

How to Check if Time Series Data is Stationary with Python

<https://machinelearningmastery.com/time-series-data-stationary-python/>

定常時系列 (Stationary Time Series)

- 定常時系列の観測地は、時間に依存しない。
- トレンドまたは季節効果を持たないなら、その時系列は定常である。
- 要約統計量 (平均, 分散) はい時間によらず一定である。
- 定常時系列はモデル化が容易. 統計的モデリングは時系列の定常性を 仮定する必要がある。

非定常過程 (Non-Stationary Time Series)

- 非定常過程は季節性, トレンド, その他時間に依存する構造を含んでいる。
- 要約統計量は時間の経過とともに変化する。
- 古典的な時系列分析・予測手法は、トレンドと季節性を除去することで非定常な 時系列データを定常化する。

定常性の種類

非定常過程の例

- **トレンド定常過程** : 定常過程にトレンド成分がある過程
- **単位根過程** : AR(1) processにおける1次項の係数が1の過程. 要するに原系列が非定常で差分系列が定常であるような過程のこと

定常性の確認方法

- プロットを見る。
- 要約統計量を見る。
- 統計的検定 (ADF検定)

時系列を2つまたはそれ以上のパーティションに分割し、各グループの平均と分散を比較する。これらの差が有意である場合は、時系列は非定常である可能性が高い。

daily-total-female-births time seriesの場合

- 平均と分散を見るから、データがガウス分布に従うと仮定する。ヒストグラムによってチェックできる。ヒストグラムを見ると、形状はガウス分布であるが、やや右尾が長い分布であることが読み取れる。
- 時系列を2つの連続する系列に分割して平均と分散を計算する。

airline-passengers time seriesの場合

ヒストグラムは明らかに定常ではない。したがって平均値、分散値はあまり意味がないことがわかる。観測値のつづれた分布は非定常過程の一つの指標になる。

季節成分は指数関数的な成長を示唆している可能性があることから、対数変換を行ってみる。平均と分散を見

る限りでは定常であるが、折れ線グラフを見ると定常とは言えない。この例において対数変換を行った結果、要約統計量が定常に見えるが、実際には定常でない場合があるから 注意しなければならない。

Augmented Dickey-Fuller test (ADF検定, 単位根検定)

時系列が単位根で表現できる = 時系列が定常でない

→ 時系列が単位根で表現できるか、を統計的検定すれば時系列が定常でないかを確認できる。

H0 : 時系列が単位根AR(p)-processに従う。 → 非定常である
H1 : 時系列が定常なAR(p)-processに従う。 → 定常である

有意水準を5% ($\alpha = 0.05$) とすると、

p-value > 0.05 : H0を棄却し, H1を受容する。 → 非定常である
p-value ≤ 0.05 : H1を採択する。 → 定常である

daily-total-female-births time seriesの場合

検定統計量は-4.808であることがわかる。その下のルックアップテーブルを見ると、1%のとき-3.449%であるから、1%の値よりも検定統計量が小さいことがわかる。これは、1%未満の有意水準で 棄却できることを意味している。よって、時系列データが定常であることが確認できる。

ADF Statistic: -4.808291
p-value: 0.000052
Critical Values:
1%: -3.449
5%: -2.870
10%: -2.571

airline-passengers time seriesの場合

検定統計量は0.815、ルックアップテーブルは5%のとき-2.884だから 帰無仮説を採択する。よって時系列が非定常であることが確認できる。

ADF Statistic: 0.815369
p-value: 0.991880
Critical Values:
1%: -3.482
5%: -2.884
10%: -2.579