, v) = Q_N \left(\frac{\left(\frac{\chi^2}{v}\right)^{1/3} - \left(1 - \frac{2}{9v}\right)}{\sqrt{2/9v}} \right) \tag{19}$$

このとき，$Q_N(u)$ は標準正規分布の上側確率である。

4.3.2　パーセント点

上側確率 α から，自由度 v の χ^2 分布のパーセント点を求めることを考える。このとき，

$$\int_{\chi^2_\alpha(v)}^{\infty} \frac{1}{2^{v/2}\Gamma(\frac{v}{2})} x^{v/2-1} e^{-x/2} dx = \alpha$$

を求めるために，自由度毎に異なる計算をする。

まず，自由度が 1 のとき。

$$\chi^2_\alpha(1) = (u_{\alpha/2})^2$$

ここでの $u_{\alpha/2}$ は標準正規分布の $100\frac{\alpha}{2}$ パーセント点。

次に，自由度が 2 のとき。

$$\chi^2_\alpha(2) = -2\log_e \alpha$$

最後に，自由度が 3 以上のとき。このときは反復法による計算になる。

1. まず，$z_L = 0, z_U = 1, z_0 = 0.5$ と置く。
2. $C = \frac{1}{z_0} - 1$ を求める。
3. この C を先ほどの上側確率計算に代入し，その値 $\alpha' = Q_{\chi^2}$ を得る。
4. α' が今回求めたい α より小さい，すなわち $\alpha' < \alpha$ であれば，z_L に z_0 の値を代入する。
5. 逆に $\alpha' > \alpha$ であれば，z_U に z_0 の値を代入する。
6. $z_1 = (z_L + Z_U)/2$ を計算する。
7. z_1 と z_0 の差が十分に小さければ，Q_{χ^2} の値が求める値である。
8. z_1 と z_0 の差が無視できないほどであれば，z_0 に z_1 の値を代入し，2 に戻る。

これはコンピュータによる反復計算になれていない人にはわかりにくいだろうから，実際に計算をしてみよう。

例 5 (自由度 7 で上位 20% のパーセント点を求めたい)**.** 収束の目安を $1e-10 = 1/10^{-10}$ ぐらいに設定しておこう。これは「十分小さい」と考えて良い。ちょっと厳しすぎるぐらいだ。

1. 初期値は $z_L^{(0)} = 0$，$z_U^{(0)} = 1$，$z_0^{(0)} = 0.5$ である。ここから，最初の $C = 1$ の上側確率，$\alpha' = 0.994829$ が得られる。
2. これは求める $\alpha = 0.2$ よりも大きいので，z_L はそのまま，$z_U^{(1)} = 0.5$ とする。
3. $z_1 = (0.0 + 0.5)/2 = 0.25$ である。z_0 と z_1 の差が 0.25 である。これは収束基準，$1e-10$ より大きいので，もう一回反復しなければならない。
4. 今回は $z_L^{(1)} = 0.0$，$z_U^{(1)} = 0.5$，$z_0^{(1)} = 0.25$ である。ここから $C = 3$ の時の上側確率，$\alpha = 0.885002$ が得られる。
5. これはまだ $\alpha = 0.2$ よりも大きいので，$z_U = 0.25$ とする。
6. $z_1 = (0.0 + 0.25)/2 = 0.125$ で，z_0 との差がまだ収束基準より大きい。もう一度反復である。
7. 今回は $z_L^{(2)} = 0.0$，$z_U^{(2)} = 0.25$，$z_0^{(2)} = 0.125$ である。ここから $C = 7$ の時の上側確率，$\alpha = 0.428800$ が得られる。
8. これもまだ大きい。$z_U^{(3)} = 0.125$ にする。さらに $z_1^{(3)} = 0.0625$ である。

このような計算を繰り返す。これはもちろん，エクセルなどを用いるよりも，*C* 言語や *Fortran* といったプログラム言語を使って計算させるべきである。実際に計算させると，この例だと *30* 回ぐらいの反復で収束する。計算プロセスを表 *4* に示しておく。

4.4 F 分布の近似

4.4.1 上側確率

自由度 (v_1, v_2) の F 分布の上側確率は以下の式で求められる。

$$Q_F(F, v_1, v_2) = \int_F^\infty \frac{1}{B\left(\frac{v_1}{2}, \frac{v_2}{2}\right)} \left(\frac{v_1}{v_2}\right)^{v_1/2} x^{v_1/2-1} \left(1 + \frac{v_1}{v_2} x\right)^{-(v_1+v_2)/2} dx$$

表 4　反復計算による近似

反復回数	C	α'	z_L	z_U	z_1
1	1	0.994829	0	0.5	0.25
2	3	0.885002	0	0.25	0.125
3	7	0.42888	0	0.125	0.0625
4	15	0.0359997	0.0625	0.125	0.09375
5	9.66667	0.208263	0.0625	0.09375	0.078125
6	11.8	0.107331	0.078125	0.09375	0.0859375
7	10.6364	0.155286	0.0859375	0.09375	0.0898438
8	10.1304	0.181302	0.0898438	0.09375	0.0917969
9	9.89362	0.194683	0.0917969	0.09375	0.0927734
10	9.77895	0.20145	0.0917969	0.0927734	0.0922852
11	9.83598	0.19806	0.0922852	0.0927734	0.0925293
12	9.80739	0.199754	0.0925293	0.0927734	0.0926514
13	9.79315	0.200602	0.0925293	0.0926514	0.0925903
14	9.80026	0.200178	0.0925293	0.0925903	0.0925598
15	9.80382	0.199966	0.0925598	0.0925903	0.0925751
16	9.80204	0.200072	0.0925598	0.0925751	0.0925674
17	9.80293	0.200019	0.0925598	0.0925674	0.0925636
18	9.80338	0.199992	0.0925636	0.0925674	0.0925655
19	9.80316	0.200005	0.0925636	0.0925655	0.0925646
20	9.80327	0.199999	0.0925646	0.0925655	0.0925651
21	9.80321	0.200002	0.0925646	0.0925651	0.0925648
22	9.80324	0.200001	0.0925646	0.0925648	0.0925647
23	9.80325	0.2	0.0925647	0.0925648	0.0925648
24	9.80325	0.2	0.0925647	0.0925648	0.0925647
25	9.80325	0.2	0.0925647	0.0925648	0.0925647
26	9.80325	0.2	0.0925647	0.0925648	0.0925648
27	9.80325	0.2	0.0925647	0.0925648	0.0925648
28	9.80325	0.2	0.0925647	0.0925648	0.0925647
29	9.80325	0.2	0.0925647	0.0925648	0.0925647
30	9.80325	0.2	0.0925647	0.0925647	0.0925647
31	9.80325	0.2	0.0925647	0.0925647	0.0925647
32	9.80325	0.2	0.0925647	0.0925647	0.0925647
33	9.80325	0.2	0.0925647	0.0925647	0.0925647

これは「不完全ベータ関数比」なるもので，以下のように表される。

$$Q_F(F, v_1, v_2) = 1 - I_x\left(\frac{v_1}{2}, \frac{v_2}{2}\right) \tag{20}$$

このとき，

$$x = \frac{v_1 F}{v_2 + v_1 F}$$

$$I_x(a, b) = \frac{1}{B(a, b)} \int_0^x t^{\alpha-1}(1-t)^{b-1} dt$$

$$B(a, b) = \frac{\Gamma(a)\dot{\Gamma}(b)}{\Gamma(a+b)}$$

さらに，不完全ベータ関数比は以下の漸化式で求められる。

$$\begin{cases} I_x\left(\frac{v_1}{2}+1, \frac{v_2}{2}\right) &= I_x\left(\frac{v_1}{2}, \frac{v_2}{2}\right) - \frac{2}{v_1} U\left(\frac{v_1}{2}, \frac{v_2}{2}\right) \\ U\left(\frac{v_1}{2}+1, \frac{v_2}{2}\right) &= \frac{v_1+v_2}{v_1} x U\left(\frac{v_1}{2}, \frac{v_2}{2}\right) \end{cases}$$

$$\begin{cases} I_x\left(\frac{v_1}{2}, \frac{v_2}{2}+1\right) &= I_x\left(\frac{v_1}{2}, \frac{v_2}{2}\right) + \frac{2}{v_2} U\left(\frac{v_1}{2}, \frac{v_2}{2}\right) \\ U\left(\frac{v_1}{2}, \frac{v_2}{2}+1\right) &= \frac{v_1+v_2}{v_2}(1-x) U\left(\frac{v_1}{2}, \frac{v_2}{2}\right) \end{cases}$$

なんだか奇妙な式に見えると思うが，上の一組は v_1 を近似していくときに，下の一組は v_2 を近似していくときに使う。また，I_x を求めるときはまず U が求まっていないといけないので，このような式になっているのである。

F 分布の近似は，まず自由度が偶数か奇数かによって，初期値を決定する。偶数なら 1 からスタートするのでいいのだが，奇数の場合は 1/2 からスタートしないと，+1 ずつ増えていくので求める値に合致しないからである。

v_1, v_2 の両方が偶数であった場合，初期値は

$$I_x(1, 1) = x$$

$$U(1, 1) = x(1-x)$$

で与えられる。v_1 が偶数で，v_2 が奇数の場合，

$$I_x\left(1, \frac{1}{2}\right) = 1 - \sqrt{1-x}$$

$$U\left(1, \frac{1}{2}\right) = \frac{x}{2}\sqrt{1-x}$$

で与えられる。v_1 が奇数，v_2 が偶数であれば，

$$I_x\left(\frac{1}{2}, 1\right) = \sqrt{x}$$

$$U\left(\frac{1}{2}, 1\right) = \frac{1}{2}\sqrt{x}(1-x)$$

最後にどちらも奇数であれば，

$$I_x\left(\frac{1}{2},\frac{1}{2}\right) = 1 - \frac{2}{\pi}\tan^{-1}\sqrt{\frac{1-x}{x}}$$

$$U\left(\frac{1}{2},\frac{1}{2}\right) = \frac{1}{\pi}\sqrt{x(1-x)}$$

である。

例 6 ($F(3,5) = 7.76$ の上側確率を求めたい)**.** 自由度がどちらも奇数であるから，初期値は

$$x = \frac{v_1 F}{v_2 + v_1 F} = \frac{7.76\times 3}{7.76\times 3 + 5} = 0.9512195$$

$$I_x^0 = 1 - \frac{2}{\pi}\tan^{-1}\sqrt{\frac{(1-x)}{x}} = 0.818929$$

$$U^0 = \frac{1}{\pi}\sqrt{x(1-x)} = 0.857313$$

である。ここから徐々に答えに近づけていく。

まず初期値を入れて得られる値は，

$$I_x^1\left(\frac{3}{2},\frac{1}{2}\right) = I_x^0 - \frac{2}{1}U^0 = 0.480852$$

$$U^1\left(\frac{3}{2},\frac{1}{2}\right) = U^0 x\frac{1+v_2}{1} = 0.199931$$

これで v_1 が 3 に達したので，v_2 を 5 まで増やしていく。

$$I_x^2\left(\frac{3}{2},\frac{3}{2}\right) = I_x^1 + \frac{2}{3}U^1 = 0.880714$$

$$U^2\left(\frac{3}{2},\frac{3}{2}\right) = \frac{3+3}{3}(1-x)U^1 = 0.141394$$

$$I_x^3\left(\frac{3}{2},\frac{5}{2}\right) = I_x^2 + \frac{2}{3}U^2 = 0.974977$$

$$U^3\left(\frac{3}{2},\frac{5}{2}\right) = \frac{3+5}{3}(1-x)U^2 = 0.049979$$

こうして，$I_x = 0.974977$ を得る。求める値は，$1.0 - I_x = 0.025023$ である。

もう少し簡単な近似計算式として，次のようなものもある。

$$Q_F(F, v_1, v_2) = Q_N\left(\frac{f_1}{f_2}\right)$$

ここで $Q_N(x)$ は標準正規分布の上側確率関数で，f_1 と f_2 はそれぞれ，

$$f_1 = \left(1 - \frac{k}{v_2}\right)F^{1/3} - \left(1 - \frac{k}{v_1}\right)$$

$$f_2 = \sqrt{\frac{k}{v_1} + \frac{kF^{2/3}}{v_2}}$$

このとき，$k = 2/9$ の定数である。

例 7 ($F(3,5) = 7.76$ の上側確率を求めたい 2)**.**

$$f_1 = \left(1 - \frac{0.222222}{5}\right) 7.76^{1/3} - \left(1 - \frac{0.22222}{3}\right) = 0.965882$$

$$f_2 = \sqrt{\frac{0.22222}{3} + \frac{0.22222 \times 7.76^{2/3}}{5}} = 0.498275$$

$$Q_N(0.965882/0.498275) = 0.0262843$$

先ほど求めた答え，0.025023 とは少し異なるが，有効桁数は 3 桁くらいなので，0.025 と 0.026 だから許される程度の誤差であるだろう。

4.4.2 パーセント点

これは χ^2 分布の時と同様，反復による近似計算になる。自由度 (v_1, v_2) の F 分布の 100α パーセント点 $F_\alpha(v_1, v_2)$ を求める式は，

$$\int_{F_\alpha(v_1,v_2)}^{\infty} \frac{1}{B\left(\frac{v_1}{2}, \frac{v_2}{2}\right)} \left(\frac{v_1}{v_2}\right)^{v_1/2} x^{v_1/2-1} \left(1 + \frac{v_1}{v_2} x\right)^{-(v_1+v_2)/2} dx = \alpha$$

である。これを求めるにはまず，不完全ベータ関数比 x の初期値として，

$$x_L = 0.0, x_U = 1.0, x_0 = 1/2$$

とし，$I_{x_0}\left(\frac{v_1}{2}, \frac{v_2}{2}\right) = \alpha'$ を求める。得られた α' と求めたい α の関係を見て，

$$\begin{cases} \alpha' < \alpha \text{のとき} & x_L \leftarrow x_0 \\ \alpha' > \alpha \text{のとき} & x_U \leftarrow x_0 \\ x_0 \leftarrow (x_L + x_U)/2 & \end{cases}$$

として更新を繰り返す。反復しても値が変化しないと収束したと見なし，その時の x_0 より，

$$F_\alpha(v_1, v_2) = \left(\frac{1}{x_0} - 1\right) \frac{v_1}{v_2}$$

とする。しかし，自由度 v_1, v_2 が共に大きい場合 $(v_i \geq 31)$ は，以下の近似式で得られる。

$$F_\alpha(v_1, v_2) = \left\{ \frac{(1-a)(1-b) + u_\alpha \sqrt{(1-a)^2 b + (1-b)^2 a - a b u_\alpha^2}}{(1-b)^2 - b u_\alpha^2} \right\}^3$$

このとき，$a = 2/9v_1, b = 2/9v_2$，また u_α は標準正規分布の上側 100α パーセント点である。

例 8 ($F(6,6)$ の 10 パーセント点を求める)**.** 収束の目安は $1e-10 = 1/10^{-10}$ ぐらいでよい。

1. 初期値は $x_L = 0.0, x_U = 1.0, x_0 = 0.5$ である。
2. $(1/0.5 - 1) \times (6/6) = 1.0$ より，$F(6,6)$ の 1.0 より上側の確率は 0.5，を得る。
3. $0.5 > 0.1$ なので，$x_U = 0.5$ とし，また $x_0 = 0.25$ としてもう一度計算。
4. 以下同様

計算プロセスを表 *5* に示す。

表 5　F 分布のパーセント点を求める反復計算

反復回数	F	α	x_L	x_U	x_0
1	1.00	0.5	0	0.5	0.25
2	3.00	0.103673	0	0.25	0.125
3	7.00	0.017161	0.125	0.25	0.1875
4	4.33333	0.049567	0.1875	0.25	0.21875
5	3.57143	0.073823	0.21875	0.25	0.234375
6	3.26667	0.08805	0.234375	0.25	0.242188
7	3.12903	0.095688	0.242188	0.25	0.246094
8	3.06349	0.099637	0.246094	0.25	0.248047
9	3.0315	0.101645	0.246094	0.248047	0.24707
10	3.04743	0.100638	0.246094	0.24707	0.246582
11	3.05545	0.100137	0.246094	0.246582	0.246338
12	3.05946	0.099887	0.246338	0.246582	0.24646
13	3.05745	0.100012	0.246338	0.24646	0.246399
14	3.05846	0.099949	0.246399	0.24646	0.246429
15	3.05796	0.099981	0.246429	0.24646	0.246445
16	3.05771	0.099996	0.246445	0.24646	0.246452
17	3.05758	0.100004	0.246445	0.246452	0.246449
18	3.05764	0.1	0.246445	0.246449	0.246447
19	3.05767	0.099998	0.246447	0.246449	0.246448
20	3.05766	0.099999	0.246448	0.246449	0.246448
21	3.05765	0.1	0.246448	0.246449	0.246448
22	3.05765	0.1	0.246448	0.246449	0.246448
⋮	⋮	⋮	⋮	⋮	⋮
32	3.05765	0.1	0.246448	0.246448	0.246448
33	3.05765	0.1	0.246448	0.246448	0.246448

4.5　t 分布の近似

4.5.1　上側確率

自由度 v の t 分布の上側確率は，以下の式で与えられる。

$$Q_t(v) = \int_t^\infty \frac{1}{\sqrt{v}B\left(\frac{v}{2}.\frac{1}{2}\right)}\left(1+\frac{x^2}{v}\right)^{-(v+1)/2} dx$$

これは F 分布の上側確立との関係を用いて，

$$Q_t(v) = \frac{1}{2}Q_F(t^2, 1, v)$$

で求められる。ただし，これは $t > 0$ の時で，$t < 0$ の場合は 1.0 から得られた値を引くこと。

Welch の検定などで，自由度に少数がついている場合は，その少数を挟む整数自由度で求めた二つの上側確率から，直線補間で求められる。自由度 ν が実数の時，

$$Q_t(\nu) = Q_t([\nu]) + (Q_t([\nu]+1) - Q_t([\nu]))(\nu - [\nu])$$

とする。ここで，$[\nu]$ は Gauss 記号と呼ばれるもので，$[x]$ とは 実数 x から整数への対応 (正確には写像 mapping) で x を超えない最大の整数を表す。即ち $x-1 < [x]x$ を満たす整数である。$x > 0$ の時は小数部分の切り捨てを意味するが，$x < 0$ の時は，例えば $[-2.5] = -3$ となることに注意。

4.5.2 パーセント点

上側確率 α から，自由度 ν の t 分布上側 100α パーセント点 $t_\alpha(\nu)$ は，

$$\int_{t_\alpha(\nu)}^{\infty} \frac{1}{\sqrt{\nu}B\left(\frac{\nu}{2},\frac{1}{2}\right)}\left(1+\frac{x^2}{\nu}\right)^{-(\nu+1)/2} ds$$

で求められる。近似式は，

$$t_\alpha(\nu) = u_\alpha + \frac{Y_1(u_\alpha)}{\nu} + \frac{Y_2(u_\alpha)}{\nu^2} + \cdots + \frac{Y_5(u_\alpha)}{\nu^5}$$

である。この u_α は標準正規分布の上側 100α パーセント点，また，

$$Y_1(u) = \frac{1}{2^2}(u^3+u)$$

$$Y_2(u) = \frac{1}{2^5\dot{3}}(5u^5+16u^3+3u)$$

$$Y_3(u) = \frac{1}{2^7\dot{3}}(3u^7+19u^5+17u^3-15u)$$

$$Y_4(u) = \frac{1}{2^{11}\dot{3}^2\dot{5}}(79u^9+776u^7+1482u^5-1920u^3-945u)$$

$$T_5(u) = \frac{1}{2^{13}\dot{3}^2\dot{5}}(27u^{11}+339u^9+930u^7-1782u^5-765u^3+17955u)$$

である。ただ，この近似法は $\nu < 5$，または $\alpha < 0.05$ の場合はうまく近似できないので，それ以上の値の時に用いられたし。