

■ 目 次

■ シリーズの刊行にあたって	iii
■ まえがき	v
Chapter 1 第 1 章 計算機科学者のための生命科学入門	1
1.1 生命に流れる情報	1
1.2 親から子へと受け継がれる情報	6
1.3 遺伝子の発現	9
1.4 遺伝子発現量の制御	13
1.5 ノンコーディング RNA	15
1.6 タンパク質	17
1.7 パスウェイ	18
1.8 生命科学という分野の特徴	19
Chapter 2 第 2 章 多重検定と無限次数多重検定法	21
2.1 仮説検定	22
2.1.1 仮説検定と分割表	22
2.1.2 フィッシャーの正確確率検定	24
2.1.3 カイ 2 乗検定	26
2.1.4 Mann-Whitney U 検定	28
2.1.5 モンテカルロ検定	30
2.1.6 偽陽性と偽陰性	31
2.2 多重検定	32
2.2.1 多重検定問題	32
2.2.2 有意水準の補正	34
2.2.3 Bonferroni 法	35
2.2.4 ステップワイズ法 (Holm 法と Hochberg 法)	37
2.2.5 Tarone 法	41

2.2.6	リサンプリング法による近似的な制御	48
2.2.7	Benjamini-Hochberg 法	51
2.2.8	分布推定による制御	55
2.3	無限次数多重検定法	57
2.3.1	頻出パターン列挙	57
2.3.2	最小サポートと P 値の下限	62
2.3.3	最小サポートと FWER	64
2.3.4	探索アルゴリズム	66
2.3.5	飽和アイテム集合を利用した補正	69
2.3.6	深さ優先探索による高速化	70
2.3.7	リサンプリング法の利用	74
2.4	生命情報における応用	76
2.4.1	転写制御因子	77
2.4.2	遺伝子発現変動	80
2.4.3	遺伝子群に対する機能解析	80
2.4.4	脳機能解析	84

第 3 章	推定量設計の理論と方法	85
3.1	バイオインフォマティクスにおける推定問題	85
3.2	記法, 記号	88
3.3	本章で取り扱う推定問題の定式化	90
3.3.1	一般形	90
3.3.2	ヘアワイズアライメント	90
3.3.3	RNA の 2 次構造予測	93
3.3.4	系統樹トポロジーの推定	95
3.4	解空間の大きさ	98
3.5	スコアの導入 — スコアモデル —	103
3.5.1	ヘアワイズアライメントのスコア	103
3.5.2	RNA の 2 次構造のスコア	105
3.6	最適解の導出	105
3.6.1	ヘアワイズアライメントの最適解	106
3.6.2	RNA の 2 次構造予測の最適解	113

3.7	スコアモデルから確率モデルへ	117
3.8	分配関数の計算方法	121
3.8.1	ヘアワイズアラインメントの分配関数	121
3.8.2	RNA 2 次構造の分配関数	123
3.9	周辺化確率	124
3.9.1	整列確率	124
3.9.2	塩基対確率	128
3.9.3	葉分割確率	128
3.10	推定方法設計の理論と方法	129
3.10.1	最尤推定とその問題点	129
3.10.2	期待利益最大化推定	130
3.10.3	利益関数と評価指標——期待精度最大化推定——	132
3.11	期待精度最大化推定の適用例	138
3.11.1	γ セントロイド RNA 2 次構造予測	138
3.11.2	γ セントロイドヘアワイズアラインメント	142
3.11.3	γ セントロイド系統樹トポロジー	143
3.12	解の不確実性とその対処法	144
3.12.1	周辺化確率の利用	146
3.12.2	確率的サンプリング	148
3.12.3	分布の可視化	149
3.12.4	クレジビリティリミット	150
3.12.5	整数スコアの分布の厳密導出手法	151
3.13	3 章のまとめ	152
付録 A	進んだ話題についての補足説明	155
A.1	任意のギャップコストの場合の動的計画法	155
A.2	局所アラインメント	156
A.3	RNA 2 次構造のエネルギーモデルと McCaskill のアルゴリズム	158
A.3.1	エネルギーモデル	158
A.3.2	McCaskill のアルゴリズム	161
A.4	バイナリ空間上の点推定問題に対する評価指標	165

■ A.5 命題 3.10 の証明	165
■ A.6 系統樹推定に関する補足	168
■ 参考文献	169
■ 索引	172