数理統計 補助資料
～尤度関数と数値的最適化～

2023年度2学期:　　月曜1限, 水曜3限
担当教員:　　　石垣　司

対数尤度関数の最大化

ロジスティック回帰モデルの尤度関数の特徴(前回の復習)

1. **対数尤度関数 $L(b|D)$ は凸関数(上に凸)であるため最適解は存在する**
2. **対数尤度関数 $L(b|D)$ を最大にする b は解析的に求めることができない**

多くの統計モデルの尤度関数は1&2の性質をもつ

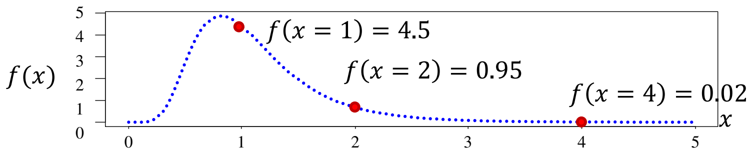
コンピュータを利用した数値的最適化が必要

– **数値的最適化: 数値計算によって最適解を見つけ出す手法の総称**

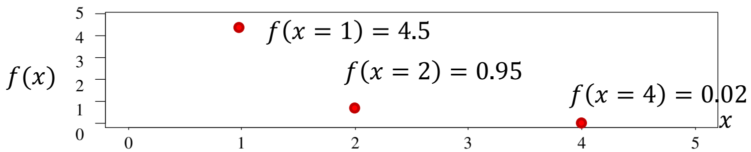
#メモ1: 経済学部生の皆さんにとっての本日の授業の意義は、統計ソフトウエアが推定結果を出力するためにどのような処理を行っているのかを理解することです。
#メモ2: 1の性質は指数型分布族という観点でまとめて議論することができます。

数値的最適化の問題設定

1. **関数 $f(x)$ の形は分からない ⇒ 青の点線の形はわからない**
2. **関数 $f(x)$ に最適解があることは分かっている ⇒ 範囲は不明**
3. **関数 $f(x)$ の導関数 $f'(x)$ は計算可能**
4. **ある値 c における関数の値 $f(x=c)$ は計算可能 ⇒ 赤点の値**



関数 $f(x)$ の形状



我々が知覚できる情報

グリッドサーチ

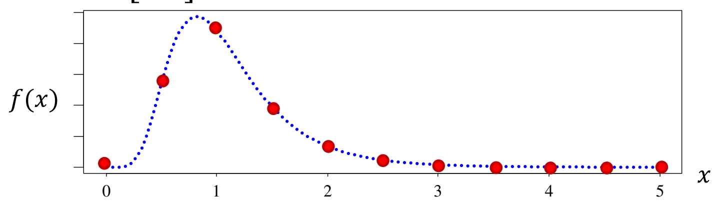
ある範囲で設定した格子点(グリッド)**の値を調べる**

– **○: 単純で直感的にも理解しやすい**
– **×: 最適解を見つけられる保証はない**
– **×: 変数の数が増えると探索点も指数的に増加**

1変数について10格子点を設定 ⇒ 10点の探索
10変数についてそれぞれ10格子点を設定 ⇒ 10^{10} 点の探索

– **前々回の授業のロジスティック回帰モデルの回帰係数ベクトル b の要素数は10**

– **例: $x=[0,5]$ の範囲で0.5刻みで10個の格子点を設定**

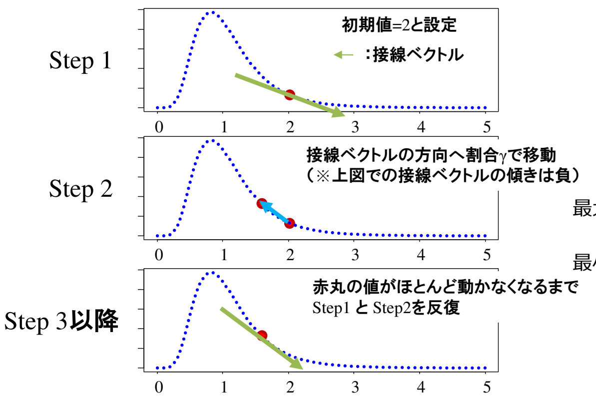


この場合、$x=1$ の値が最も大きいので、最適解として $x=1$ を採用

最急降下法(山登り法)

関数 $f(x)$ の1階導関数を利用した反復計算

- ○: 1回の計算量は少ない
- ×: 反復回数が大きくなる傾向
- ×: 初期値に反復回数が依存



【アルゴリズム】

初期設定

Set $k=0$

Set initial value of $x^{(k=0)}$

Set parameter γ

収束するまで反復計算

最大化問題

$$x^{(k+1)} = x^{(k)} + \gamma \frac{\partial f\left(x^{(k)}\right)}{\partial x^{(k)}}$$

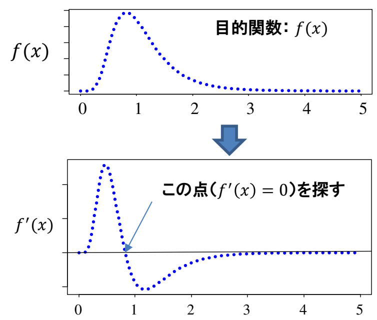
最小化問題

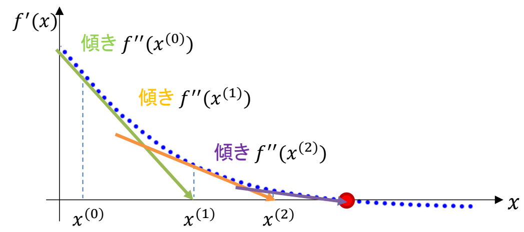
$$x^{(k+1)} = x^{(k)} - \gamma \frac{\partial f\left(x^{(k)}\right)}{\partial x^{(k)}}$$

ニュートン法(Newton-Raphson Method) #1

2階導関数を用いて $f'(x)=0$ となる点を探索

- ○: 最急降下法と比べ反復回数が少ない傾向
- ×: 初期値で解が変化。無限ループ化の可能性





更新アルゴリズム

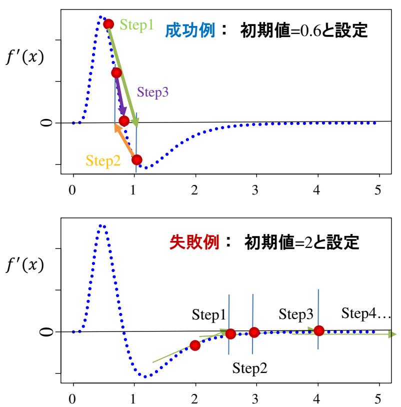
check!

$$x^{(k+1)} = x^{(k)} - \{f''(x)\}^{-1} f'(x)$$

ニュートン法(Newton-Raphson Method) #2

2階導関数を用いて $f'(x)=0$ となる点を探索

- ○: 最急降下法と比べ反復回数が少ない傾向
- ×: 初期値で解が変化。無限ループ化の可能性



【アルゴリズム】

初期設定

Set $k=0$

Set initial value of $x^{(k=0)}$

収束するまで反復計算

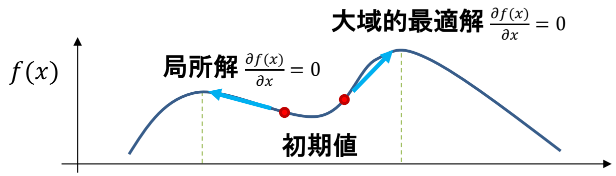
$$x^{(k+1)} = x^{(k)} - H^{-1}\frac{\partial f\left(x^{(k)}\right)}{\partial x^{(k)}}$$

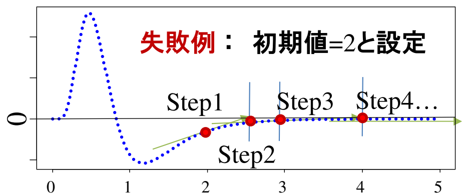
H: ヘッセ行列($f(x)$ の2階微分行列)

#メモ: 多くの場合、対数尤度関数では安定した解が得られる。ここでは、失敗例を示すためにあえて失敗しやすい関数を用いた

最急降下法・ニュートン法の弱点

1. 微分できない関数では利用できない
2. $\frac{\partial f(x)}{\partial x} = 0$ の点が複数ある場合, 解は初期値に依存
3. 関数形によっては計算が収束しないことがある



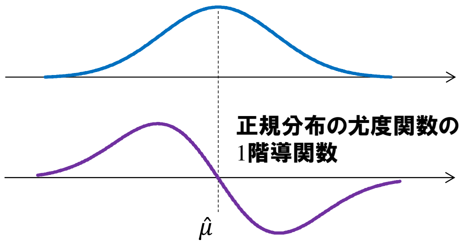


対数尤度関数の数値的最適化

凸関数とニュートン法

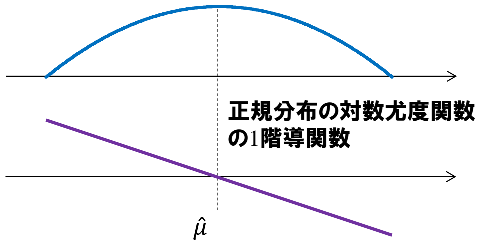
– 非凸関数の尤度関数ではニュートン法が失敗する可能性あり
– 対数尤度関数が凸関数の場合, ニュートン法で比較的簡単に最大解を見つけることが可能

正規分布の尤度関数(非凸関数)



この場合, 初期値に依存して求解に失敗

正規分布の対数尤度関数(凸関数)



この場合、1回の計算で最大解にたどり着く

演習問題

単純なグリッドサーチでは効率的な数値最適化を実現できない理由をもう少し詳しく考えてみよう。

1. ある尤度関数 $f(x)$ の最大点をグリッドサーチで探索する。ここで x は 1 から 100 の自然数を動く $(x \in \{1, \cdots, 100\})$。グリッドサーチの格子点を $x = \{10,20,30,40,50,60,70,80,90,100\}$ の 10 点に定めたとき, このグリッドサーチで探索できる点の数は, x が動く全空間(点の数)の何パーセントであるか答えなさい

2. ある尤度関数 $f(\boldsymbol{x})$ の最大点をグリッドサーチで探索する。ここで $\boldsymbol{x} = [x_1, \cdots, x_{10}]^T$ であり, 各 x_i は 1 から 100 の自然数を動く $(x_i \in \{1, \cdots, 100\})$。各 x_i に対するグリッドサーチの格子点を $x_i = \{10,20,30,40,50,60,70,80,90,100\}$ の 10 点に定めたとき, このグリッドサーチで探索できる点の数は, $\boldsymbol{x}$ が動く全空間(点の数)の何パーセントであるか答えなさい