IoT 時代の
データ処理の基本と実践

―スマホ内蔵センサ取得データを用いて―

博士（工学） 田中　博
博士（情報科学） 五百蔵重典
共著

コロナ社

ま　え　が　き

数十年前までは，いつでも，どこでも，だれとでも，を目標としていた通信システムは，今や当然のものとなり，通信速度の高速化はいうに及ばず，機器の小型化，低消費電力化を実現し，多くのユーザがその利便性を享受している。さらに今日では，IoT（Internet of Things）という言葉に代表されるように，人ばかりではなく，人とモノ，モノとモノがつながり，データをやりとりしている。各種センサの小型化とともにそれらのデータをワイヤレスで伝送することにより，センサ設置の負荷やコストが大幅に抑えられ，結果としてデータの収集がきわめて容易に，そして大規模なデータになっている。

このような中で，データの分類やそれらの本質を把握すること，データから有意な情報を抽出することなどのために，得られたデータに対する処理技術がますます重要になってきている。これにより客観的に事象を分析すること，課題解決の指針を得ることも可能であり，そして，それらの巧拙によって大きな差異が生じる時代であるといえる。

本書はこのような背景を踏まえ，得られたデータに対して，客観的な解析，処理を行うために必要となる基本的なデータ処理，解析の技法を述べるものである。これから専門的な統計解析，機械学習，信号処理を学んでいく，あるいは研究対象とする学生が，その前準備として，前提となる基礎知識やデータ処理技術の基本を身につけることのできる教科書あるいは参考書となることを目的として執筆している。

第 1 章では基本中の基本である各種統計量について述べ，第 2 章では代表的なデータ解析手法である回帰分析，第 3 章は推定と検定について，その手法とともに得られたデータの有意性の判定方法などを例を示して説明している。第 4 章では，三つの基本的なデータの判別方法を取り上げ，それらの考え方を示

している。これは機械学習の基礎となるものである。第5章では時系列データに対する代表的な解析手法であるフーリエ解析について述べ，そして第6章では信号処理技術として，ディジタルフィルタの構成とその効果を示す。

本書では可能な限り，具体的な例を示して読者の理解を高めるように配慮している。数学的な厳密さよりも，実際の問題への適用という観点を重視している。最初から順を追って読んでいくことが望ましいが，興味を持った章から読み，必要となる知識を他の章から得るように読み進めていくことも可能である。スマートフォン（スマホ）には多数のセンサが実装されているが，その中から加速度センサを取り上げ，実際に取得したデータを用いて解析している。データ取得のためのプログラムもAndroid端末用であるが，ダウンロード可能である。実際に自分のスマホでデータを取得してデータ処理方法を体験することにより，より理解が深まると思われる。また，本書で取り上げたExcelのデータをコロナ社の本書籍詳細ページ（http://www.coronasha.co.jp/np/isbn/9784339028805/）よりダウンロードできる。理解を深めるために，必要に応じて活用していただきたい。読者が本書を通して基本知識を習得し，つぎのステップにつなげることができれば，本書はその目的を十分に達したものと考えている。

最後に，本書の企画を担当し，執筆の機会をくださったコロナ社にこの場を借りてお礼を申し上げたい。スマホのセンサデータの取得方法として，MIT App Inventorの紹介とともに，日頃より多くの示唆をくださる神奈川工科大学情報工学科教授の山本富士男先生に深謝したい。また，執筆にあたり，プログラムや図面の作成に寄与してくれた本学学生の金田一将君，門倉 丈君，岡安優奈さんに感謝する。

2018年1月

著者しるす

目　　次

1. 統計処理の基本

2. 回帰分析

3. 推定と検定

4. データの分類

5. フーリエ解析

タと異なっているためである。そこで，物理的意味を考慮した標準偏差が定義される。

$$\sigma = \sqrt{V} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mu\right)^2} \tag{1.10}$$

ここで，x_i は各データ，n はデータ数，μ は平均値である。標準偏差は σ（小文字のシグマ）で表されるが，総和を示す Σ（大文字のシグマ）とは同じシグマでも意味が異なる。

取り扱う物理量によるが，一般的に平均値だけではデータの性質を示す値としては不十分であることが多い。データの性質を報告する際は，最低限，データ数，平均値，そしてこの標準偏差あるいは分散を明記することが好ましいと思われる。

—— 事例：加速度データの基本統計量

本事例では，実データから統計量を求める一例を示すために，スマートフォン（スマホ）内に実装された加速度センサで取得したデータを用いる。スマホに内蔵された加速度の検出軸方向を**図 1.3** に示す。

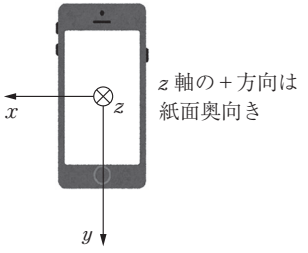

図 1.3　スマートフォンの加速度の検出軸方向

スマホの所持方法の相違による，検出軸の検出値が異なるという影響をなくしたい。そのために，取得データに対して式 (1.11) に示す 3 軸合成と呼ばれる処理を適用する。A は 3 軸合成値と呼ばれることがある。

$$A = \sqrt{a_x^{\,2} + a_y^{\,2} + a_z^{\,2}} \tag{1.11}$$

ここで，a_x は x 軸方向の加速度，a_y は y 軸方向の加速度，a_z は z 軸方向の加速度である。

スマホを手に持って小さくゆっくり振ったときのデータと，大きく速く振ったときのデータを取得し，式 (1.11) の処理を施すと，**図 1.4** のそれぞれの時系列データになる。なお，ここでの横軸はデータを取得した順番を示す整数

—— 事例：サーミスタの抵抗–温度特性 ——

サーミスタとは，温度による抵抗値の変化を用いて，温度測定を行うセンサである[†1]。**表 2.1** の抵抗–温度特性が仕様として提示されている場合の抵抗–温度特性直線を求める[†2]。

表 2.1　サーミスタの抵抗–温度特性

温　度〔℃〕	抵　抗〔kΩ〕
−10	3.651
0	2.449
10	1.684
20	1.184
25	1.000
30	0.848 6
40	0.618 9

温度と抵抗値の回帰直線を，Excel の機能を用いて求める。近似曲線は，1次（回帰直線）ではなく，m 次の回帰曲線にすることもできる。ただし，通常は3次までにとどめることが多い[†3]。1次式と3次式の場合の結果を，**図**

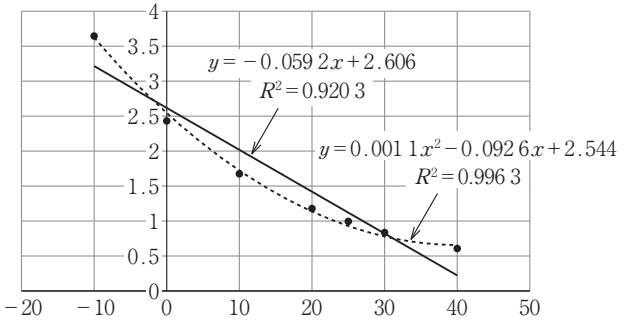

図 2.13　サーミスタの抵抗–温度特性と回帰直線と曲線

†1　マイコンの入力は電圧信号であり，各種センサの物理的変化（この場合は抵抗値）を回路によって電圧変化として検出して，マイコン内での処理によって温度として計測する。

†2　高精度サーミスタ特性表による。(http://akizukidenshi.com/download/ds/semitec/at-thms.pdf)

†3　オーバーフィッティングと呼ばれる，対象の特性に過剰に合わせることを回避するためである。オーバーフィッティングのある状態で，値を予測すると，近似誤差が大きくなることがある。

れは，図 3.8 からも明らかである。

最近は方法 2 を用いる場合が多いようなので，以降の説明は方法 2 で述べる。具体的な手順は以下である。

STEP1：仮説（対立仮説もしくは帰無仮説の）設定

STEP2：統計量 T の算出

STEP3：統計量 T に対応する p 値の算出（p 値は T から自動的に決まる）

STEP4：p 値と有意水準との比較

ここで，有意水準は 0.05 または 0.01 であり，通常は 0.05 がとられる。

STEP5：仮説の判定（棄却もしくは，採択）

ここで，1 標本 1 集団の母平均の検定用の統計量 T は，式 (3.7) で与えられる。

$$T = \frac{\overline{x} - m_0}{\frac{u}{\sqrt{n}}} \tag{3.7}$$

ここで，$\overline{x}$ は標本平均，m_0 は比較値，u は標本標準偏差，n は標本数である。

3.9.2　p 値の算出方法

p 値は正規分布と t 分布（前述したが，標本数が少ない場合に，正規分布に従う母集団の平均値を推定するときに使用される）によって異なる値となる（表 3.2）。標本数 n が 100 以上の場合は標準正規分布を利用する Z 検定を用い，標本数 n が 100 未満の場合は t 分布を利用する t 検定を用いる。p 値は，表 3.1 で示した Excel の関数を用いて，両側検定，片側検定それぞれにおいて **表 3.3** で求められる。

ここで，NORM.S.DIST 関数は，標準正規分布の値を，T.DIST.2T，T.DIST.RT 関数はそれぞれ，両側 t 分布，右側 t 分布の値を返す関数である。なお，後者二つの関数は Excel の旧バージョンでは，それぞれ T.DIST(T, n−1, 2)，T.DIST(T, n−1, 1) として提供されていたものである（最新版の Excel 2016 でも使

4.3 線形判別分析

前節で述べた相関比が最大となるように二つのグループを分けることができれば，それは一つの基準で最もよく二つのグループに分けることになる。そのための数学的に明確な判断基準を求めるのが判別分析であり，直線で2グループに分ける方法を線形判別分析という。

ここで，つぎの例を考える。これはある検査1と検査2の結果とガンの有無の評価結果である（**表 4.4**）。ここで，検査結果であるガンの有無は数値，すなわち1はガン有，2はガン無で表している。グラフで表現すると**図 4.7** になる。グラフから明らかに，検査結果1および2の結果が低いことがガンの可能性が高い傾向にあることが把握できる。この二つを分離する直線を求めることができれば，二つのグループに分離できることになる。そして，新たなサンプルに対して，その直線を適用することによって，どちらのグループに属するものかを判別することができる。

数学的にガンの有無の判別を行う。すなわち，二つのグループに分割する直

表 4.4　検査結果とガンの有無

被検査者	検査結果1	検査結果2	ガンの有無
A	3	2	1
B	4	1	1
C	2	2	1
D	2	3	1
E	4	5	1
F	4	4	2
G	5	8	2
H	3	6	2
I	6	7	2
J	5	4	2

1：ガン有
2：ガン無

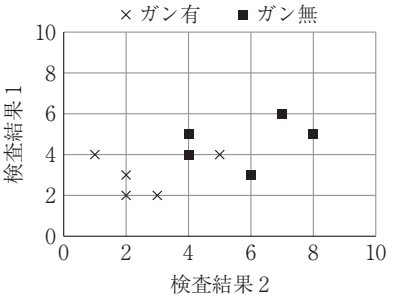

図 4.7　検査結果とガンの有無

【解答】

周波数分解能は式 (5.72) より，$f_s/N=1\,000/512=1.95$ Hz である。また，解析できる信号の最高周波数は，ナイキスト周波数である 1 kHz/2 の 500 Hz である。

離散フーリエ変換のスペクトルは周期的であり，その周期はポイント数 N である。負のスペクトルは，$k=N/2$ から $k=N-1$ に現れる。すなわち，振幅スペクトル，パワースペクトルは $k=N/2$ を中心とした左右対称である。フーリエ係数としては，c_0，c_1，…，c_{511} であるが，c_{257} 以降は負のスペクトルで意味を持たないものである。よって，$k=256$ である。 ◆

事例：加速度データのフーリエ変換

加速度データに対してフーリエ変換を適用した例を示す。ここでは，① 肩を中心にゆっくり円を描く形で腕を回転させた場合と，② 肘を中心に，同じく円を描く形で速く回転させた場合での加速度データを取得した。得られた 3 軸方向の加速度を合成した $\left(\sqrt{a_x^{\,2}+a_y^{\,2}+a_z^{\,2}}\right)$ データに対して，FFT を適用した。FFT の実行方法は 5.9.1 項で述べたとおりである。

ゆっくり回したときの結果を**図 5.18** に示す。このときのポイント数は，$N=64$ である。データ取得時のサンプリング周期は約 0.1 s†，したがってサンプリング周波数は 10 Hz，ナイキスト周波数は 5Hz である。確かに，ナイキスト周波数を中心にしてエイリアスが発生していることがわかる。物理的に意味がある区間 0 Hz ～ 5 Hz を表示したものが**図 5.19** である。ピーク周波数は 1.2

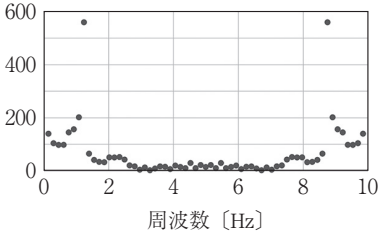

図 5.18　FFT の結果（エイリアスの確認）

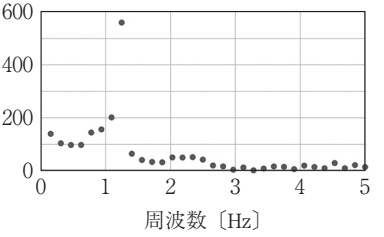

図 5.19　FFT の結果（$N=64$）

† サンプリング間隔が一定であるべきだが，スマートフォンで取得できるセンサデータのサンプリング間隔は必ずしも一定間隔ではない。ここでは，近似的に 0.1s として FFT を行った結果を示す。

が確認できる。

出力信号の周波数特性は，伝達関数の定義から入力信号の周波数特性×フィルタの周波数特性で求められる。周波数特性における縦軸は dB（デシベル）表示であることが多く，その場合は，乗算ではなく，加算で計算できる†。これは dB の定義が対数変換された値であり，乗算が対数計算では加算となることによる。このことから，ローパスでは高い周波数成分が，ハイパスでは低い周波数成分が，そしてバンドパスでは低い周波数と高い周波数成分が除去されることが周波数特性からも確認できる。

この三つの周波数成分を含む信号をこれらのフィルタに通した結果を求める。IIR フィルタ構造に対応する C 言語プログラムを**図 6.15** に示す。INDEX_SIZE はフィルタへの入力データの要素数，FILTER_TAP はフィルタタップ数である。ループ文を用いることで，容易にフィルタを構成することができる。各フィルタ（順にローパス，ハイパス，バンドパス）を通した後の信号波形を**図 6.16** に示す。三つの周波数成分を持つ信号に対して，各パラメータを用いてその効果が実現できる。このグラフを得るための，式 (5.72) で示される入

```
//フィルタリング
for(i=0; i<INDEX_SIZE; i++){

    for(j=0; j<=FILTER_TAP; j++){
        if(i-j>=0){
            y[i] += b[j] * x[i-j];
        }
    }
    for(j=1; j<=FILTER_TAP; j++){
        if(i-j>=0){
            y[i] += -a[j] * y[i-j];
        }
    }

}
```

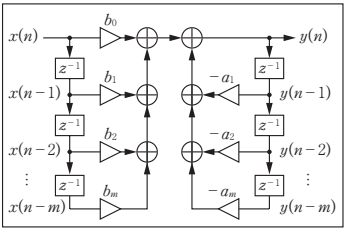

図 6.15　フィルタ構造に対応するプログラム

† $\log ab = \log a + \log b$，$\log a/b = \log a - \log b$ である。このように，乗算，除算は対数値に変換すると加算，減算となり，最後に対数値を真数に戻すことにより，途中の複雑な乗除算演算を容易にできる。

索　　引

―― 著 者 略 歴 ――

田中　博（たなか　ひろし）
1983 年　北海道大学工学部精密工学科卒業
1985 年　北海道大学大学院工学研究科博士前期課程修了（精密工学専攻）
1985 年　日本電信電話株式会社勤務
1994 年　博士（工学）（北海道大学）
1994 年～97 年　宇宙開発事業団（現宇宙航空研究開発機構）出向
2006 年　神奈川工科大学教授
現在に至る

五百蔵　重典（いおろい　しげのり）
1993 年　東京理科大学理学部応用数学科卒業
1993 年　株式会社 PFU 勤務
1996 年　北陸先端科学技術大学院大学情報科学研究科博士前期課程修了（情報システム学専攻）
1999 年　北陸先端科学技術大学院大学情報科学研究科博士後期課程修了（情報システム学専攻），博士（情報科学）
1999 年　神奈川工科大学助手
2005 年　神奈川工科大学講師
2008 年　神奈川工科大学准教授
2013 年　神奈川工科大学教授
現在に至る

IoT 時代のデータ処理の基本と実践
―スマホ内蔵センサ取得データを用いて―

Basics and Practice of the Data Handling of the IoT Era
― Using Acquisition Data from a Sensor Built-in Smartphone ―

2018 年 3 月 20 日　初版第 1 刷発行　★

検印省略

著　者　田　中　博
　　五百蔵　重　典
発行者　株式会社　コロナ社
　　代表者　牛来真也
印刷所　新日本印刷株式会社
製本所　有限会社　愛千製本所

112-0011　東京都文京区千石 4-46-10
発 行 所　株式会社　**コ ロ ナ 社**
CORONA PUBLISHING CO., LTD.
Tokyo Japan
振替 00140-8-14844・電話(03)3941-3131(代)
ホームページ http://www.coronasha.co.jp

ISBN 978-4-339-02880-5　C3055　Printed in Japan　（齋藤）