

Hommachi.Stan / DB.Stats

2016年10月31日 第1回

みどり本 第1章・第2章

Hommachi.Stan の進め方

基礎固めのために『みどり本』を概観して、その後『StanとRでベイズ統計モデリング』を読み込んでいく。
急がず、じっくりやる。

第1章

重要だと思ったところ

観測されるデータは自然現象（や社会現象）のごく一部を切り取ったもの。そこに見られるパターンを要約・整理したり予測したりするために統計モデルが必要。統計モデルを作るのが統計モデリング。

第2章

重要だと思ったところ

データ解析で最も重要なのはデータを様々な方法で図示すること。統計モデリングの第1歩は、データがどんな確率分布から発生したかを考えること。統計モデルを推定したら、うまく図示することが重要。確率分布の選び方は次のとおり・説明したい量は離散か連続か？・説明したい量の範囲は？・説明したい量の標本分散と標本平均の関係は？

用語

確率分布：

データに見られる様々な「ばらつき」を表現する統計モデルの本質的な部品。確率変数の値とそれが出現する確率を対応させたもの。

確率変数：

個体ごとにばらつく変数。サイコロの場合、出る目が確率変数（1～6）

パラメーター：

確率分布の形をきめるもの。

ばらつきを示す標本統計量：

標本分散, 標本標準偏差

ポアソン分布：

- 確率変数は負でない整数（カウントデータ）
- パラメーターは平均 λ
- 平均と分散は等しい

最尤推定法（最尤推定）：

尤度を最大にするパラメータを推定すること。尤度とは「あてはまりの良さ」を表す統計量。パラメータの関数。尤度の実態は「あるパラメータの値を決めたときにすべての個体についての $p(y_i|\lambda)$ の積。ある事象が起こりうる確率が最大になるパラメータを見つけるため。尤度関数は計算しづらいので対数を取る。