

## 出生死滅過程・Yule 過程

分岐年代を推定するうえで、期待される「系図のかたち」は重要な情報となる。ここでは系図のかたちとは分岐順序のことではなく、ある分岐イベントから次の分岐イベントまでの待ち時間（分岐イベント間隔）の分布のしかたを意味する。例えば図 6.10 の3つの系図は、分岐順序は同一だが分岐イベント間隔の分布に違いがある。この図が示しているように、種内変異においては過去の集団動態（あるいは選択）が系図のかたちに影響を与える。それでは種間の系統樹では、系図のかたちはどのように決まってくるのだろうか？

オクスフォード大学の Sean Nee らは、出生死滅過程に基づき、系統樹内の系統数が時間とともにどのように変動をするのかについて理論的な研究を行った(Nee et al. 1992, 1994, Nee 2001)。**出生死滅過程**とは単位時間あたりの系統ごとの種分化率( $\lambda$ )と絶滅率( $\epsilon$ ; 註 1)の期待値の一定性を仮定した連続時間マルコフ過程である。**Yule 過程**(Yule, 1924)は純出生過程ともよばれ、出生死滅過程の特殊なケース( $\epsilon=0$ )である。分子系統では、現在生き残っている系統の遺伝情報のみから過去の歴史を推定するため、系統数は系統樹の根の年代から現在に向かって増加し続ける(註 2)。この増加パターンは、分岐イベント間隔の分布により変動するが、種分化率と絶滅率の影響を大きく受ける。なお、ここでいう絶滅率は、恒常的に起きている絶滅の確率のことである。しかし、恐竜などの絶滅で有名な白亜紀-古第三紀境界のように、生物の歴史では多くの様々な系統がほぼ同時期に一斉に絶滅する「大量絶滅」という現象が繰り返し起きたことが知られている。これは上記の恒常的な絶滅率の確率的なゆらぎでは説明できないことが解っている(Raup, 1992)。Nee et al. (1994)は、大量絶滅を生き延びる確率  $\rho$  も系統樹の系統数の増加パターンに大きな影響を与えることを示した。すなわち種間系統樹では、「系図のかたち」は種分化率、絶滅率そして大量絶滅を生き延びる率により大きく規定される。なお  $\rho$  は、現存する全系統のうち、実際にどれだけの系統を標本として抽出しているのかを示す標本抽出率という意味でも用いることができる。現在ではこちらの意味で用いられることが多いため  $\rho$  を標本抽出率とする。

Nee et al. (1994)によると、時間  $t$  に出現した系統が現在まで絶滅せずに子孫を残し、これが標本中に 1 系統以上含まれる確率は

$$P(0, t) = \frac{\rho(\lambda - \epsilon)}{\rho\lambda + [\lambda(1 - \rho) - \epsilon]e^{(\epsilon - \lambda)t}}$$

である。

ここから時間  $t$  に出現した系統の子孫が標本中に 1 系統のみ含まれる確率  $P_1(t)$  は

$$P_1(t) = \frac{1}{\rho} P(0, t)^2 e^{(\epsilon - \lambda)t}$$

となる。出生死滅過程おける時間  $t$  で、新たな系統が出現する確率分布は以下の分布に従う。

$$g(t) = \frac{\lambda P_1(t)}{1 - \frac{1}{\rho} P(0, t_{root}) e^{(\epsilon - \lambda)t_{root}}}$$

$t_{\text{root}}$  は根の年代である。なお  $\lambda=\varepsilon$  の場合、新たな系統が出現する確率分布は以下のような式になる

$$g(t) = \frac{1 + \rho\lambda t_1}{t_1(1 + \rho\lambda t)^2}$$

新たな系統の出現する点は、系統樹上では内部ノードとして表現される。図 1 に見られるように標本抽出率が高い場合、種分化率が絶滅率と比較して高ければ( $\lambda=10, \varepsilon=1, \rho=1$ )、内部ノードの密度分布は末端ノード側に偏る。その一方種分化率が大きくない場合( $\lambda=1, \varepsilon=1, \rho=1$ )は均一分布に近づく。標本抽出率も確率分布に大きく影響を与え、種分化率が高くても標本抽出率が低ければ( $\lambda=10, \varepsilon=5, \rho=0.001$ )、ノードの密度分布は根側に偏る。Yang and Rannala (1997)は、こうした出生死滅過程に基づく内部ノードの確率密度分布を、分岐年代の事前確率分布として用いている。なお図 16.3 の事前確率分布は MCMCTREE のデフォルトオプションである  $\lambda=1, \varepsilon=1, \rho=0$  という条件で推定したものである。これは図 1 に示すように均一分布となる。

分岐年代推定の過程では各内部ノードの年代の事前確率は、分岐の順序にしたがい密度分布に応じて配置される。様々な種分化率、絶滅率、標本抽出率に基づく内部ノードの分布に、化石記録による制約条件を加味したハチドリの分岐年代の事前確率分布を図 2 に示す。各ノードの年代の事前確率分布は信用区間が非常に大きい。図 1 の内部ノードの分布からも予想されるように、種分化率が高い場合は、標本抽出率が高いと内部ノードの年代はより若い年代で推定され、標本抽出率が低いとより古い年代で推定される傾向にある。これに対して事後確率分布のほうでは、信用区間は短くなり、その中央値も出生死滅過程の影響は限定的である。これはデータから推定される尤度関数がシャープで、出生死滅過程における種分化率や絶滅率などの仮定に対し頑健な推定が行われたことを意味する。もしも事後確率が事前確率の影響を強く受けすぎる場合は、データが分岐年代を推定するうえで十分な情報量を持っていないと考えられる。

註 1: 出生死滅過程で用いる絶滅率には Kendall (1948) など伝統的に  $\mu$  が用いられるが、本稿においては本文で突然変異率の意味で  $\mu$  を用いているので、ここでは extinction (絶滅) の頭文字 e に相当するギリシア文字である  $\varepsilon$  を当てている。

註 2: 進化速度が非常に速い RNA ウイルスは、数年～数十年間という時間スケールでも、塩基配列に大きな違いが生じるため、採取された年代が違う標本を用いればすでに絶滅した系統を含めた系統樹推定を行うことができる。また古代 DNA を用いた場合、2021 年現在の技術では過去数十万年以内に限られるものの、絶滅した個体や集団、種の遺伝情報も用いて系統樹推定を行うことができる。このような例外があるものの

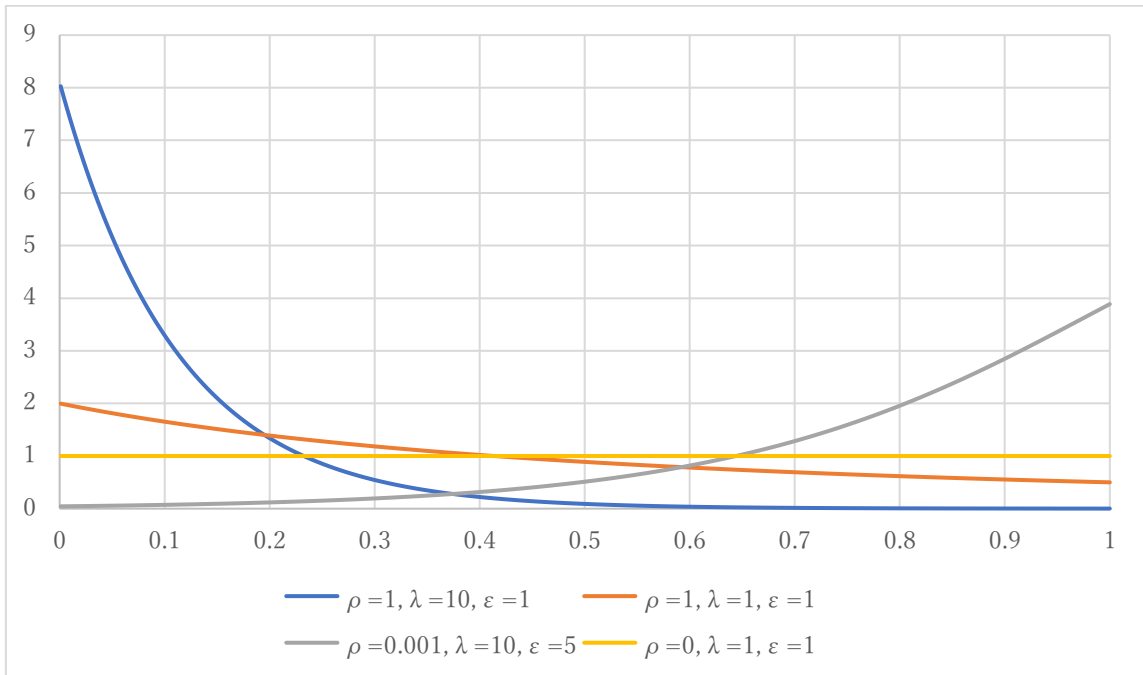


図 1: 出生死滅過程に基づく内部ノードの分布

X 軸は分岐年代（根の年代を 1、末端節の年代を 0 とする）を示す。種分化率( $\lambda$ )、絶滅率( $\epsilon$ )、標本抽出率( $\rho$ )の値により内部ノードの分布は大きく変化する。

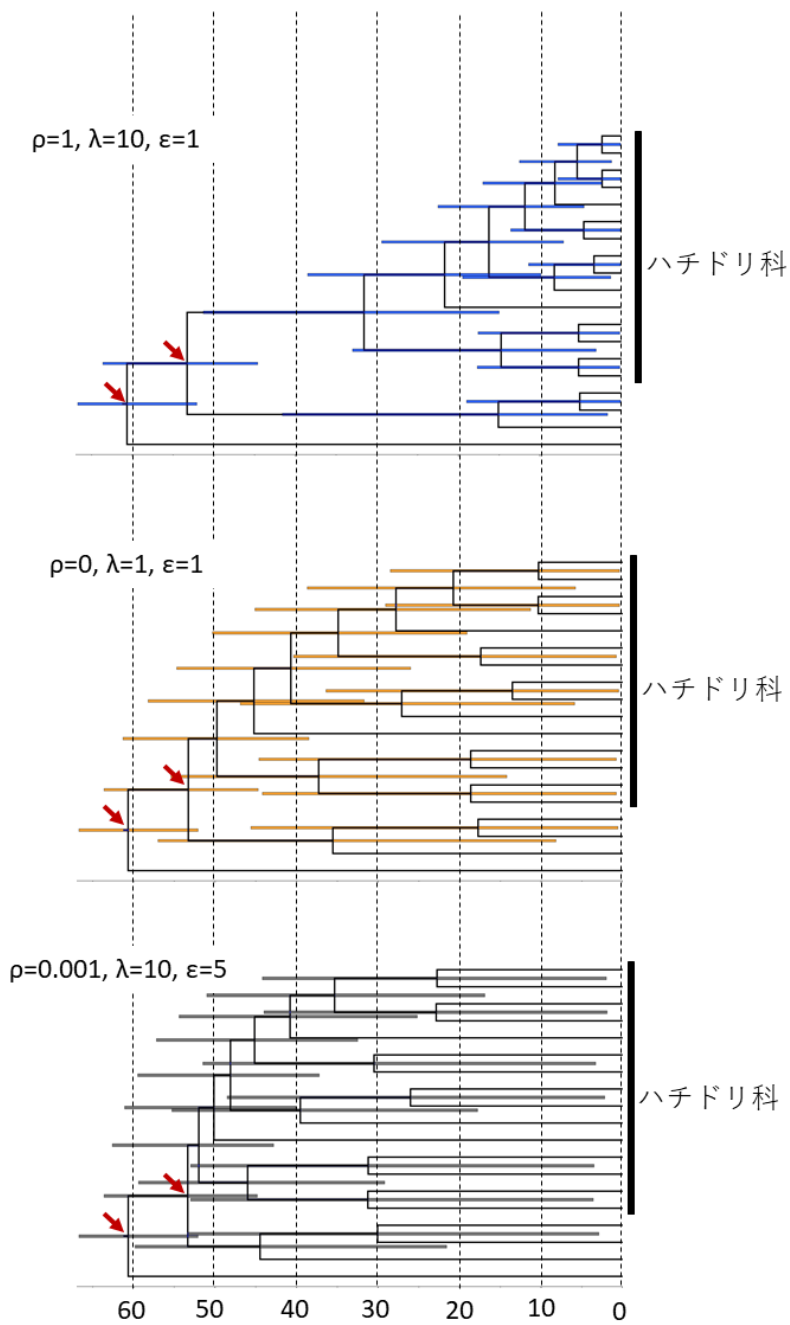


図 2:分岐年代の事前確率

分岐年代の事前確率分布を示した時間軸付き系統樹。ノード上の帯は 95%最高事前密度信用区間を示す。種分化率( $\lambda$ )、絶滅率( $\epsilon$ )、標本抽出率( $\rho$ )の値により事前分布は大きく変化する。赤い矢印で示したノードは化石記録の制約を入れたノードである。

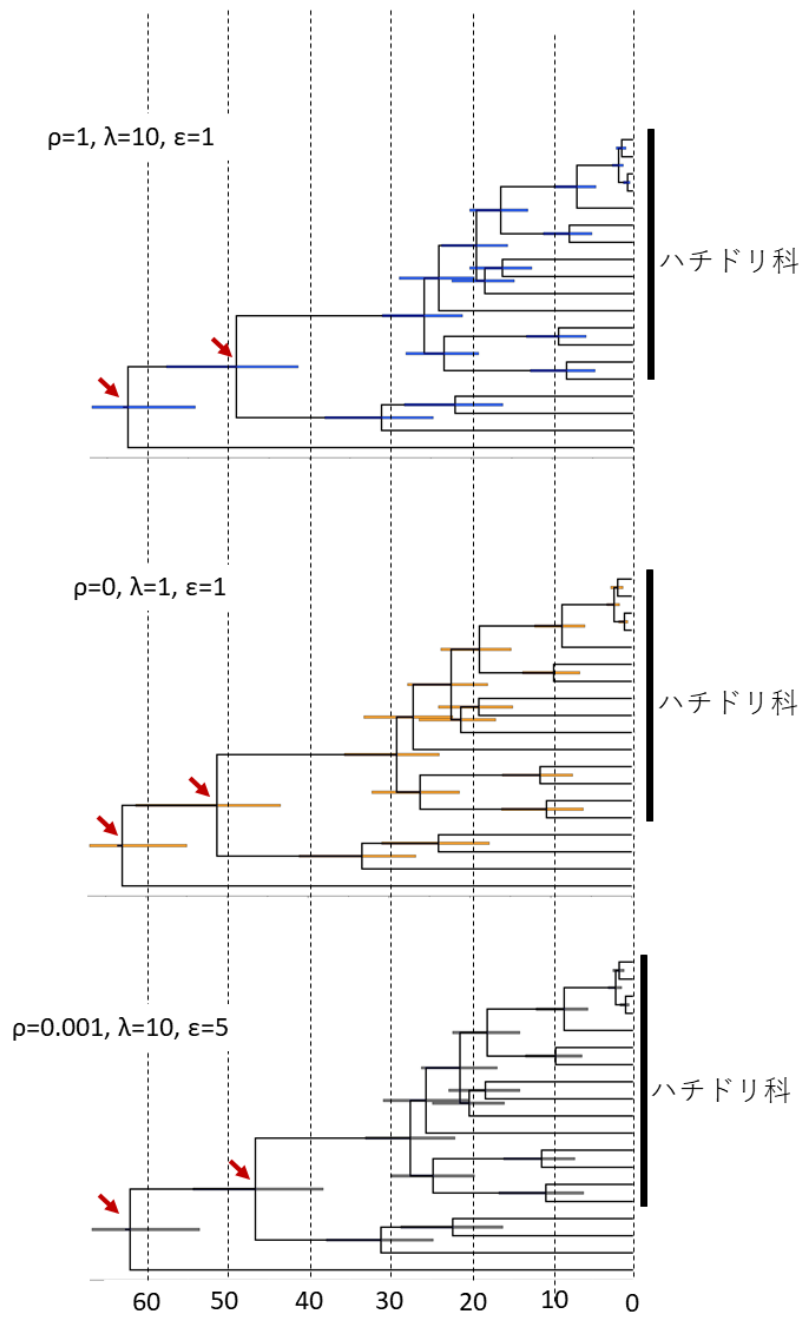


図 3:分岐年代の事後確率

分岐年代の事後確率分布を示す時間軸付き系統樹。ノード上の帯は 95%最高事後密度信用区間を示す。種分化率( $\lambda$ )、絶滅率( $\epsilon$ )、標本抽出率( $\rho$ )の影響は限定的である。赤い矢印で示したノードは化石記録の制約を入れたノードである。

Kendall, D.G. (1948). On the generalized “birth and death” process. *Ann. Math. Stat.* **19**, 1-15.

Nee, S.P. Harvey, H., Mooers, A. Ø. (1992). Tempo and mode of evolution revealed from molecular phylogenies. *Proc. Natl. Acad. Sci. USA* **89**, 8322–8326.

Nee, S., May, R.M., Harvey, P.H. (1994). The reconstructed evolutionary process. *Philos. Trans. R. Soc. Lond. B.* **344**, 305-3 11

Nee, S. (2001) Inferring speciation rate from phylogenies. *Evolution.* **55**, 661–668  
INFERRING

Raup, D.M. (1992) *Extinction: Bad Genes or Bad Luck?*. ISBN 978-0-393-30927-0.

Yang, Z. and Rannala, B (1997). Bayesian Phylogenetic Inference Using DNA Sequences: A Markov Chain Monte Carlo Method. *Mol. Biol. Evol.* **14**, 717–724

Yule, G. U. (1924). A Mathematical Theory of Evolution, Based on the Conclusions of Dr. J. C. Willis, F.R.S. *Philosophical Transactions of the Royal Society B: Biological Sciences.* 213 (402–410): 21–23