確率統計学
（データ解析学）

書き込み式ノート（Ver.1）

担当教員：綴木　馴

●順列

一般に，　個の異なるものから，任意に　個とって，
1列に並べる順列の数は

で与えられ，「パーミュテーション　，　」
または「　」と読む.
特に　のとき，すなわち
異なる　個のものを全部1列に並べるときの順列の数は，

である．　は　の階乗という．この記号を使うと

と書くことができる．ただし，　と定義する．

●組み合わせ

与えられた複数個のものから，順序付けはしないでいくつか選んだ組を，
組み合わせ（combination）という.

一般に，$\boxed{}$ 個の異なるものから，任意に $\boxed{}$ 個とった組み合わせの数は

$$\boxed{}$$

で与えられ，「コンビネーション $\boxed{}$ 」または「 $\boxed{}$ 」と読む.

コンビネーションについては以下が成り立つ

$$\boxed{}$$

●確率に使う語句の定義

1. 測定する操作を $\boxed{}$ と言う.
2. 試行による結果を $\boxed{}$ と言う.
3. 起こりうる結果の全体を $\boxed{}$ と言う.
4. 標本空間でそれ以上にわけられない事象を $\boxed{}$ と言う.
5. 二つ以上の根元事象を含むものを $\boxed{}$ と言う.

例えば，サイコロ振りの場合，事象「一つの目が出る」は根元事象であり，事象「偶数目が出る」は3通りあるから結合事象である.

●確率の定義

ある試行において，標本空間の大きさが $\boxed{}$ 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 $\boxed{}$ をとり，
$\boxed{}$ の起こる場合の数が $\boxed{}$ であるとき，
$\boxed{}$ の確率 $\boxed{}$ を

$\boxed{\phantom{P(A)=\frac{a}{N}}}$ と定義する.

この確率を $\boxed{\phantom{\text{数学的確率}}}$ という.

例題1

5枚の100円玉を投げて，3枚が表，2枚が裏となる確率を求めよ.
ただし，表が出るのも裏が出るのも同様に確からしいと考えて良い.

(解)

5枚投げるというのは，2個の異なるもの（表と裏）を繰り返しを許して5回投げることと同じだから，起こりうる全ての場合は，$\boxed{}$ となる.

3枚表になるのは，5個の異なるもの（5回の試行）から任意に3個（5回のうち3回）とったときの組み合わせの数に等しいので

$\boxed{}$ 通り.

数学的確率の定義を使うと，問題の確率は

$\boxed{}$ となる.

●経験的確率

野球の打率では，多くの試合の成績から，安打の出る確率をおしはかるのである．
この様な場合には，確率を次のように定義する．

$\boxed{\quad}$ 回試行を行った結果，ある事象 $\boxed{\quad}$ が $\boxed{\quad}$ 回起こったとする．
$\boxed{\quad}$ を大きくしていくとき，$\boxed{\quad}$ が一定の値 $\boxed{\quad}$ に近づけば，
$\boxed{\quad}$ の確率 $\boxed{\quad}$ を，

$$\boxed{\qquad\qquad}$$

とする．

この確率を**経験的確率**または**統計的確率**という．実際は試行を無限界行うことは不可能である．しかし，$\boxed{\quad}$ が十分大きいとき，$\boxed{\quad}$ を確率 $\boxed{\quad}$ としても，その誤差はかなり小さいと期待して良い．これを**大数の法則**という．

例題２

打率0.333≒1/3のバッターが，ある試合の第1,第2打席で凡退した.
第3打席でヒットを打つ確率はいくらであるか答えよ.

(解)

2回凡退したからといって3回目で必ずヒットを打つわけでは無い.
打率は若干下がっているが，やはり $\boxed{}$ 程度であるのに変わりは無い.

よって，ヒットを打つ確率は約 $\boxed{}$ である.

確率の公理

公理とは，その他の命題を導きだすための
前提として導入される最も基本的な仮定

サイコロを振って，いつ6の目が出るかという試行を考える．
この場合，いつ6が出るか分からないから，標本空間の大きさは
無限である．

すなわち，第　回目に6の目が出るという事象を　とすると，
それは　と限りなく存在するのである．

この様な加算無限（数えられるが有限個ではない）の場合にも，
確率の規則が当てはまる様にすることを，これから考える．

確率の公理

標本空間を　　とすると,　　は1つの集合であり,
事象　　は　　の部分集合になっている.

すなわち,　　と書ける.

　　の中で,　　の起こらない事象は余事象といい,　　と書く.

決して起こらない事象は空事象といい,　　と書く.

事象　　または　　が起こる事象を和事象といい　　と書き,
　　と　　が同時に起こる事象を積事象といい　　と書く.

確率の公理

さらに，$\boxed{}$ と $\boxed{}$ が同時に起こらないときは $\boxed{}$ と書く．

このとき，$\boxed{}$ と $\boxed{}$ は互いに排反(exclusive)である，という．

例えば，サイコロ振りで $\boxed{}$ を偶数目，$\boxed{}$ を3の目とすると，$\boxed{}$ と $\boxed{}$ は互いに排反な事象である．

また，根元事象はすべて互いに排反である．

確率の公理

標本空間 [ ] の各事象 [ ] に対して，次の3つの条件を満たす
実数 [ ] が存在するとき, [ ] を事象 [ ] が起こる確率という.

確率の公理

(1) [ ]

(2) [ ] , [ ]

(3) [ ] が互いに排反な事象のとき,
[ ]

確率の公理

(1)は確率は0と1の間であらわす，と言う意味.

(2)は全てが起こる確率は1で，何も起こらない確率は0である，と言う意味.

(3)は，経験的確率の性質を一般的に書いたものである.
例えば，サイコロ振りで偶数目が出るという事象 $\boxed{}$ の起こる確率 $\boxed{}$ は $\boxed{}$.
3の目が出る事象 $\boxed{}$ の起こる確率 $\boxed{}$ は $\boxed{}$ である.
$\boxed{}$ と $\boxed{}$ は同時に起こらない.
偶数目または3の目の出るという事象 $\boxed{}$ の確率 $\boxed{}$ は $\boxed{}$ である.
結局，$\boxed{}$ が成り立っている.
(3)では事象の個数が無限であってもよいとしているのである.

もちろん，この規則は公理であるから証明のできる様な筋合いのものではない.
要するに，この公理を認めて，これから確率の性質を調べよう！と言う事である.

確率に関する公式

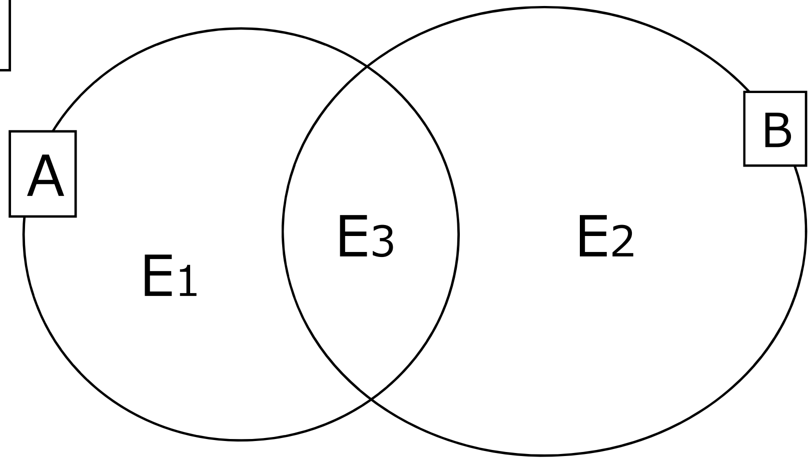


図1

ある事象　と　を考える．図1のように，

とすると，　　　　　　　は互いに排反であり，

と表わせる．

確率に関する公式

確率の公理(3)を使うと

また

結局，次の式が成り立つ.

加法公式

確率に関する公式（例題１）

よく切ったトランプ53枚（ジョーカーも入っている）から，1枚をとりだして，
そのカードがスペードである（A）か，絵札である（B）かの確率を求めよ．
ただし，ジョーカーは絵札ではない．

解

スペードである確率は　　　　，
絵札である確率は
スペードの絵札である確率は　　　　である．

よって，求める確率は

余事象の起こる確率

ある事象 $\boxed{}$ の余事象 $\boxed{}$ を考えると，

よって，確率の公理（２）と（３）より，

したがって，余事象の起こる確率に対して，

確率に関する公式（例題２）

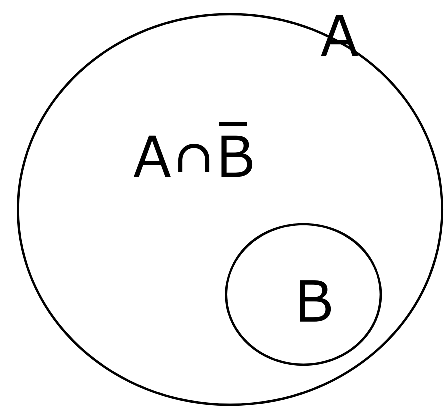
雨が降る確率が60%のとき，雨が降らない確率を求めよ

雨が降らない確率は

すなわち，　　　　　　である.

確率に関する単調性の性質

図



図のように，ある事象 [　] と [　] について，[　　　] のときは，
[　　　　] だから，

[　　　　　　　　]

と表せる．[　] と [　　　] は共通部分を持たないから確率の公理（３）より，

[　　　　　　　　]

ところが，確率の公理（１）より，[　　　　] であるので，（次ページに続く）

確率に関する単調性の性質

結局，次の確率に関する単調性の性質が成り立つ.

ある事象が起こる確率は，その一部が起こる確率より，
「決して小さくない」という事である.

完全確率の定理

標本空間 　 が　 個の根元事象 　 から成り立っているとする.

すなわち,

根元事象は互いに排反だから，確率の公理（３）から,

が言える．また確率の公理（２）から, 　 であるから,

が成立する．すなわち，全ての根元事象の確率の和は１である.

この性質を完全確率の定理という.

完全確率の定理（例題１）

〇×（まるばつ）式の問題が10題あるばあいにおいて，
でたらめに印しをつけたとき，
少なくとも1問正解する確率を求めよ．

解

少なくとも１問正解する確率を $\boxed{}$ とおき，
全て間違う確率を $\boxed{}$ とすれば，

条件付き確率

2つの事象$\boxed{\quad}$，$\boxed{\quad}$があって，$\boxed{\quad}$が起こったという条件のもとで$\boxed{\quad}$が起こると言う事象を$\boxed{\qquad\qquad}$であらわす.

その確率$\boxed{\qquad\qquad}$を，条件$\boxed{\quad}$のもとでの$\boxed{\quad}$の条件付き確率（conditional probability）といい，

$\boxed{\qquad\qquad\qquad\qquad}$

で，定義する.

乗法定理

でなければ，条件のもとでの の条件付き確率も定義出来る．すなわち，

上式と，条件付き確率の分母を払うと，次の乗法定理が得られる．

●例題

くじ引きをするときの，引く順番と当たる確率を調べてみる．例えば，
10本中3本の当たりくじがあるとし，最初に引いた人が当たるという事象を$\boxed{\quad}$，
2番目に引いた人が当たるという事象を$\boxed{\quad}$とする．

$\boxed{\quad}$が起こる確率は，$\boxed{\qquad\qquad}$．$\boxed{\quad}$が起こらない確率は$\boxed{\qquad\qquad}$である．

$\boxed{\quad}$が起こるのは，2つの場合がある．$\boxed{\quad}$が起こったときに，$\boxed{\quad}$が起こるのは，
$\boxed{\quad}$本中$\boxed{\quad}$本だから，条件付き確率として，$\boxed{\qquad\qquad}$．
$\boxed{\quad}$が起こらなかったときに，$\boxed{\quad}$が起こるのは$\boxed{\quad}$本中$\boxed{\quad}$本であるから，
$\boxed{\qquad\qquad}$．乗法定理を使うと，

$\boxed{\qquad\qquad\qquad\qquad}$

事象$\boxed{\qquad}$と$\boxed{\qquad}$は互いに排反であり，$\boxed{\qquad\qquad}$
であるから，確率の公理より，

$\boxed{\qquad\qquad\qquad\qquad}$

結局，最初に引いても次に引いても確率は同じ．

ベイズの定理

たとえば，いくつかの機械から作られた多数の同種の製品から，
作為なしに1つ取り出して，それが不良品であったとき，その製品がどの機械で作られたのかと言う確率を知りたいときがある.

すなわち，ある結果が起こったとき，それがどの原因によるのかを調べたいのである.

→ベイズの定理を使って調べる.

ベイズの定理

いま，機械が3つあるとする．3つの機械をA，B，Cで区別し，
一つの製品がこれらから作られたものであるという事象を$\boxed{}$，$\boxed{}$，$\boxed{}$ で表す．
ただし，機械は製品をそれぞれ，$\boxed{}$，$\boxed{}$，$\boxed{}$の割合で作っているとする．

製品を一つ取り出したときに，不良品であるなしにかかわらず，
どの製品から作られたかというのは分かっているのである．

不良品である事象を$\boxed{}$としたとき，それぞれの機械が不良品を作る確率

$\boxed{}$

も分かっているとする．ある原因によって結果が生じる
因果関係の確率が分かっているのである．

このとき，実際に結果$\boxed{}$が起こったとき，その結果が$\boxed{}$，$\boxed{}$，$\boxed{}$ の
どの原因によるかという確率 $\boxed{}$，$\boxed{}$，$\boxed{}$ を求める．

ベイズの定理

などは，結果と関係がないので，存在の確率（または事前確率）といい，
などは，結果が起こったときの原因を考えているから，
原因の確率（または事後確率）という．

，，は互いに排反は事象だから，結果が原因による
確率の比は，でで，でで，でが起こる確率の比に等しく，

となることが直感的にわかるだろう．実際，たとえばについては，
乗法の定理から，

であるから，

と書ける．

ベイズの定理

ところが，　　　であり，

は，それぞれ排反な事象だから，

である．

やはり乗法定理から，　　　が成り立つので，

と書ける．したがって，

ベイズの定理

一般に，ある事象 $\boxed{}$ が，$\boxed{}$ 個の互いに排反で全ての場合をつくす原因，$\boxed{}$ によっているとき，そのうち1つの $\boxed{}$ によって起こる確率 $\boxed{}$ は，

$$\boxed{\phantom{P(H_i \mid A) = \frac{P(H_i)P(A \mid H_i)}{\sum_{j=1}^{n} P(H_j)P(A \mid H_j)}}}$$

で表される．これを**ベイズ（Bayes）の定理**という．

ベイズの定理（例題）

3つの機械のうち，生産量の10%を A，30%を B，60%を C が占めているとする.
また，不良品である割合が，Aは3%，Bは2%，Cは1%であるとする.
一つの製品を取り出したら不良品であったとき，それがAの製品である確率を求めよ.

統計的独立

条件付き確率 は, が起こったときに が起こる確率であった.
しかし, が に何の影響も及ぼさないこともある.
例えば，2個のサイコロを振るとき，一方のサイコロの目が，他方のサイコロの出方に関係することはない．この様なとき,

乗法定理は，このとき

と書ける.

上式が成立しているとき，事象 と 事象 は確率的に独立であるという.

一般に, 個の事象 があるとき，それから取り出した任意個の事象 に対して

が成立しているとき，事象 は互いに統計的に独立である.

統計的独立

統計的独立の定義は，事象だけでなく，試行に対しても考えられる．

例えば，サイコロを何回も振るときのように，一定の条件のもとで，
同じ試行を繰り返すとき，1回ごとの試行は他の試行に影響しない．

この様な試行を，　　　　　　　または，ベルヌーイ(Berunoulli)試行という．

確率変数

標本空間の中の根元事象に対して適当な数値を対応させた変数 $\boxed{}$ を考えて，
その変数がどの数値をとるかは偶然に支配されるけれども，
$\boxed{}$ がある特定の数値 $\boxed{}$ をとる確率，すなわち，ある根元事象の起こる確率が定まっているとき， $\boxed{}$ を確率変数（random variable）という．

確率変数は試行を行って初めて値が決まる変数であり，
単なる数値と区別して，一般的に大文字であらわすことが多い（ので大文字にする）

また，おのおのの試行の結果は（根元事象に割り振った）単なる数値であるので，
一般的に小文字であらわすことが多い（ので小文字にする）．

確率変数（離散的な場合）

確率変数 $\boxed{}$ がとびとびの値

$\boxed{}$

しかとらないとき, $\boxed{}$ を離散変数という.

例えば, サイコロ振りの場合は,

$\boxed{}$

である. これは有限個であるが, 無限個あっても構わない. また, $\boxed{}$ が連続した値をとるときは, 連続変数という（これは後でやる, 連続の方が重要）

$\boxed{}$

$\boxed{}$ は変数全体の集合.
$\boxed{}$ はその要素.

確率密度関数（離散的な場合）

確率変数　が離散的で有限個の値をとるとする.
　が　となる確率を

と表す．サイコロ振りのときは,　である.

このとき,　のとる値それぞれに対して確率　の数値が定まるから,
確率は関数の形で,

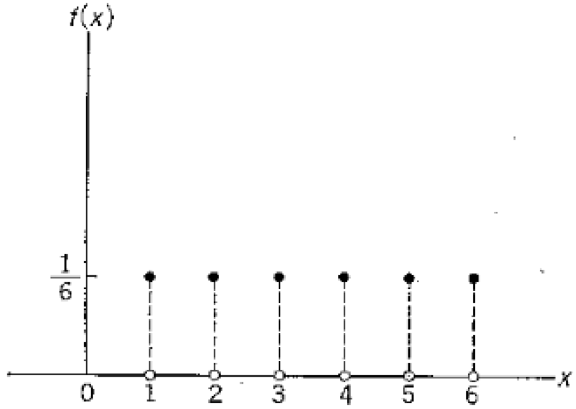
と書くことができる.　のときは確率が0だから,　としたのである.

このようにして定められた関数を確率密度関数(probability density)という.

確率密度関数（離散的な場合）

離散的な場合には確率密度関数を特に確率密度ということもある.

サイコロ振りの場合の確率密度のグラフは下図のようになる.



確率の和は1であるから,

が成り立つ.

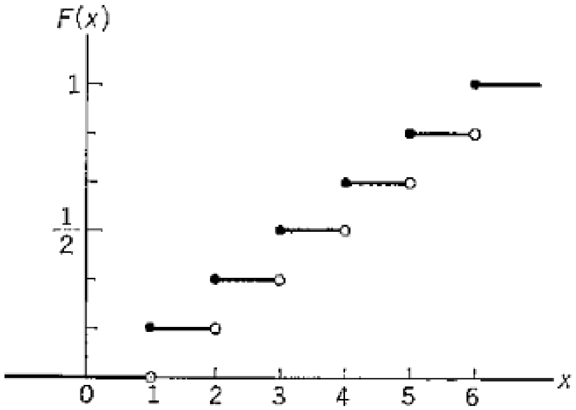
分布関数

確率変数　のとる値が　以下である確率に対しても，

という関数を考えることができる．たとえば，サイコロ振りのとき，
　は　または　となる確率だから，

　である．

　もやはり　である．この様な関数　を
分布関数(distribution function)という．サイコロ振りの場合の分布関数のグラフは



となる．

図からも分かるように，離散的な場合の分布関数には，
次ページの性質がある．

分布関数

(1) $\boxed{}$

(2) $\boxed{}$ は $\boxed{}$ について減少することのない階段状の関数であり，

$\boxed{}$

(3) $\boxed{}$

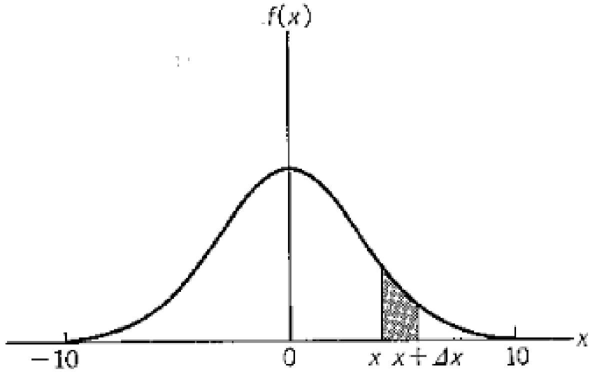
これらの関数の定義は，確率変数が無限個の離散的な値をとるときにも当てはまるが，これからは，必要なとき以外は有限個の場合を考えることにする．

確率変数（連続的な場合）

確率変数 が連続的な値をとり，その値の範囲が であるとする．

このとき， が と の間にある確率が，

となるような関数 がやはり確率密度である．



白色ノイズの確率密度のグラフは左図のようになり，
陰部分の面積が に相当している．

が小さいとき，この面積は で近似できるから

とかける．

確率変数（連続的な場合）

離散変数のときと比べて本質的に違う点は，確率がある区間での積分で表されているため，区間の端が含まれているかいないかは確率に関係しない.

したがって，［空欄］は，

［空欄］

［空欄］と書いても同じことである.

この事実は測定に誤差のある現実問題において妥当なものである.
例えば，電気的な雑音電圧の測定で，2.0Vの値を得たいと思ったとき，
これはちょうど2.0Vであると言うわけではなく，たとえば，
1.95V～2.05Vの間にあるのを2.0Vとしているのである.

確率密度（連続的な場合）

確率変数$\boxed{}$のとる値が$\boxed{}$と$\boxed{}$の間に限られているとすると，
全ての確率の和は１であるので，確率密度は，

$$\boxed{}$$

を満たしている．区間$\boxed{}$の外では，$\boxed{}$であるので，

$$\boxed{\phantom{\int_{-\infty}^{\infty} f(x)\,dx = 1}}$$

と，書くこともできる．上の2つの式は40ページにおける図の
山の面積が１であることを意味している．

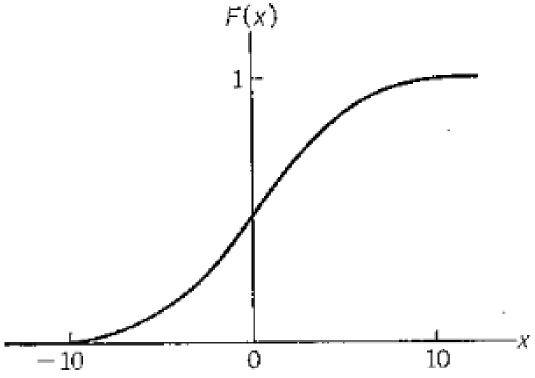
分布関数（連続的な場合）

離散的なときと同様に, のとる値が 以下でもある確率に対して,
離散ように定義された関数 を

連続的な場合も考えることができる. 確率密度を用いて表現すると,

と書ける.

この関数をやはり分布関数という.



40ページの図に対する分布関数のグラフを書くと,
左図のようになる.

分布関数（連続的な場合）

この分布関数は離散的なときと同様の性質をもつ.
すなわち，離散値の場合

に対応するものが,

であり,

に対応するものが

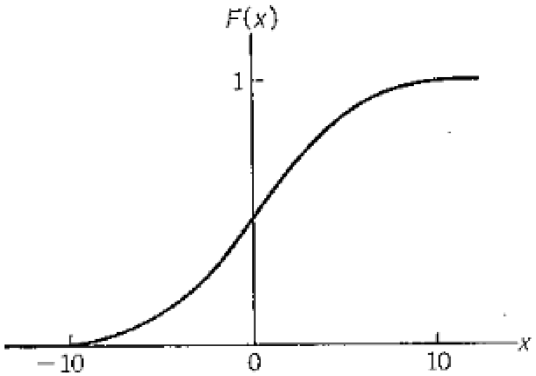
である.

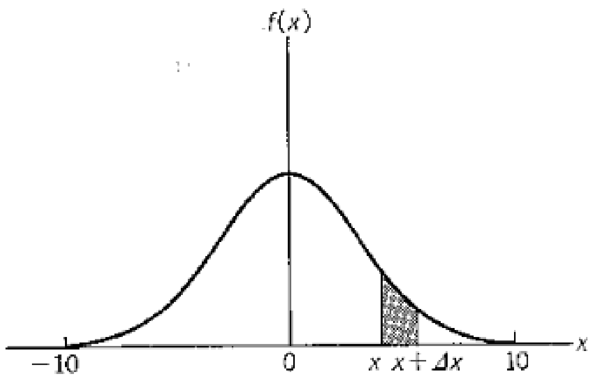
離散的な場合から連続的な場合への移行は，総和から積分への極限操作である.

確率密度と分布関数の関係

を　　について微分すると,

が得られる．これは，左図の勾配（傾き）が右図の　　の値になっていることを意味している．





例題（一様分布関数）

関数

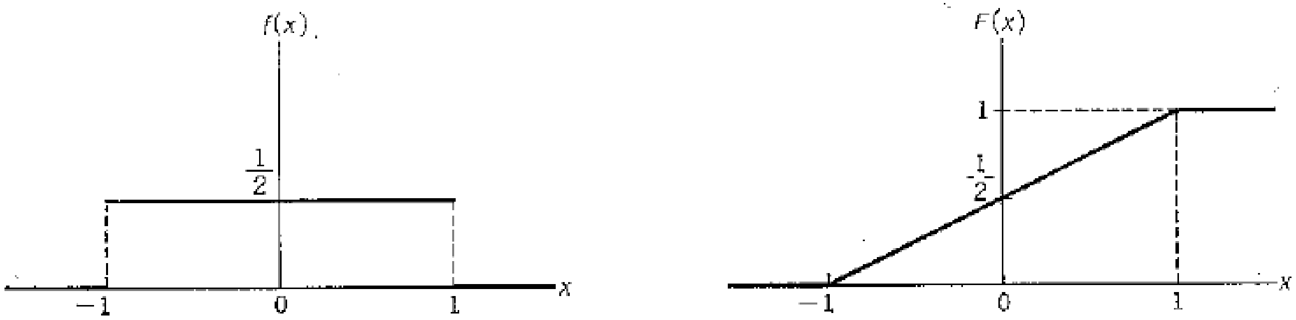
が確率密度となるようにｃの値を決め，を求めよ．また，の図も書け．
（この分布は確率変数が0でない値をとる範囲で確率密度が一定であり，
とくに一様分布という）．

解

例題（一様分布関数）の続き

解

得られた結果を図で示すと，下図の様になる.



(a) 確率密度　　(b) 分布関数

分散と期待値

確率密度または分布関数が与えられると，標本空間の中の各事象がどのような割合で起こるかが完全に分かる．
つまり，これらの関数は確率変数がどういう値をとるかについての確率的な情報を全て含んでいるのである．

ある確率変数に対して確率密度または分布関数がわかっているとき，その確率変数は与えられた確率分布（probability distribution）に従っているとか，簡単に，分布が分かっている，と言う．

しかし，現実の問題では余り細かい情報まで知る必要はなく，
確率変数がだいたいどのくらいの値をとるか，または，
大体どの程度の範囲に存在しているかといったことだけ分かれば良い場合も多い．
そのような場合には確率密度 $\boxed{}$ や分布関数 $\boxed{}$ よりももっと簡単な量を用いた方が便利である．それがこれから定義する，期待値や分散である．

期待値

確率変数　　の平均（mean or average）または期待値（expectation value）は次の式で与えられる.

平均（mean）の頭文字 m のギリシア文字が　　（ミュー）なのである.

上式での　　は離散的なときの確率密度であり,
　　は連続のときの確率密度である.

例題

ある宝くじは，1枚300円であり，1000万枚につき，

6000万円	5本,	1500万円	10本,	1000万円	10本,
100万円	60本,	10万円	595本,	7万円	90本,
1万円	2000本,	3000円	10万本,	400円	100万本

の当たりくじがある．1枚買ったとき，当たる金額の期待値を求めよ．

例えば，6000万円当たる確率は$\boxed{\qquad\qquad}$である．
他の当たり方も同様に計算して，

分散

確率変数 が大体どの程度の範囲にあるかを示すのに使われる量
を分散（variance）といいい，次の式で定義する．

また，分散の平方根をたったもの を標準偏差（standard deviation）という．

s のギリシア文字が （シグマ）なのである．

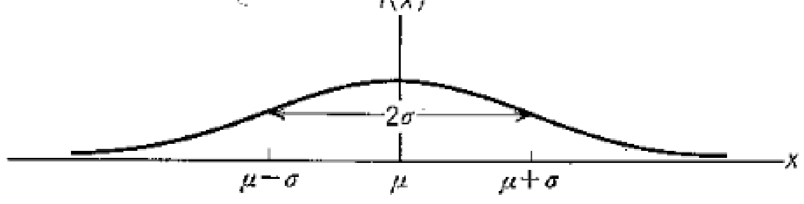
標準偏差は，分布が平均からどのくらいの幅にあるのかを示すめやすになっている．

例題１

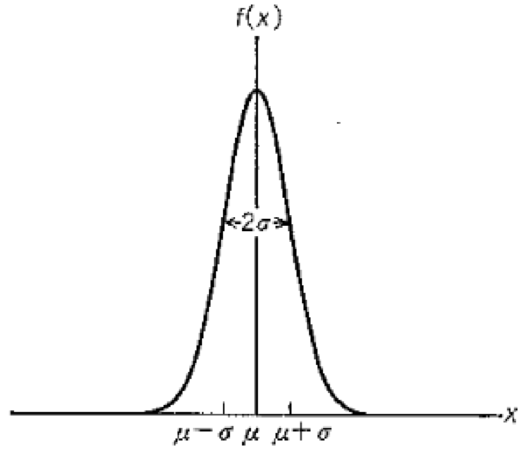
サイコロ振りの場合の平均と標準偏差を求めよ.

標準偏差の意味

分散または標準偏差が大きいときは，その分布のばらつきの程度は大きく（図a）
標準偏差が小さいときは，分布はほとんど平均値のまわりに密集している（図b）



(a) σが大きいとき



(b) σが小さいとき

正規分布

連続変数　に対して，

の形の分布を平均　，分散　の正規分布（normal distribution）といい，
　と表す．

特に，平均が 0 分散が 1 の分布

を，標準正規分布という．

正規分布

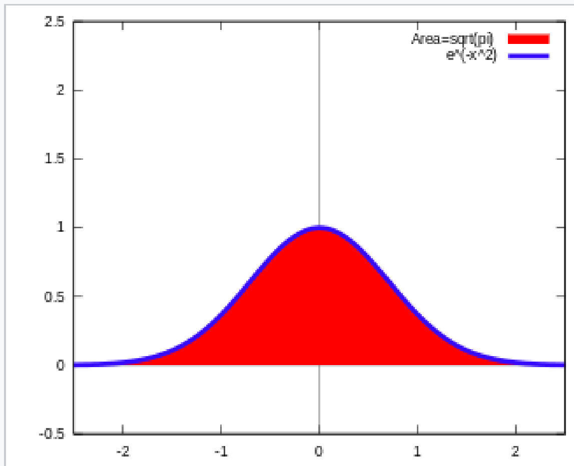
正規分布は，理工系の広い分野であらわれる．
正規分布は実用上最も重要な分布である．

正規分布はガウス（C.F.Gauss：IQ=250）が測定誤差の研究の中で見つけたので
ガウス分布または，誤差法則と呼ぶこともある．

ガウス積分と呼ばれる物があり，次が成り立つ．

$$\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$$

導出は大学1年生レベルの積分で習うべきものであり，
ここでは，あえて証明しない．しかしながら，極めて重要な
積分なので，覚えておかないと理系人として非常識である．



関数 $y = \exp(-x^2)$ のグラフと x 軸で囲まれた部分の面積 (= $\sqrt{\pi}$) がガウス積分を表す

中心極限定理

中心極限定理は，一般的な分布に対して成り立つ強力な定理である．
そのために，正規分布は実際的な統計処理を行う際に，非常に良く顔を出す．
実用的に重要な中心極限定理の一つの形として，次の命題がある．

確率変数　　が互いに独立で，平均　　，分散　　をもつ
同一の分布に従っているとする．　　の単純平均，

に対して

とすると，　　を大きくしたとき，　　の分布は標準正規分布　　
に近づく．

標準化変換

確率変数$\boxed{}$が正規分布$\boxed{}$に従っているときは,

の変換を行えばよい．すると，$\boxed{}$ は$\boxed{}$に従うことになり，
この変換を**標準化変換**と言う．