

# バイオインフォマティクスの基礎 2

慶應義塾大学富田研究室

平成 21 年 5 月 4 日



# 目次

第 1 章	ベイズの定理	5
1.1	はじめに . . . . .	5
1.2	タンパク質間相互作用予測への応用例 . . . . .	5



# 第1章 ベイズの定理

2007/5/9 斎藤輪太郎

## 1.1 はじめに

本章では、ベイズの定理を用いた応用について述べます。

## 1.2 タンパク質間相互作用予測への応用例

それではベイズの定理を応用して、タンパク質間相互作用の予測を試みましょう。Iをタンパク質間相互作用を表す記号とします。I = P なら相互作用するという事実を、I =  $\neg$ P なら相互作用しないという事実を表します。また、 $P(I = P)$  は与えられたタンパク質の組が相互作用する確率を表します。 $P(I = P)$  と  $P(I = \neg P)$  をまとめて、 $P(I)$  と表します。

次に  $E_1, E_2, E_3, \dots$  で相互作用予測の根拠となる事象を表します。例えば  $E_1$  が遺伝子発現相関を表すとすれば、 $E_1 = 0.7$  は遺伝子発現相関が 0.7 であるという事象を表し、 $P(0.7 \leq E_1 < 0.8)$  は遺伝子発現相関が 0.7 以上 0.8 未満となる確率を表します。 $E_1$  に設定可能な条件式は様々で、それに対する確率も様々になります。それらの確率をまとめて、 $P(E_1)$  と表します。

$$P(I|E) = P(E|I)P(I)/P(E) \quad (1.1)$$

このとき、 $P(I = G_P|E)$  では正しい計算になりません。何故なら、 $G_P$  以外にも PPI はたくさんあるからです。

また評価するためには、正解率を出さなければならない。すなわち  $G_P$  だけでなく、 $G_N$  も必要となる。リスクを最低にすることを考えると、ベイズ判別を用いて、

$$\frac{P(I|E)}{P(\neg I|E)} = \frac{P(E|I)P(I)}{P(E|\neg I)P(\neg I)} \quad (1.2)$$

但し  $P(E|I = P)$  は  $P(E|G_P)$  で、 $P(E|\neg I)$  は  $P(E|G_N)$  で代用。