Michael A. McCarthyの主張

生態学の解析では古典的統計学（頻度主義）よりベイズ法の方がより適切だ！！

- 1) ベイズ法は数学的な論法に矛盾がない。古典的統計学はデータに関する確率論的記述は論理的だが、仮説については論理的でない。ベイズ法は物事のある状態に対して確率論的予測を行うことができるが、古典統計学ではサンプルデータの仮定的な反復から得られる長期的な平均に関してのみ予測を行うことができる。
- 2) ベイズ法では適切な事前情報を利用してパラメータ推定が可能だが、古典的統計学ではサンプルデータに含まれる情報以外のものを利用することができない。

ベイズ法の4要素

prior(事前情報) + data(データ) -model posterior(事後情報)

「事前分布」と「データ」から予測を行うのは日常生活では当たり前のこと。朝、家を出る前にニュースで天気予報を聞く（事前情報）玄関を出て空模様を見る（データ）、ごによごによ考える（model）、「雨が降りそう」（posterior）と結論、傘を持って出る。

Example 1

ある町の池に生息するカエルの生態について研究をしている生態学者がいます。この研究者はある池を初めて訪れ、池の周りを歩き、カエルの鳴き声が聞こえないか 20分間調査しましたが、カエルの声は聞こえませんでした。研究者は経験的に、たとえカエルが実際には生息していたとしても、カエルを見つけられる確率は 80%にすぎないことを知っていたので、一回目の調査でカエルが見つからなかったことに特に驚きませんでした。この情報をもとにして、「この池にカエルがいる確率」についてどのような結論が得られるか考えてみましょう。

頻度主義的方法 1：仮説検定

帰無仮説 1：この池にカエルは生息していない

対立仮説 1；この池にカエルは生息している

p 値：帰無仮説が成立する時に、このデータ（カエルの声が聞こえない）が得られる確率
「池にカエルが生息していない」時に、「カエルの声が聞こえない」確率は 1。

p 値 = 1 により帰無仮説 1 は棄却できない。結論：カエルは生息していないっぽい…。

帰無仮説 2：この池にカエルが生息している

対立仮説 2：この池にカエルが生息していない

p 値：「この池にカエルが生息している」ときに、「カエルの声が聞こえない確率」

p 値 = 0 . 2 により帰無仮説 2 は棄却できない。結論：カエルは生息しているっぽい…。

仮説検定の欠点...

- 1) どのような帰無仮説を設定するかによって結論が異なってしまう。ここでは検出力に関して何も考慮されていないのが原因(詳細は2章で...)
- 2) p値は必ずしも帰無仮説に関する信頼できる指標とはならない。p値はデータが起こりうる頻度についての情報を与えるもので、帰無仮説が成立する確率に関して直接的な情報はもたらさない。

情報論的方法

尤度に基づいた情報論的方法は仮説検定に対する代替法の1つである。複数の仮説を立ててそのうちどれが最もらしいかを検討する。

仮説A: この池にカエルは生息していない

仮説P: この池にカエルは生息している

これらの仮説のもとで、観測データ(カエルの声が聞こえない)が得られる確率は?

仮説Aの場合: 「池にカエルが生息しない」時、「カエルの声が聞こえない」確率

$$\Pr(D|H_A) = 1$$

仮説Bの場合: 「池にカエルが生息している」時「カエルの声が聞こえない」確率

$$\Pr(D|H_B) = 0.2$$

仮説Aと仮説Bの比較: $\Pr(D|H_B) / \Pr(D|H_A) = 0.2/1 = 1/5$

結論: この池にカエルが生息している確率は生息していない確率よりもかなり低い。

ベイズ推定法

H_A : この池にカエルは生息していない

H_B : この池にカエルは生息している

「カエルが生息していない」確率: $\Pr(H_A) = 0.5$

「カエルが生息している」確率: $\Pr(H_B) = 0.5$

「池にカエルが生息しない」時、「カエルの声が聞こえない」確率: $\Pr(D|H_A) = 1$

「池にカエルが生息している」時「カエルの声が聞こえない」確率: $\Pr(D|H_B) = 0.2$

観測データ(カエルの声が聞こえない)が得られる確率:

$$\Pr(D) = \Pr(H_A) * \Pr(D|H_A) + \Pr(H_B) * \Pr(D|H_B) = 0.5 * 1 + 0.5 * 0.2 = 0.6$$

観測データ(カエルの声が聞こえない)が得られた時、 H_B が成立する確率:

$$\Pr(H_B|D) = \Pr(H_B) * \Pr(D|H_B) / \Pr(D) = 0.5 * 0.2 / 0.6 = 1/6$$

観測データ(カエルの声が聞こえない)が得られた時、 H_A が成立する確率:

$$\Pr(H_A|D) = \Pr(H_A) * \Pr(D|H_A) / \Pr(D) = 0.5 * 1.0 / 0.6 = 5/6$$

$\Pr(H_a) = \Pr(H_b) = 0.5$ の時は尤度に基づく方法と同じ結果になる。
WinBUGSを使って解いてみよう...

```
box1.5
model
{
  prior <- 0.5
  present ~ dbern(prior)
  prob_detect <- 0.8 * present
  detected ~ dbern(prob_detect)
}
```

Example2

ユーカリの残存林の樹木直径を推定してみよう…。連続データの場合のベイズ推定では仮説が無限に考えられるが、ベイズ推定では変数の不確実性を連続的な分布で表すことでこの問題を克服できる(らしい)。計算は大変なので、WinBUGSにお任せ…。

10本の木をランダムに選んで直径を測定。すでに他の残存林の43サイトで計測した結果(平均 53cm 標準偏差 5cm) が得られているので、これを事前分布として利用する。データの精度として標準偏差を利用するが、WinBUGSでは1/分散 ($1/\text{sd}^2$) を利用する。

頻度主義の結果：平均値 56.3cm 信頼区間 47.9-64.3cm ($56.3 \pm 1.96 * 4.3$)

ベイズ推定の結果：平均値 54.9cm 信頼区間 48.5-61.3cm

(無情報事前分布(分散のととも大きい正規分布)にすると、頻度主義と同じ結果になる。)

box1.8

model

```
{  
  m ~ dnorm(53, 0.04)  
  stdev <- sd(Y[])  
  prec <- 1/(stdev*stdev)  
  for(i in 1:10)  
  {  
    Y[i] ~ dnorm(m, prec)  
  }  
}
```