

確率モデルとゲノム多様性解析

間野 修平 数理・推論研究系 准教授

1 粒子系の確率モデル

粒子系の確率モデルの確率解析において、マルコフ過程の双対性が有用であることが知られています。粒子系の確率モデルによって生成されるデータは一般に複雑な構造を持っていますが、ある種の問題については、双対性が母集団と標本の従うマルコフ過程を関係づけていて、非常に有用であることが分かってきました。最近の進捗は次の通りです。

1.1 染色体の2座位の遺伝モデルにおける双対性

染色体の2座位の遺伝モデルとしてのWright-Fisher拡散は40年以上前に提案された基本的なモデルですが、同時確率密度が求まっていないなど、性質がよくわかっていません。以前、母集団の時間発展を表す拡散過程とAncestral Recombination Graph(ARG)とよばれる標本を過去に遡った系統関係を表現するランダムグラフを生成するマルコフ過程が双対であることを見出しました。全てのモーメントを求めれば密度が求まるのですが、モーメントの従う漸化式をARGにより表現し、系統的に求める手法を見出しました。ARGのいくつかの性質も調べました。モーメントを特殊関数により展開することで、一部の結果は解析的に表現できました。

1.2 配列の複写を伴う遺伝モデルにおける双対性

遺伝子には複数の遺伝子からなるものが多く、遺伝子変換とよばれるメンバー間で配列の複写により一様化する現象があります。遺伝子変換にはATがGCに移される傾向が逆よりも高い偏りがあります。有利な自然淘汰の働く遺伝子は塩基置換が速いのですが、ゲノムからそのような遺伝子を探すとほとんどは淘汰ではなく遺伝子変換の偏りの結果であるという警戒すべき結果が報告されました。この現象を拡散過程としてモデル化し、双対になる標本が従うランダムグラフを導き、インポートサンプリングとパーフェクトシミュレーションを用いて尤度を計算する方法を考えました。

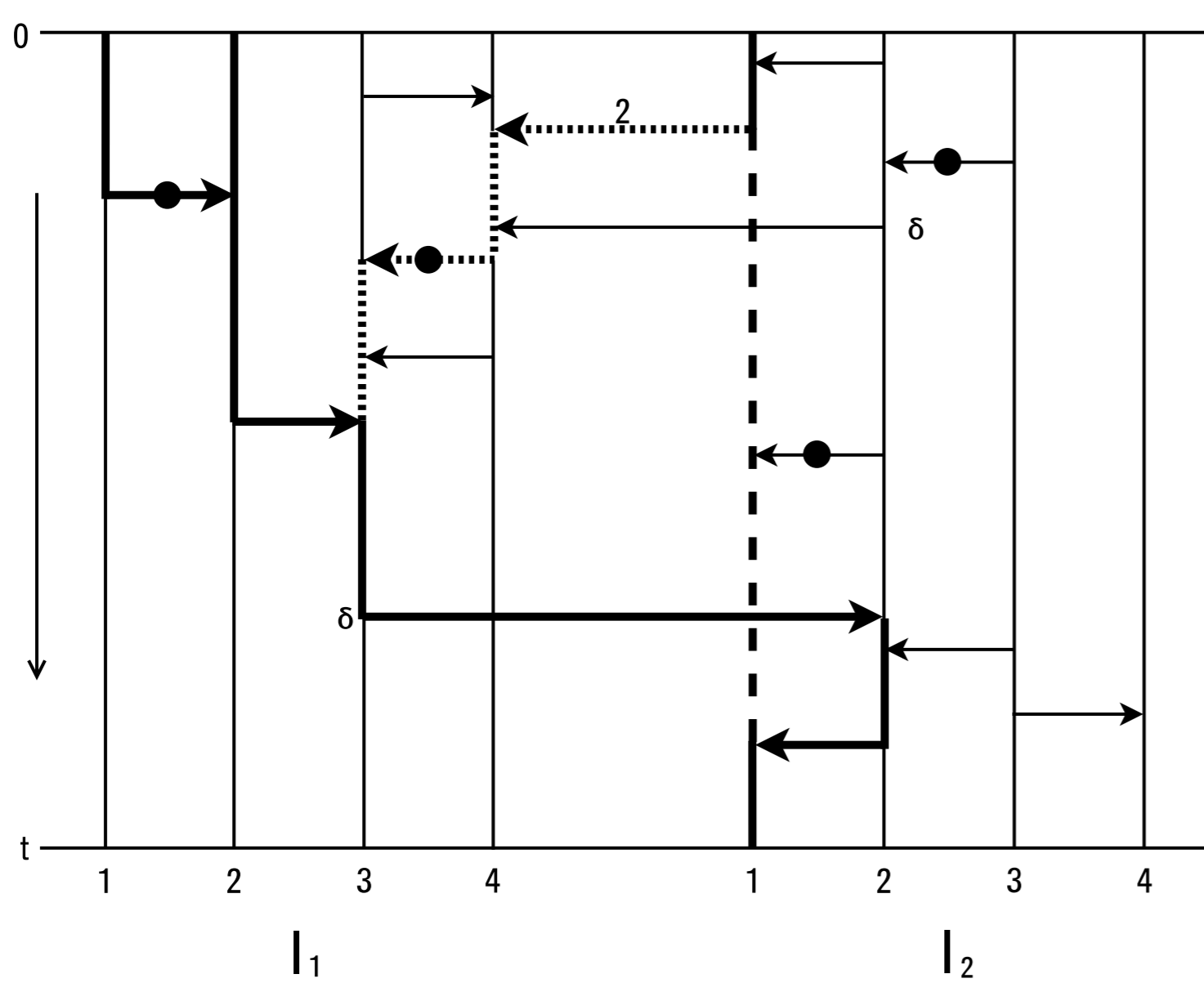


図. 完全グラフをなすノードがあり、各ノードは完全グラフをなすサイトからなる。異なるノードのサイト間の置換に偏りがある投票者モデル。

拡散とサイズ過程はモーメント双対: $\mathbb{E}_{\mathbf{p}}[\prod_{i=1}^d (X_i(t))^{n_i}] = \mathbb{E}_{\mathbf{n}}[\prod_{i=1}^d p_i^{A_{n,i}(t)}]$.

1.3 ある投票者モデルにおける双対性

伊藤名誉教授らは、グラフ上に分布する粒子があり、無作為に2つ抽出した粒子が近接ノードに由来すれば一方を他方のノードに移す、という投票者モデルを提案しました。あるグラフの拡散極限で無限個のマルチ

ンゲールがあり、それを用いて吸収確率を求めています。この粒子系の双対になる標本が従うランダムグラフを導き、考えています。

2 ゲノム多様性と生物多様性

人類学など生物多様性と進化にかかわる問題をゲノム多様性解析によって統計学的に明らかにすることを目指しています。

2.1 日本人成立の二重構造

日本人が日本列島と大陸が陸続きであった2万年前頃に移入してきた縄文人と、4世紀頃までの千年間に移入してきた弥生人の混合として成立したという仮説は良く知られています。弥生人が現在の大陸の民族と同一起源であることは疑いありませんが、縄文人の起源はわかっていません。日本人と漢民族について、全ゲノムの400万か所を超える一塩基の個人差が取得されています。そこで、弥生人を漢民族と仮定し、縄文人と弥生人のゲノムが混合して日本人のゲノムが成立する過程を隠れマルコフモデルで表し、階層ベイズモデルを用いてパラメタを推定しました。その結果、混合は150世代ほど前から始まり、縄文人の割合が45%程度であること、縄文人と弥生人で最も明確な違いを示すのは色素関連遺伝子であること、縄文人は現在網羅的ゲノム個人差データが取得されている世界の50以上の民族のどれにも似ていないことがわかりました。

2.2 転移性配列を指標とした生物地理

マリナーとよばれる転移性配列は、生物の界を超えてゲノム間を転移します。琉球大学の研究グループにより、日本列島と琉球弧の様々な種に転移した配列が発見されています。転移性配列の塩基置換の蓄積から転移年代が推定できると便利ですので、その方法を確立しました。転移性配列は入れ子構造をもつことが多いですが、その組み合わせで配列を階層的に分類し、階層ごとに共通祖先の年代を系図モデルにもとづく棄却法で推定し、それぞれの転移の年代を限定しました。この手法をゲノムが既知のカイコとその野生種であるクワコに適用し、転移配列の内外で置換の蓄積速度はほぼ同じであること、カイコの分岐は700万年前頃であること、ヨナクニサンへの転入は日本列島と琉球弧が陸続きで現在より温暖であった40万年前頃であることがわかりました。

3 ゲノム多様性と医療

社会的要請が強く緊急性のある医療における問題の解決にゲノム多様性解析により貢献することを目指しています。

日本のC型肝炎キャリアは200万人にのぼります。発症する方の7割は感染が持続し、一部の方が肝硬変、肝ガンに至ります。治療の目的は肝ガンを防ぐことです。インターフェロンとリバビリンの併用治療が著効ですが、半数の方には効きません。副作用があり、経済的負担も大きいので、著効と予測される方だけに治療を施すことが望ましいです。一昨年、インターロイキンの遺伝子にある一塩基の個人差の有無が奏功と強い関連をもつことが国立国際医療センターの研究グループにより報告され、現在の治療ガイドラインに反映されています。この個人差で7-8割の患者さんの奏功を正しく判別できますが、精度の高上のために、国立国際医療センターと共同で、臨床データと全ゲノムの90万か所ほどの個人差を解析しています。LASSOとCARTにより構築した判別規則によれば、上記報告データの交互確認では1割誤判別が減少しました。グラフィカルモデルにより、インターロイキン、発ガン関連遺伝子、ウイルスの殻の形成に必要なコレステロールが奏功に関わることが示されました。