第 2 章

統計的な考え方の基礎〜確率と確率分布

2.1　本章のテーマ

前章で，統計学とは「不確実性を考慮した論理的推論である」と述べた。不確実性とは，絶対にないとも絶対にあるとも決まっていない，ファジーな確率をもっているということである。

しかし，ここで簡単に「確率」と言ってしまったが，さて，改めて確率とは何かと訊かれたら，答えに詰まってしまうのではないだろうか。

そこで，本章では，あらゆる統計的な考え方の基礎となる「確率」というものを徹底的に考えてみることにする。高校数学の復習になってしまうかもしれないが，頭を整理しておくという意味で，役に立つのではないだろうか。このテーマについて，もっと詳しく知りたい方は，伏見 (1987) など，確率論の本を参照されたい。

2.2　確率的な現象を統計的事象と呼ぶ

2.2.1　どういう現象が確率的か？

- サイコロを振ったときの目：振ってみるまでは 1 から 6 のどれが出るかはわからない。どの目がでる可能性も等しいから。
- 天気予報：「明日の天気予報は晴れ」といっても「必ず晴れる」とは限らない。「曇ったり雨が降ったりする可能性も少しはあるが，晴れる可能性が高い」ことを意味する。
- 喫煙と肺がんの関係：「タバコを吸うと肺がんになる」という命題は，「タバコ

を吸った人と吸わなかった人を比べて，肺がんになった人の割合が吸った人の方で高い」という関係を示す。タバコを吸っても肺がんにならない人もいるし，吸わなくても肺がんになる人もいる。

こういう「不確かさ」に潜む法則性（長期間繰り返し観察したり，大集団で観察すると見られる）を考える学問を確率論と呼ぶ。大雑把に言えば，この種の法則性をもつ現象を，「統計的事象」と呼び，統計的事象の確かさの度合いを示すのに便利なモノサシが「確率」である。そこで，「確率」をきちんと定義してみることにする。その前に，いくつかの準備が必要である。

2.3 準備その１：「標本空間」

統計的事象を捉えるには，「どんなことが起こりうるか」という範囲を定めることが必要である。

現象は一般に多面的で，様々な観察方法がある。以下３点によって統計的現象を捉えた，記号化された結果の集合のことを「標本空間」と呼ぶ。

- 観察を行う側面を特定する
- 起こりうる結果の範囲を規定する
- その範囲内の各結果に記号を対応させる

個々の結果の起こりうる可能性を示す数値（これを「確率」と呼ぶ）を考える。もっとも単純には「どの結果も同程度に起こる」と考える。各結果に対応付けられた確率は０から１までの数値であり，各確率の値の総和は１にならねばならない。たとえば，サイコロの目では，標本空間は｛1，２，３，４，５，６｝である。

2.4 準備その２：「事象」

問題は，個々の結果の可能性よりも，いくつかの結果が複合された集合（これを「事象」と呼ぶ）の起こる可能性がどのくらいか，ということである。つまり，「事象」＝「標本空間の部分集合」である。

サイコロの例では，「目が偶数（丁半博打でいえば丁）」とか「目が５以上」とか「目が１」とかいうことが事象といえる。

ある事象の確率は，その事象に含まれる各結果の生起確率の和である。したがって，各結果の生起確率が等しい場合は，その事象に含まれる結果の場合の数をすべての場合の数で割ると，その事象の確率になる。サイコロの例では，「目が５以上」という事象の確率は，2/6=0.333... である。

2.5 準備その３：余事象・和事象・積事象・排反事象

起こりうるすべての結果の集合を「全事象」という。つまり，全事象は標本空間に等しい。

決して起こらない事象を「空事象」といい，空集合 ϕ で表す。

事象 E に対して，E が起こらないという事象を E の「余事象」という。E の余事象を $\bar{E}$ と書く。サイコロの例では，「目が偶数」という事象の余事象は「目が奇数」である。事象 E が起こる確率を $Pr(E)$、余事象の起こる確率を $Pr(\bar{E})$ という記号で書くと $Pr(E)+Pr(\bar{E})=1$ が常に成立する。

事象 E と F の少なくとも一方が起こるという事象を，E と F の「和事象」といい，$E\cup F$ で表す。

事象 E と F の両方が起こるという事象を，E と F の「積事象」といい，$E\cap F$ で表す。

事象 E が起これば F は決して起こらないとき，E と F は「排反事象」であるという。E と F が排反事象なら，$E\cap F=\phi$ である。

2.6 準備その４：相互排反性と加法定理

サイコロで考えると，１回振ったとき「偶数の目が出る」という事象 E が起こる確率は，2,4,6 の場合の数３を，1,2,3,4,5,6 の場合の数６で割った値なので $Pr(E)=0.5$ である。

では，２回振って「少なくとも１回は偶数の目」の確率はどうなるだろうか？ まず，0.5+0.5=1.0 ではないのは自明である（1.0 ということは，必ずそうなるということだから）。ここで大切なのは，『１回目に「偶数の目が出る」事象 E_1 と２回目に「偶数の目が出る」事象 E_2 とは排反ではない』ことに注意することである。集合のベン図（図 2.1）から考えると，$Pr(E_1\cup E_2)=Pr(E_1)+Pr(E_2)-Pr(E_1\cap E_2)$ であることが直感的にわかる。この式を「加法法則」と呼ぶ。ベン図をよく見ると，「２回とも奇数」の余事象なので，$1-Pr(\bar{E}_1\cap\bar{E}_2)=1-9/36=0.75$ と考えてよいこともわかるだろう。因みに，事象 E と事象 F が排反なら，$Pr(E_1\cap E_2)=0$ なので，$Pr(E\cup F)=Pr(E)+Pr(F)$ という「加法定理」が成立する。

2.7 準備その５：事象の独立性と乗法定理

事象 E が起こったときに事象 F が起こる確率を，「E が起こったときの F の条件付き確率」といい，$Pr(F|E)$ と書く。

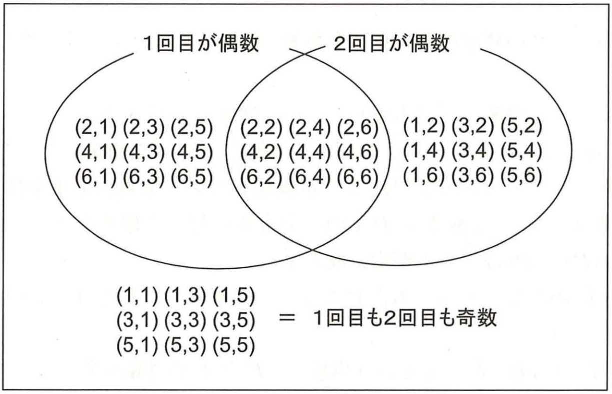

図 **2.1**　サイコロを 2 回振って「少なくとも 1 回は偶数の目」の確率を考えるためのベン図

「E が起こった」うちの「E も F も起こった」場合なので，$Pr(F|E) = Pr(F \cap E)/Pr(E)$ である。

E と F が互いに無関係（＝独立）なら，$Pr(F|E) = Pr(F)$。逆にいえば，$Pr(F) = Pr(F|E)$ のときに事象 E と事象 F は互いに独立であるという。独立でないとき「従属である」という。

上記 2 つの式から，E と F が独立なら，$Pr(F \cap E) = Pr(F) \times Pr(E)$ という「乗法定理」が成立する。

2.8　確率を定義するための 4 種類のアプローチ

確率を定義するには，以下の 4 種類のアプローチがある。

- 操作的アプローチ（統計的定義）：数多く試したときの相対度数の極限。たとえば，事象 E が起こる確率 $Pr(E)$ は，N 回試したときに N_1 回事象 E が起こるとして，N_1/N という相対度数が，N を無限大にしたときに漸近する値である。
- 対称的確率：サイコロの場合，6 通りの目の出る確率はどれも等しくなければならず，その和は 1 でなくてはならないので，たとえば 1 の目が出る確率は

1/6 となる。限定的かつ循環論法。

- 公理的客観確率：標本空間の各要素を e_i として，$Pr(e_i) \geq 0$ かつ $Pr(e_1) + \mathrm{Pr}(e_2) + \ldots + Pr(e_N) = 1$ かつ事象 E が起こる確率 $Pr(E) = \sum \mathrm{Pr}(e_i)$ を公理とする[*1]。
- 主観確率：観念的にも二度と繰り返すことのできない事象についての「見込み」を扱う。決定理論において重要。

公理的客観確率

より厳密に定義するならば，次のようになる（伏見，1987より）。
確率の議論をする際に考える事象群 $\mathfrak{A}$ は，次の条件を満たしていなければならない。

【B1】標本空間 Ω が $\mathfrak{A}$ に含まれている。

【B2】事象 A が $\mathfrak{A}$ に含まれているならば，A の余事象 $\bar{A}$ も $\mathfrak{A}$ に含まれている。

【B3】$A_1, A_2, \ldots$ が $\mathfrak{A}$ に含まれているならば，それらの和事象 $A_1 \cup A_2 \cup \ldots \left(= \bigcup_{i=1}^{\infty} A_i\right)$ も $\mathfrak{A}$ に含まれている。

これら3つの条件をすべて満たす $\mathfrak{A}$ のことをボレル集合体という。このとき，以下の条件も自然に成立する。

【B4】空集合 ϕ が $\mathfrak{A}$ に含まれている。（ϕ は標本空間の余事象なので，【B1】と【B2】より自明である）

【B5】$A_1, A_2, \ldots$ が $\mathfrak{A}$ に含まれているならば，それらの積事象 $A_1 \cap A_2 \cap \ldots \left(= \bigcap_{i=1}^{\infty} A_i\right)$ も $\mathfrak{A}$ に含まれている。（集合論におけるド・モルガンの法則〔＝積事象の余事象は余事象の和事象に等しい〕と【B2】と【B3】より自明である）

こうしてボレル集合体 $\mathfrak{A}$ を定めた上で，$\mathfrak{A}$ に含まれる個々の事象が起こる確率（確率測度ということもある）を定義することができる。
事象 A の起こる確率を $Pr(A)$ という記号で書くと，確率 $Pr(\cdot)$ は，次の性質をもつものとして定義できる（これが公理的客観確率の厳密な定義である）。

【P1】任意の事象 A の確率は 0 と 1 の間の実数である。

【P2】標本空間全体 Ω の確率は 1 である。

【P3】$A_1, A_2, \ldots$ が互いに排反な事象であるならば，それらの和事象の確率は，それらの確率の和に等しい（「完全加法性」と呼ばれる）。

まず標本空間を考え，その部分集合の集まりの一種としてボレル集合体を考え，最後にボレル集合体の要素に実数値を対応させる関数で性質【P1】～【P3】を満たすものとして確率を定めたので，これら3つを組にして $(\Omega, \mathfrak{A}, P)$ と書き，これを確率空間と呼ぶ。

[*1] 要素は互いに排反であるということ。当然である。

2.9　大数の法則（操作的接近の根拠）

操作的アプローチが成り立つ様子を，コンピュータを使って調べてみよう。R のプログラムでは次のようになる[*2]。

```
a <- c(100,1000,10000,10000)
op <- par(mfrow=c(2,2))
for (i in 1:4){
y <- as.integer(runif(a[i],1,7))
for (j in c(1:length(y))){if (y[j]>6) y[j]<-6}
s <- paste("n=", as.integer(a[i]))
barplot(table(y),main=s)}
par(op)
```

```
a <- c(100,1000,10000,100000)
op <- par(mfrow=c(2,2))
for (i in 1:4) {
y <- as.integer(runif(a[i],1,7))
for (j in c(1:length(y))) { if (y[j]>6) y[j]<-6 }
s <- paste("n=",as.integer(a[i]))
barplot(table(y),main=s) }
par(op)
```

このプログラムで試行回数を増やすと，図 2.2 のように，サイコロの特定の目が出る割合が，ある一定値に近づくことがわかる。「1 の目が出る」事象 E_1 が起こる確率 $Pr(E_1) = p$ とおけば，N 回サイコロを振って N_1 回 1 の目が出たとして，任意の正の小さな数 ε に対して，$\lim_{N\to\infty} Pr(|N_1/N - p| < \varepsilon) = 1$ ということで，これをベルヌーイ (Bernoulli) の大数の法則という。

2.10　確率変数・期待値・分散の感覚的把握

確率変数と期待値について，まず感覚的に把握することは重要である。そこで，鈴木 (1995) に掲載されている，スロットマシンの例を紹介しよう。

スロットマシンでは，ごくたまに，投入金額の何十倍ものコインが出てくることがある。マシン利用者全員に返ってくる賞金の合計を利用回数で割った値が，1 回に期待される賞金額で，これを賭け金で割った値を「賞金還元率」と呼ぶ。言い換えると，1 から賞金還元率を引いた値が，賭け事の胴元が儲けると期待される値である。一般に，賞金額が $x_1, x_2, x_3, \ldots$ で，その賞金が得られる確率が $p_1, p_2, p_3, \ldots$ のように設定されたスロットマシンの期待賞金額 M は，$M = x_1p_1 + x_2p_2 + x_3p_3 + \ldots$ で与え

[*2] `runif(N,min,max)` は区間 [min,max] の N 個の一様乱数を発生させる関数。min と max が省略されたときは [0,1] と仮定される。[1,7] を as.integer で切り捨てると，ちょうど 7 の場合はサイコロにない目が出ることになってしまうので，ちょうど 7 の場合のみ強制的に 6 にしている。また，テキストファイルとしてプログラム全体を予め用意し，`source()` で読み込んで実行させる場合は問題ないが，R のコンソールで 1 行ずつ入力していると，2 行目の `par()` を入力した後で，フォーカスがグラフィックウィンドウに移ってしまうので，コンソールウィンドウをクリックして入力可能な状態に戻す必要がある。付録（167 ページ）に Windows で実行した様子を示した。

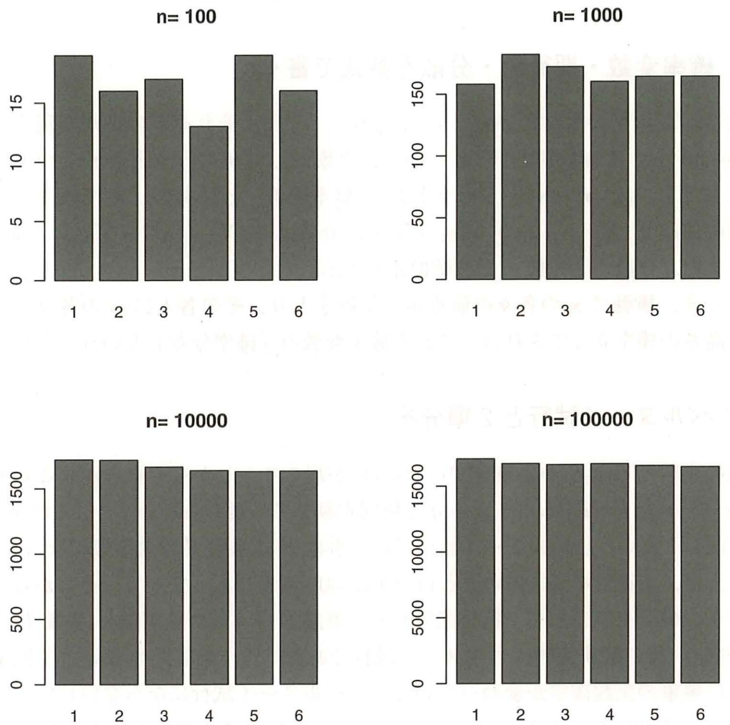

図 **2.2** 大数の法則のシミュレーション

られる。このスロットマシンのようなものを確率変数といい，期待賞金を期待値と呼ぶ（厳密には後述）。

期待賞金が同じでも，値動きの幅が小さいと一喜一憂の程度が小さく，逆に幅が大きいと滅多に当たらないが当たったときの喜びは大きくなる。つまり，ギャンブル性は，値動きの幅と，チャンスの大きさに依存している。

各賞金がどれくらい期待賞金から隔たりがあり，それを獲得できる可能性がどれくらいあるのかを見積もれば，ギャンブル性が表せる。

マシンのギャンブル性を V とおけば，$V = \sum$（期待値からの隔たり）$\times$（可能性）という値が定義できる。この V を「分散」と呼ぶ。このとき，各賞金額 x と期待値

M の隔たりは，差を二乗した値 $D = (x - M)^2$ で表す。

2.11　確率変数・期待値・分散を数式で書く

一般に，とりうる値の集合 $\mathbf{x} = (x_1, x_2, x_3, \ldots)$ と，それぞれの値が実現する確率 $\mathbf{p} = (p_1, p_2, p_3,\ dots)$ が与えられていて，事象として $\mathbf{x}$ のうちどれか 1 つの値のみ実現するとき，$(\mathbf{x}, \mathbf{p})$ という 1 セットを，「確率変数」と呼んで，X で表す。このとき，期待値は $E(X) = \mu = \sum x_i p_i$ であり，分散は $V(X) = \sigma^2 = \sum (x_i - \mu)^2 p_i$ となる。また，分散の平方根 σ を標準偏差と呼ぶ。

このとき，横軸に $\mathbf{x}$ の各々の値を示す位置をとり，その各々に $\mathbf{p}$ の各々の可能性を示す高さの棒を立ててみれば，これが確率変数の「確率分布」ということになる。

2.12　ベルヌーイ試行と 2 項分布

1 回の実験で事象 S か事象 F のどちらかが起こり，しかもそれらが起こる可能性が，$Pr(S) = p, Pr(F) = 1 - p = q$ で何回実験しても変わらないとき，これを「ベルヌーイ試行」という。ベルヌーイ試行では，事象 F は事象 S の余事象になっている。

たとえば，不透明な袋に黒い玉と白い玉が 500 個ずつ入っていて，そこから中を見ないで 1 つの玉を取り出して色を記録して（事象 S は「玉の色が黒」，事象 F は「玉の色が白」）袋に戻す実験はベルヌーイ試行である（注：袋に戻さないと 1 回実験するごとに事象の生起確率が変わっていくのでベルヌーイ試行にならない）。

ベルヌーイ試行を n 回行って，S がちょうど k 回起こる確率は，$Pr(X = k) ={}_n C_k p^k q^{n-k}$ である。${}_n C_k$ は言うまでもなく n 個のものから k 個を取り出す組み合わせの数である。2 項係数と呼ばれる。このような確率変数 X は，「2 項分布に従う」といい，$X \sim B(n, p)$ と表す。$E(X) = np, V(X) = npq$ である。

2.13　2 項分布のシミュレーション

正二十面体（各面には 1 から 20 までの数字が割り振られている）サイコロを n 回 $(n = 4, 10, 20, 50)$ 投げたときの，1 から 4 までの目が出る回数を 1 試行と考えれば，これはベルヌーイ試行である。1 回投げたときに 1 から 4 までの目が出る確率は 0.2 であるとして（＝母比率を 0.2 とする），試行 1000 セットの度数分布を図 2.3 に示す。これを描いた R のプログラムは下記の通り[*3]。

[*3] `runif(n,1,21)` では 21 が出る可能性がゼロではないが，R-1.7.0 と 1.7.1 での乱数生成アルゴリズムの標準であるメルセンヌツイスター法では $1/(2^{19937} - 1)$ しかないので無視して差し支えな

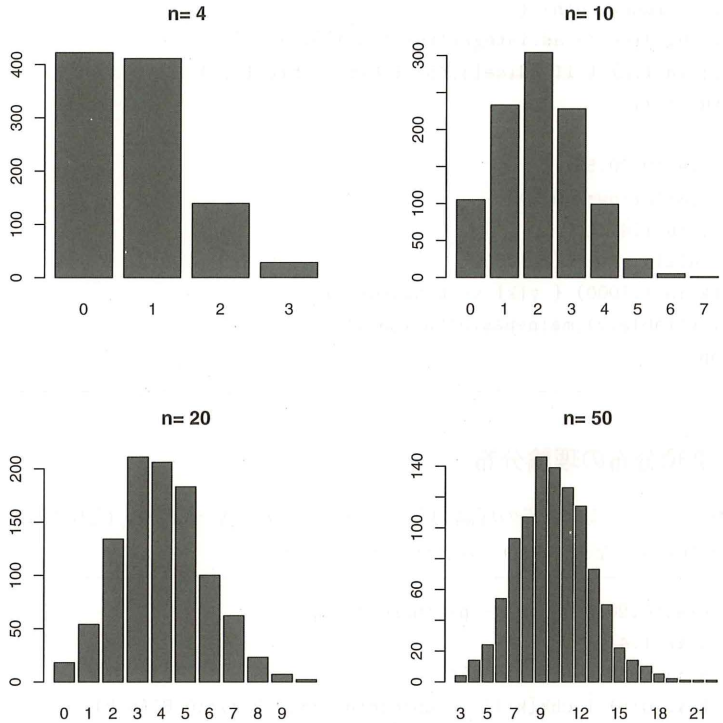

図 **2.3**　2項分布のシミュレーション

い。

```
times <- function(n) {
hit <- 0; dice <- as.integer(runif(n,1,21))
for (j in 1:n) { if (dice[j]<5) { hit <- hit+1 } }
return(hit)}

a <- c(4,10,20,50)
op <- par(mfrow=c(2,2))
for (i in 1:4) {
nx <- a[i]; y <- c(1:1000)
for (k in 1:1000) { y[k] <- times(nx) }
barplot(table(y),main=paste("n=",nx)) }
par(op)
```

times <- function(n) {
hit <- 0; dice <- as.integer(runif(n,1,21))
for (j in 1:n) { if (dice[j]<5) {hit <- hit+1 }}
return(hit)}
a <- c(4,10,20,50)
op <- par(mfrow=c(2,2))
for (i in 1:4) {
nx <- a[i]; y <- c(1:1000)
for (k in 1:1000) { y[k] <- times(nx) }
barplot(table(y),main=paste("n=",nx)) }
par(op)

2.14　2項分布の理論分布

この例で，各 n についての理論的な確率分布は，$Pr(X = k) =_n C_k 0.2^k 0.8^{n-k}$ より図 2.4 のようになる。R のプログラムは下記の通り。

```
a <- c(4,10,20,50); op <- par(mfrow=c(2,2))
for (i in 1:4) {
n <- a[i]; chk <- c(1:(n+1))
for (k in 0:n) { chk[k+1] <- choose(n,k)*(0.2^k)*(0.8^(n-k)) }
barplot(chk,col='red',main=paste("n=",n)) }
par(op)
```

ただし，R には様々な確率分布についての関数があり，`choose(n,k)*(0.2^k)*(0.8^(n-k))` は `dbinom(k,n,0.2)` と同値である。このように，確率変数が取りうる各値に対して，その値をとる確率を与える関数を確率密度関数という。値が小さいほうからそれを全部足した値を与える関数（つまり，その確率変数の標本空間の下限から各値までの確率密度関数の定積分）を分布関数（あるいは確率母関数，累積確率密度関数）と呼ぶ。

2.15　正規分布

n が非常に大きい場合は，2項分布 $B(n,p)$ の確率 $Pr(X = np + d)$ という値が，

$$\frac{1}{\sqrt{2\pi npq}} \exp\left(-\frac{d^2}{2npq}\right)$$

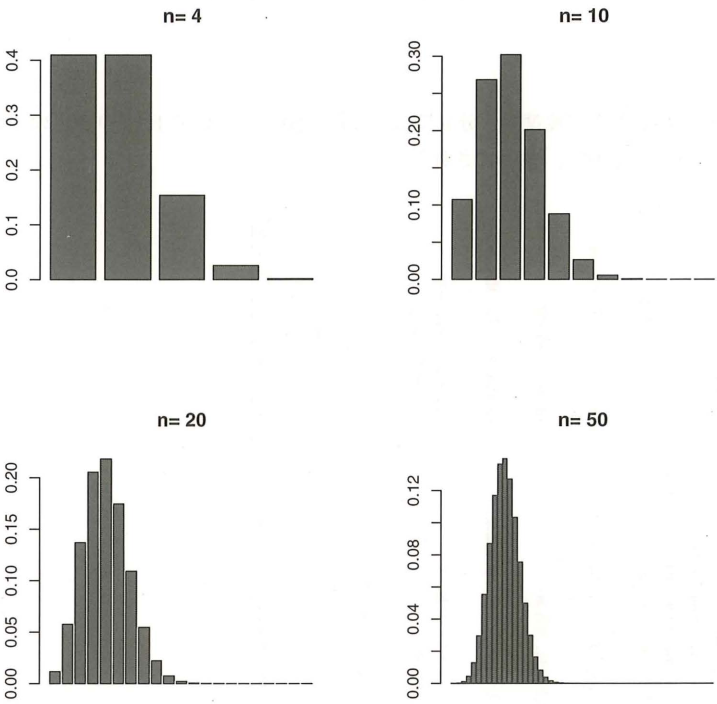

図 **2.4**　2 項分布の理論分布

で近似できる。一般にこの極限（n を無限大に限りなく近づけた場合）である，

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

という形をもつ確率分布を正規分布と呼び，$N(\mu,\sigma^2)$ と書く。

$z=(x-\mu)/\sigma$ と置けば，

$$Pr(Z=z)=\frac{1}{\sqrt{2\pi}}\exp\left(\frac{-z^2}{2}\right)$$

となる。これを標準正規分布と呼び，$N(0,1)$ と書く。統計学でよく使われる確率分

布であるカイ二乗分布とか t 分布とか F 分布は，正規分布から導かれる分布である。

参考

よく使われる確率密度関数，分布関数，分位点関数について R での表現の一覧を下表にまとめておくので，参考にされたい。

分布の種類	確率密度関数 (probability density function)	分布関数＝確率母関数＝累積確率密度関数 (distribution function = probability generating function cumulative probability density function)(4)	分位点関数 (quartile function)
カイ二乗分布	dchisq(カイ二乗値, 自由度)	pchisq(カイ二乗値, 自由度)	qchisq(%, 自由度)
2項分布	dbinom(生起回数, 試行回数, 母比率)	pbinom(生起回数, 試行回数, 母比率)	qbinom(%, 試行回数, 母比率)
ポアソン分布	dpois(生起回数, 期待値)	ppois(生起回数, 期待値)	qpois(%, 期待値)
正規分布 (1)	dnorm(Zスコア, 平均, 標準偏差)	pnorm(Zスコア, 平均, 標準偏差)	qnorm(%, 平均, 標準偏差)
対数正規分布 (2)	dlnorm(Zスコア, 対数平均, 対数標準偏差)	plnorm(Zスコア, 対数平均, 対数標準偏差)	qlnorm(%, 対数平均, 対数標準偏差)
一様分布 (3)	dunif(値, 最小値, 最大値)	punif(値, 最小値, 最大値)	qunif(%, 最小値, 最大値)
t 分布	dt(t 値, 自由度)	pt(t 値, 自由度)	qt(%, 自由度)
F 分布	df(F 値, 第1自由度, 第2自由度)	pf(F 値, 第1自由度, 第2自由度)	qf(%, 第1自由度, 第2自由度)

(1) 平均と標準偏差は省略可能。省略時は標準正規分布 (平均 0, 標準偏差 1) になる。

(2) 対数平均と対数標準偏差は省略可能。省略時は対数平均 0, 対数標準偏差 1 になる。なお，対数平均と

は自然対数をとった値の平均，対数標準偏差とは自然対数をとった値の標準偏差をいう。dlnorm(1) は dnorm(0) と等しい。対数をとる以上，当然のことながらゼロを含むデータについては計算できない。

(3) 閉区間である（区間に両端を含む）。省略時は 0 と 1 になる。

(4) R には，これらの分布関数に従う乱数を生成する関数もある。たとえば，0 から 1 までの一様乱数を 1000 個生成する関数は `runif(1000,0,1)` である。試行回数 100 回，母比率 0.2 の 2 項分布に従う乱数を 1000 個発生させるには，`rbinom(1000,100,0.2)` とすれば良い。