Uogólnione modele liniowe

Laboratorium nr 1

1.1 Rozważamy dwa zbiory danych pochodzących z rozkładu dwumianowego, z $n = 25$ i $y = 10$ oraz $n = 50$ i $y = 20$ odpowiednio (y oznacza liczbę sukcesów).

(a) Narysować na jednym wykresie znormalizowane log-wiarogodności (jako funkcje od p, czyli prawdopodobieństwa sukcesu) odpowiadające badanym zbiorom (normalizacja ma polegać na odjęciu od funkcji log-wiarogodności tej samej funkcji, ale obliczonej w wartości estymatora NW). Zauważyć różnicę w krzywiźnie związaną z liczbą obserwacji w zbiorze. Jaka jest interpretacja statystyczna krzywizny funkcji log-wiarogodności?

(b) Użyć instrukcji `nlm` do numerycznego znalezienia maksimum (nieznormalizowanej) funkcji log-wiarogodności. (Uwaga: `nlm` znajduje minima, a nie maksima funkcji.) Eksperymentować z różnymi wartościami początkowymi.

(c) Porównać wartość odwrotności hesjanu w maksimum z wartością obserwowanej informacji w tym punkcie i ze standardowym estymatorem wariancji.

1.2 Dopasować do danych ze zbioru **bliss**

conc	dead	number
0	2	30
1	8	30
2	15	30
3	23	30
4	27	30

model logistyczny $y \sim$ conc. Jak zmienia się szansa zgonu przy zwiększeniu conc o 1?

1.3 Utworzyć rozwiniętą kopię zbioru **bliss** (w postaci danych niegrupowanych), na przykład za pomocą instrukcji `rep`. Dopasować model logistyczny. Porównać współczynniki z uzyskanymi w punkcie 1.

1.4 Zaprogramować procedurę iteracyjnej estymacji parametrów na przykładzie zbioru **bliss**. Wypisać kolejne przybliżenia parametrów β_1, β_2, dopasowane wartości $\hat{y}_i = n_i\hat{\pi}_i$, kolejne przybliżenia macierzy informacji Fishera oraz dewiacji dla kilku pierwszych iteracji. Eksperymentować z różnymi wartościami początkowymi.