応用数学：動画講義の要約

「第1章：線形代数」

「第2章：確率・統計」

「第3章：情報理論」

について記載

「第 1 章：線形代数」

スカラーとベクトル

スカラー：普通の数。通常の演算の対象となる。

ベクトル：大きさと方向を持つ。スカラーの組で表示される

例：(2,3)

行列：ベクトルを並べたもの。横方向が行、縦方向が列

行列 と 行列 の積

X ：n行p列、Ｙ：p行m列 とすると以下の要素のn行m列の行列となる。

　　Σ Σ

　　Σ Σ

……………………

　　Σ Σ

行列 と ベクトルの積

ベクトルは行数または列数が1の行列として扱うことができるので

行列 と 行列 の積と同様。

固有値と固有ベクトル

行列A , ベクトルx ( x ≠ 0)の間に

Ax = λx を満たすスカラーλが存在する場合、

xを行列Aの固有ベクトル,λを行列Aの固有値　という。

固有値，固有ベクトルの求め方:(Eは単位行列)

正方行列Aの固有方程式det(A-λE)=0を未知数λの方程式として解いた解が、

固有値である。n次正方行列の場合、n個存在する。

各固有値を連立方程式(A-λE)x = 0 に代入して，対応する固有ベクトルxを求める。

固有ベクトルの定数倍も固有ベクトルになるので、任意定数が付くことになる。

固有値分解

正方行列Aの固有値が,, , 固有ベクトルが とすると

固有値を対角線上に並べた対角行列Λ ,

それに対応する固有ベクトルを並べた行列V ()について

AV = VΛが成り立つので、A=VΛと変形できる。

これをAの固有値分解という。

⇒　単純なベクトルの組み合わせに変換できる、

行列の累乗の計算が容易になる というメリットがある。

　特異値分解

　 正方行列以外でも固有値分解に似た概念がある。

　 M を m×n行列とする。ある実数σ に対し、

Mv = σu

u = σv ( はMの転置行列 )

という条件を満たす単位ベクトル u(m次元) と 単位ベクトル v(m次元) の組が

存在するとき、実数σを（ベクトル u, v に対応する）行列 M の特異値と呼ぶ。

またベクトル u, v を、それぞれ σ の左特異ベクトルと右特異ベクトルと呼ぶ。

このような特殊な単位ベクトルがある場合は以下の分解が可能である。

これを特異値分解という。

M = US

( S : M の特異値を対角成分に並べた対角行列

U : 左特異ベクトルを並べた行列

V : 右特異ベクトルを並べた行列

)

MV = US より M = US

U = Vより = V

これらの積は

M = USVS となり、これを固有値分解すれば左特異ベクトルと

特異値の2乗が得られる。

「第 2 章：確率・統計」

・条件付確率

ある事象X=xが起こるという条件下での別の事象Y=yとなる確率。P(Y=y|X=x) と表す。

事後確率ともいう。

P(Y=y|X=x) = P(Y=y,X=x) / P(X=x)

P(Y=y,X=x): 事象X=xと事象Y=yが両方起きる確率 :

P(X=x) : 事象X=xが起きる確率

・独立事象の同時確率

事象 X=x と事象 Y=yがお互いの発生には因果関係がない場合、

事象 X=x と事象 Y=yは独立であるという。

事象 X=x と事象 Y=yが同時に発生する確率は以下のように各々の確率の積となる。

P(Y=y,X=x) = P(X=x) P(Y=y) = P(X=x ,Y=y)

・ベイズ則 (ベイズの定理)

確率および条件付確率について、以下が成り立つ。

P(X=x |Y=y) P(Y=y) = P(Y=y|X=x) P(X=x)

ベイズの定理を使えば、事後確率P(X=x |Y=y)は下記に従って計算される。

P(X=x |Y=y) = P(Y=y|X=x) P(X=x)/ P(Y=y)

・確率変数と確率分布

確率変数: 事象と結び付けられた数値

事象そのものを指すと解釈する場合も多い

確率分布

•事象の発生する確率の分布

•離散値であれば表に示せる

例：コインを4枚なげるとする。

事象 : (表4,裏0), (表3,裏1), (表2,裏2), (表1,裏3), (表0,裏4)

確率変数: 裏0,表1として数値化

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 事象 | (表4,裏0) | (表3,裏1) | (表2,裏2) | (表1,裏3 | (表0,裏4) |
| 確率変数 | 4 | 3 | 2 | 1 | 0 |

確率分布: 事象と対応する確率

1200回投げて、事象の発生した回数が75,300,450,300,75だったとすると

下表のようになる。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 事象 | (表4,裏0) | (表3,裏1) | (表2,裏2) | (表1,裏3 | (表0,裏4) |
| 確率変数 | 4 | 3 | 2 | 1 | 0 |
| 事象の発生した回数 | 75 | 300 | 450 | 300 | 75 |
| 確率分布:  事象と対応する確率 | 1/16 | 4/16 | 6/16 | 4/16 | 1/16 |

・期待値

ある確率分布における確率変数の平均値 = Σ(確率変数×確率)

確率変数をf(x) 確率変数をP(x) とすると

期待値E(f) = P(X=) 連続値の場合P(X=x)f(X=x)dx

上記の例だと

4×1/16 + 3×4/16 + 2 ×6/16 + 1×4/16 + 0×1/16 =(4+12+12+4) /16=2.0

・分散、共分散

分散(variance): データの散らばりの度合いを表す値

データの各々の値が，期待値からどれだけズレているのか平均したもの

Var(f) = E( (f(X=x)–E(f)) \*\*2 ) = E(f(X=x)\*\*2)-E(f)\*\*2

　を標準偏差と呼び、データの散らばりの度合いを示す値である。

共分散(covariance): ２つのデータ系列の傾向の違い

正の値 ⇒ 似た傾向

負の値 ⇒ 逆の傾向

ゼロに近い ⇒ 関係性に乏し い

Cov(f,g) = E((f(X=x)–E(f))(g(Y=y)–E(g))) = E(fg)–E(f)E(g)

・様々な確率分布

ベルヌーイ分布: 確率変数が0、又は1の値をとる確率分布。

・コイントスのイメージだが、裏と表で出る割合が等しくなくとも扱える。

　　　　　・またと「Aという現象が起きたか、それ以外が起きたか」という場合に

使えるので大変重要な分布である。

・確率変数が1である確率がμであるとき,0である確率は1-μとなる。

・P(X=μ) =

マルチヌーイ（カテゴリカル）分布

確率変数が3個以上ある。

⇒確率変数がn個の場合、P(X=k) = とすると = 1 である。

さいころを転がすイメージ

各面の出る割合が等しくなくとも扱える。

二項分布

ベルヌーイ分布の多試行版：

コインn回投げて,何回表が出るかというイメージ

P(x|λ,n) = n!/( x!(n-x)! ) ×

( λ : ベルヌーイ分布のμに相当する。)

ガウス分布(正規分布)

釣鐘型の連続確率分布

[平均](https://ja.wikipedia.org/wiki/%E5%B9%B3%E5%9D%87)を *μ*, [分散](https://ja.wikipedia.org/wiki/%E5%88%86%E6%95%A3_(%E7%A2%BA%E7%8E%87%E8%AB%96))を *σ*2 > 0 とする正規分布とは次の形の分布である。



* 確率分布であるので、右辺を-∞**< x <** ∞ で積分すると1になる。

「第 3 章：情報理論」

自己情報量

P(x)の確率で起きる事象の自己情報量は以下の式で定義される。

logの底を2にしたとき、単位はbit

I(x) = -log(P(x)

⇒ 確率P(x)が低いほど情報量が多い。

たまにしか起きない事象が発生したという情報は重要度が高いという考え方。

⇒ Aが起こる確率をP(A)、Bが起こる確率をP(B)とすると

A,B両方が起こる確率は P(A) \* P(B)なので、

自己情報量は -log( P(A) \* P(B) ) = -logP(A)-logP(B)

各々の自己情報量の和となる。

シャノン・エントロピー (平均情報量)

自己情報量の平均 = 自己情報量の期待値

H(x) = E(I(x)) = -E(log(P(x)) = -Σ(P(x)log(P(x))

ベルヌーイ分布の場合、確率P(x)とエントロピーH(x)の関係は下図となる。



カルバック・ライブラー ダイバージェンス ( KLダイバージェンス )

・同じ事象・確率変数における異なる確率分布 P,Q の違いを表す

・Qの自己情報量からPの自己情報量を引いて平均を取ったもので、

分布間の距離のように考えることができる。

QとPが変わらない。⇒ 0

QとPの差が大きい ⇒ 大きな値となる。

Dkl(P||Q) = Ex～p[log(P(x)/Q(x))] = Ex～p[logP(x) - logQ(x)]

交差エントロピー

KLダイバージェンス の一部をとりだしたもの。

Q についての自己情報量を P の分布で平均している。

H(P,Q) = H(P) + Dkl(P||Q) = -Ex～p[log(Q(x)]

QがPが大きく変わらない ⇒ H(P,Q)とH(P)は大きく変わらない

QとPの差が大きい ⇒ H(P,Q)とH(P)は大きく変わる。