

Mapleによる統計的な推定法

- 最尤推定とモーメント法 -

数式処理システムである Maple には、統計解析のための **Statistics** パッケージが標準で組み込まれています。**Statistics** パッケージの特徴は、Maple の持つ最大の特徴であるあらゆる種類の数式モデルとその処理機能を組み合わせて、極めて自由度の高い統計解析が行えることです。

本資料では、Maple が用意しているいろいろな統計解析の中から、統計的な推定 (Estimation) として広く知られている最尤法とモーメント法を紹介します。

Maple で用意されている統計分布

Maple の **Statistics** パッケージを読み込むと、次の表に示す 30 種以上の統計分布が利用できます。

Bernoulli 分布	Beta 分布	二項分布	コーシー分布
離散一様分布	経験分布	アーラン分布	誤差分布
フィッシャーF分布	ガンマ分布	幾何分布	ガンベル分布
逆ガウス分布	ラプラス分布	ロジスティック分布	対数正規分布
Moyal 分布	負の二項分布	非心ベータ分布	非心 χ 二乗分布
非心 StudentT 分布	正規 (ガウス) 分布	パレート分布	ポアソン分布
レイリー分布	StudentT 分布	三角分布	一様分布
χ 二乗分布	指数分布	超幾何分布	von Mises 分布
マクスウェル分布	非心F分布	ベキ分布	ワイブル分布

Table 1: 標準で利用可能な統計分布

各統計分布の詳細は、次のコマンドを実行して表示されるヘルプページから参照可能です。

```
[> help("Statistics,Distributions");
```

また、既存の分布やユーザ独自の関数からカスタムの分布を作ることも可能です。詳しくは[こちらのヘルプページ](#)を参照してみてください。

最尤推定

まず、統計パッケージ **Statistics** を現在のセッションに読み込みます。

```
> with(Statistics)
[AbsoluteDeviation, AgglomeratedPlot, AreaChart, AutoCorrelation, BarChart,
Bootstrap, BoxPlot, BubblePlot, CDF, CGF, CentralMoment,
CharacteristicFunction, ChiSquareGoodnessOfFitTest, ChiSquareIndependenceTest,
ChiSquareSuitableModelTest, ColumnGraph, Correlation, CorrelationMatrix,
(2.1)
```

Count, CountMissing, Covariance, CovarianceMatrix, CrossCorrelation, Cumulant, CumulantGeneratingFunction, CumulativeDistributionFunction, CumulativeProduct, CumulativeSum, CumulativeSumChart, DataSummary, Decile, DensityPlot, DiscreteValueMap, Discretize, Distribution, ErrorPlot, EvaluateToFloat, Excise, ExpectedValue, ExponentialFit, ExponentialSmoothing, FailureRate, FisherInformation, Fit, FivePointSummary, FrequencyPlot, FrequencyTable, GeometricMean, HarmonicMean, HazardRate, Histogram, Information, InteractiveDataAnalysis, InterquartileRange, InverseSurvivalFunction, Join, KernelDensity, KernelDensityPlot, KernelDensitySample, Kurtosis, Likelihood, LikelihoodRatioStatistic, LineChart, LinearFilter, LinearFit, LogLikelihood, LogarithmicFit, MGF, MLE, MakeProcedure, MaximumLikelihoodEstimate, Mean, MeanDeviation, Median, MedianDeviation, MillsRatio, Mode, Moment, MomentGeneratingFunction, MovingAverage, MovingMedian, MovingStatistic, NonlinearFit, NormalPlot, OneSampleChiSquareTest, OneSampleTTest, OneSampleZTest, OneWayANOVA, OrderByRank, OrderStatistic, PDF, Percentile, PieChart, PointPlot, PolynomialFit, PowerFit, Probability, ProbabilityDensityFunction, ProbabilityFunction, ProbabilityPlot, ProfileLikelihood, ProfileLogLikelihood, QuadraticMean, Quantile, QuantilePlot, Quartile, RandomVariable, Range, Rank, Remove, RemoveInRange, RemoveNonNumeric, Sample, ScatterPlot, ScatterPlot3D, Score, Select, SelectInRange, SelectNonNumeric, ShapiroWilkWTest, Shuffle, Skewness, Sort, SplitByColumn, StandardDeviation, StandardError, StandardizedMoment, SunflowerPlot, Support, SurfacePlot, SurvivalFunction, SymmetryPlot, Tally, TallyInto, Trim, TrimmedMean, TwoSampleFTest, TwoSamplePairedTTest, TwoSampleTTest, TwoSampleZTest, Variance, Variation, WeightedMovingAverage, Winsorize, WinsorizedMean]

この章では、最尤法を用いて統計パラメータの推定を行います。いま、例としてパレート分布に従う 3000 個のランダムな数値データを変数 S に用意します。

```
> S := Sample(Pareto(5.0, 7.0), 3000)
```

$$S := \begin{bmatrix} 1 \dots 3000 \text{ Vector}_{\text{row}} \\ \text{Data Type: float}_8 \\ \text{Storage: rectangular} \\ \text{Order: Fortran_order} \end{bmatrix} \quad (2.2)$$

```
> S[1..3]
```

$$\left[6.36161176160045 \quad 7.00694729625251 \quad 5.09795037995579 \right] \quad (2.3)$$

なお、パレート分布の確率密度関数は次で与えられます：

```
> X := RandomVariable(Pareto(a, c)) :
```

```
> PDF(X, u)
```

(2.4)

$$f(u) = \begin{cases} 0 & u < a \\ \frac{c a^c}{u^{c+1}} & \text{otherwise} \end{cases} \quad (2.4)$$

パレート分布の平均と分散は次の通りです：

$$\begin{aligned} &> \text{Mean}(X), \text{Variance}(X) \\ &\left[\begin{array}{l} \left\{ \begin{array}{ll} \text{undefined} & c \leq 1 \\ \frac{c a}{-1 + c} & \text{otherwise} \end{array} \right\}, \left\{ \begin{array}{ll} \text{undefined} & c \leq 2 \\ \frac{c a^2}{(-1 + c)^2 (-2 + c)} & \text{otherwise} \end{array} \right\} \end{array} \right] \end{aligned} \quad (2.5)$$

変数 S の割当て部分にあるように、今回パレート分布の位置 (location) 及び形状 (shape) パラメータはそれぞれ $a=5.0, c=7.0$ としています。

最尤法を用いたパラメータの推定として、ここではパレート分布の形状パラメータが未知であったと仮定します。すなわち、以下のパラメータ c を最尤法で推定します。

$$\begin{aligned} &> \text{Pareto}(5.0, c) \\ &\left[\text{Pareto}(5.0, c) \right] \end{aligned} \quad (2.6)$$

最尤法の定義から考えれば、これは観測点のデータ S に対して、 $\text{Pareto}(5.0, c)$ のスコア関数 F を求めて

$$\begin{aligned} &> F := \text{Score}(\text{Pareto}(5.0, c), S) \\ &\left[F := -429.637635901464 + \frac{3000}{c} \right] \end{aligned} \quad (2.7)$$

これを c について解くことで求めることが可能です。

$$\begin{aligned} &> \text{fsolve}(F=0, c) \\ &\left[6.982628497 \right] \end{aligned} \quad (2.8)$$

これと等価な処理が **Statistics** パッケージの **MaximumLikelihoodEstimate** 関数です。

$$\begin{aligned} &> M := \text{MaximumLikelihoodEstimate}(\text{Pareto}(5.0, c), S) \\ &\left[M := 6.98262849742118 \right] \end{aligned} \quad (2.9)$$

求めた最大尤度の信頼区間を考えるため、サンプル S に対する $\text{Pareto}(5.0, c)$ の尤度比統計量を計算します。

$$\begin{aligned} &> L := \text{LikelihoodRatioStatistic}(\text{Pareto}(5.0, c), S) \\ &\left[L := 5660.552525 - 6000. \ln(c) + 859.2752718 c \right] \end{aligned} \quad (2.10)$$

一般に、尤度比の統計量は、一つの確率変数に対する χ^2 乗分布に従うことにより、例えば95%の信頼区間はその分位数を計算し次の手順で求められます。

```
[ > t := Quantile(ChiSquare(1), 0.95)
      t := 3.84145606580278 (2.11)
```

従って、この最尤推定による約 95%の信頼区間は次のように計算できます。

```
[ > CI := fsolve(L=t, c=-∞..M) ..fsolve(L=t, c=M..∞)
      CI := 6.735734364 ..7.235483321 (2.12)
```

今回、本資料ではパレート分布による最尤推定を行いました。Maple の持つ様々な統計分布に対しても同様な手順で推定を行うことが可能です。

モーメント法によるパラメータ推定

前章では最尤法を使ったパラメータの推定を行いました。ここではモーメント（積率）を使った推定を行ってみます。

ここでも、前回同様にある適当な分布に従ったランダムデータを用意します。

```
[ > restart
  > with(Statistics)
  [AbsoluteDeviation, AgglomeratedPlot, AreaChart, AutoCorrelation, BarChart,
    Bootstrap, BoxPlot, BubblePlot, CDF, CGF, CentralMoment,
    CharacteristicFunction, ChiSquareGoodnessOfFitTest, ChiSquareIndependenceTest,
    ChiSquareSuitableModelTest, ColumnGraph, Correlation, CorrelationMatrix,
    Count, CountMissing, Covariance, CovarianceMatrix, CrossCorrelation, Cumulant,
    CumulantGeneratingFunction, CumulativeDistributionFunction,
    CumulativeProduct, CumulativeSum, CumulativeSumChart, DataSummary, Decile,
    DensityPlot, DiscreteValueMap, Discretize, Distribution, ErrorPlot,
    EvaluateToFloat, Excise, ExpectedValue, ExponentialFit, ExponentialSmoothing,
    FailureRate, FisherInformation, Fit, FivePointSummary, FrequencyPlot,
    FrequencyTable, GeometricMean, HarmonicMean, HazardRate, Histogram,
    Information, InteractiveDataAnalysis, InterquartileRange, InverseSurvivalFunction,
    Join, KernelDensity, KernelDensityPlot, KernelDensitySample, Kurtosis, Likelihood,
    LikelihoodRatioStatistic, LineChart, LinearFilter, LinearFit, LogLikelihood,
    LogarithmicFit, MGF, MLE, MakeProcedure, MaximumLikelihoodEstimate, Mean,
    MeanDeviation, Median, MedianDeviation, MillsRatio, Mode, Moment,
    MomentGeneratingFunction, MovingAverage, MovingMedian, MovingStatistic,
    NonlinearFit, NormalPlot, OneSampleChiSquareTest, OneSampleTTest,
    OneSampleZTest, OneWayANOVA, OrderByRank, OrderStatistic, PDF, Percentile,
    PieChart, PointPlot, PolynomialFit, PowerFit, Probability,
    ProbabilityDensityFunction, ProbabilityFunction, ProbabilityPlot,
    ProfileLikelihood, ProfileLogLikelihood, QuadraticMean, Quantile, QuantilePlot,
    Quartile, RandomVariable, Range, Rank, Remove, RemoveInRange,
    RemoveNonNumeric, Sample, ScatterPlot, ScatterPlot3D, Score, Select,
    SelectInRange, SelectNonNumeric, ShapiroWilkWTest, Shuffle, Skewness, Sort,
    SplitByColumn, StandardDeviation, StandardError, StandardizedMoment,
    SunflowerPlot, Support, SurfacePlot, SurvivalFunction, SymmetryPlot, Tally,
```

(3.1)

