

情報生物学 (1 / 1)

[1] 二つの遺伝子の発現量 X と Y の間に次の関係があると仮定する。

$$Y = aX + b + Z$$

ここで a, b は未知の定数で、 Z は環境等の効果により確率的に変動する量 (確率変数) でその平均値は 0 ($E[Z]=0$ 、ここで $E[W]$ は確率変数 W の平均値を表す) とする。この二遺伝子の発現量を n 個の個体で調べ、データ (X_i, Y_i) , ($i=1, \dots, n$) を得た。このデータから a, b を推定したい。次の問いに答えなさい。(計 50 点)

- (1) $Q(a, b) = \sum_{i=1}^n (Y_i - aX_i - b)^2$ を最小にするような a, b の推定値 \hat{a}, \hat{b} を求めなさい。(15 点)
- (2) \hat{a}, \hat{b} が不偏推定量 (unbiased estimator、 $E[\hat{a}] = a, E[\hat{b}] = b$) であることを示しなさい。(15 点)
- (3) 次のような性質を満たす a の推定量を線形不偏推定量と呼ぶ。

$$\tilde{a} = \sum_{i=1}^n c_i Y_i, \quad E[\tilde{a}] = a \quad (\text{ここで } c_i \text{ は } X_j (j=1, \dots, n) \text{ 等によって決まる定数})$$

任意の線形不偏推定量 \tilde{a} について、次の式が成り立つことを示しなさい。(10 点)

$$\text{Cov}(\tilde{a} - \hat{a}, \hat{a}) = 0 \quad (\text{Cov}(X, Y) \text{ は } X, Y \text{ の共分散を表す。})$$

- (4) (3) の結果を利用して、 \hat{a} が線形不偏推定量の中で分散 ($\text{Var}(\tilde{a})$) を最小にする推定量であることを示しなさい。(10 点)

[2] 系統樹に関する次の問いに答えなさい。(計 50 点)

それぞれの分岐点で二つの枝 (branch) に分岐する系統樹を考える。共通祖先の場所が決まっていない系統樹を無根系統樹 (unrooted tree、図 1)、決まっている系統樹を有根系統樹 (rooted tree、図 2) と呼ぶ。

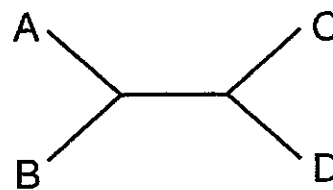


図1 4種の無根系統樹

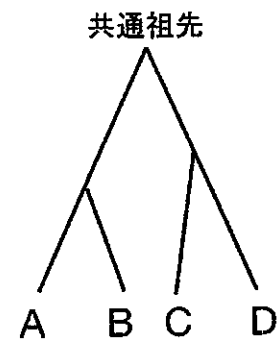


図2 4種の有根系統樹

- (1) 無根系統樹で表される系統関係が何種類あるかを知

りたい。種数が 3 および 4 の場合にその数を答えなさい。(10 点)

- (2) n 種 (n は任意の自然数) の無根系統樹の枝 (branch) の総数を求めなさい。(10 点)

- (3) n 種の無根系統樹は何種類あるか、その数を求めなさい。また n 種の有根系統樹は何種類あるか、その数を答えなさい。(10 点)

- (4) 塩基配列等を使って推定される系統樹 (分子系統樹) の信頼性を測る時に使われるブートストラップ法とはどのような方法か。図等 (右図参照) も使用して簡単に説明しなさい。但し L は塩基数を表す。(20 点)

種 1	ATCCGGGAAG
種 2	ACCCGGGGAG
種 3	ATCGGGAGTG
種 4	ATCCGGGATG