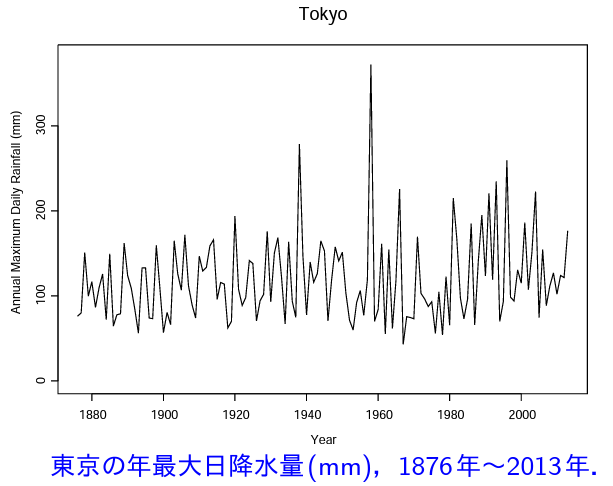
#### 極値統計学

# 極値統計学の目的

下図の降水量データの様に，与えられたデータから（与えられた空間や時間の中で）『どの様な大きな値がどれくらいの確率で出現するのか？』を知りたい．そのためには『極値データの確率構造』を明らかにしないといけない．適切な統計モデルを作成しデータ解析を行う．



上記の目的を達成するための統計学が『**極値統計学**』となる．大きな値の出現に対して情報を持っている極値データのみを考える．データに適合させる分布は，**一般極値分布**と**一般パレート分布**の2種類が一般的である．

参考文献：[極値統計学（高橋倫也）](chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/viewer.html?pdfurl=http%3A%2F%2Fcoop-math.ism.ac.jp%2Ffiles%2F129%2F03Takahashi.pdf&clen=992870&chunk=true)

|  |  |
| --- | --- |
| **極値統計学** | **数理統計学** |
| データの端の推測 | データの中心（平均）の推測 |

データのデザイン方法としては，3つの方法が考えられる．

|  |  |  |
| --- | --- | --- |
| **データの形式** | **説明** | **モデル** |
| ブロック最大値データ | 全データを群に分け各群の最大値 | 一般極値分布 |
| 上位個データ | 全データを群に分け各群の上位個 | 極値上位個同時分布 |
| 閾値超過データ | 全データのうちある閾値の超過分 | 一般パレート分布  点過程モデル（ポアソン過程） |
| 時系列データ | 時刻に従い記録されたデータ |  |

【極値統計量の基準化と極値分布】

|  |
| --- |
| を基準化：数列 と退化していない分布 を持つ確率変数が存在して，  すなわち  ：極値分布(extreme value distribution)  分布は極値分布の値吸引領域に属する：．  ()：基準化定数 ． |

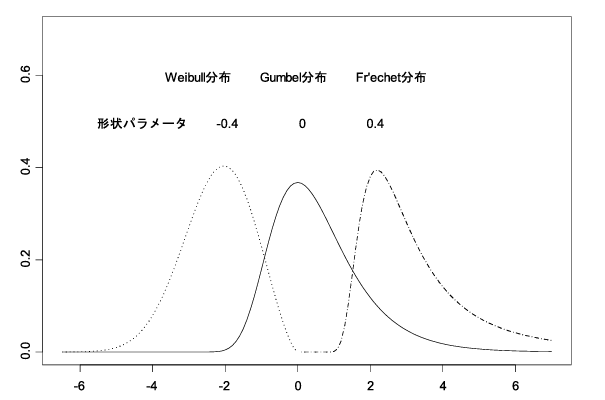
## ブロック最大値データ

【一般極値分布（定義4.6）】

|  |
| --- |
| 定数，，に対し，分布関数  を持つ分布を**一般極値分布**（Generalize extreme value distribution）と言い，**GEV()**と表す． |

【一般極値分布GEV()の3つの型について】

|  |
| --- |
| のとき，ワイブル分布  のとき，グンベル分布  のとき，ワイブル分布 |



母集団分布が適当な条件を満たし，独立同分布なデータが十分多い，つまりブロックの大きさが十分大きいとき，

とおくと

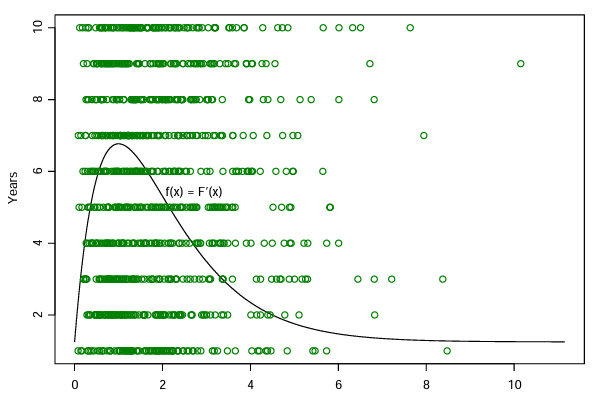
ただし，

結果として，が十分大きいときの最大値データは漸近的に極値分布に従うとみなすことができる．

また，一般極値分布の密度関数は，

【最尤法によるの推定】

ブロックに分け，各ブロックの要素数をとする



の実現値がとしたとき，尤度関数は

対数尤度は，のとき

|  |
| --- |
| ならば，  となるがあれば， |

のとき，対数尤度は

|  |
| --- |
|  |

解析的に解くことができないため，数値計算法を用いる．実際に母数を推定する手順は以下の通り．

|  |  |
| --- | --- |
| **手順** | **方法** |
| １ | ブロック数を決め，全データを個のブロックに等分する．一部で欠損値が出ても構わない． |
| ２ | 各ブロック内最大値を求める． |
| ３ | データ()から母数を最尤法で推定する（数値計算法）． |

参考文献：

[Numerical Methods of Statistics（John F. Monahan, North Carolina State University）](https://www.cambridge.org/core/books/numerical-methods-of-statistics/ED2D1845F52AF845CCF560E3526B9B56)

ブロック数をどうやって決めるか…

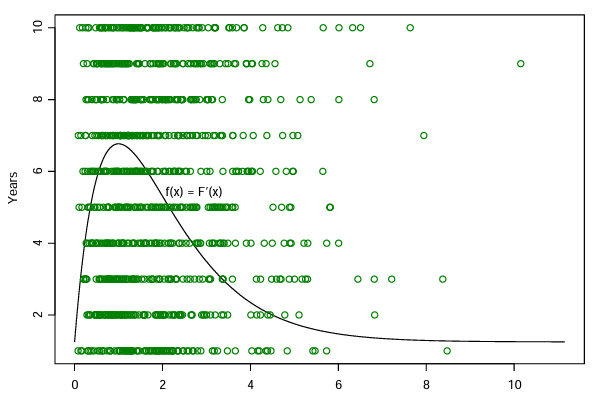
|  |  |  |
| --- | --- | --- |
| ブロック数を大きくする | ➡ | ブロック内のデータサイズは小さくなるため，収束が悪くなる．  （極値統計学は「極限定理」が重要！！） |
| ブロック数を小さくする | ➡ | ブロック内のデータ数は大きくなるが，が小さいために母数を最尤法で推定するためのデータサイズが小さく，推定量の分散が大きくなる． |

表にまとめると以下の通り．

|  |  |  |
| --- | --- | --- |
| **ブロック数** | **収束** | **推定量分散** |
| 小 | ◎ | × |
| 大 | × | ◎ |

## 上位個データ

各ブロックに分け，ブロック内の上位個データを利用



一般極値分布について，に従う独立同分布の順序統計量を以下の通りにする．

何をやっているかというと➡各ブロックを値が大きいものから順に振り替えている．

通常の極限定理により，（吸引係数）が存在し，

【上位個データにおける同時分布関数】

|  |
| --- |
| 上位個の順序統計量を基準化した同時分布関数の極限に対し， |

【上位個データにおける同時密度関数】

|  |
| --- |
|  |

【最尤法によるの推定】

|  |  |
| --- | --- |
| **手順** | **方法** |
| １ | ブロックの要素を大きさの順に並び替え➡ |
| ２ | これらの実現値をベクトルで表現．， |
| ３ | サンプルが十分大きいとき極限定理より，上位個データは漸近的に極限分布に従うとみなすことができるので，尤度関数は |
| 4 | 対数尤度を取り，数値計算で最大となる母数（）を推定する． |

## 閾値超過データ