The Fractional Coalescent

Somayeh Mashayekhi[*,1] **and Peter Beerli**[†]
[*†]Department of Scientific Computing, Florida State University, Tallahassee, FL 32306

ABSTRACT A new approach to the coalescent is introduced. The *fractional* coalescent (f-coalescent) extends the Kingman's n-coalescent allowing more variability in the patterns of the waiting times. These changes in the waiting times could be caused by population size changes or selection processes. The f-coalescent is derived by applying fractional calculus, in particular the fractional Poisson process which itself is based on the expansion to the fractional waiting times of the conventional Poisson process. The f-coalescent introduces an additional parameter α affecting the steepness of the waiting time function. With an $\alpha = 1$, Kingman's n-coalescent and the f-coalescent are equivalent. The f-coalescent has been implemented in the population genetic model inference software MIGRATE. With an $\alpha = 1$, as expected, both, the n-coalescent and f-coalescent implementation deliver the same results. When either data is simulated under models that allow for $\alpha < 1$ or for real datasets, then Bayes factor comparisons show an improved model fit of the f-coalescent over the n-coalescent.

KEYWORDS Coalescent; Fractional Calculus; Population genetics; Bayesian inference

Introduction

In 1982, Kingman (1982 a,b) introduced the n-coalescent. The n-coalescent describes the probability density function of a genealogy of samples embedded in a population with fixed size. Extensions to this probabilistic description of the genealogical process include changing population size (Kuhner *et al.* 1998) , immigration (Notohara 1990; Wilkinson-Herbots 1998), population divergence (Takahata 1988), selection (Neuhauser and Krone 1997), and recombination (Griffiths and Marjoram 1996). These theoretical advances resulted in several widely use computer packages that estimate various population parameters (for example, (Beerli 2006; Hey and Nielsen 2004; Kuhner 2006)). While the waiting times for events in the n-coalescent are exponentially distributed, The extension of this distribution can be introduced by using Fractional calculus. Fractional calculus have attracted considerable interest because of their ability to model complex phenomena such as continuum and statistical mechanics (Carpinteri and Mainardi 2014), visco-elastic materials (Bagley and Torvik 1985), and solid mechanics (Rossikhin and Shitikova 1997). Our work explores the use of the fractional Poisson process in the context of the coalescent. The fractional Poisson process was derived from the standard Poisson process for two different cases: the time-fractional Poisson process and the space-fractional Poisson process. These two different cases can be combined to introduce the time–space-fractional Poisson process (Orsingher and Polito 2012). We focus on the time-fractional Poisson process. Time-fractional generalizations of the homogeneous Poisson process are based on the substitution of the integer-order derivative operator appearing in Cauchy problems with a fractional-order derivative operator such as the Riemann-Liouville fractional derivative (Laskin 2003) or the Caputo fractional derivative (Beghin *et al.* 2009) (see the Appendix The Mittag-Leffler Functions). Properties of this time-fractional Poisson process have been studied by (Meerschaert *et al.* 2011) and (Politi *et al.* 2011). Laskin (2003) introduced a generalization of the time-fractional Poisson process expressing waiting times using Mittag-Leffler functions, which we use for our discussion.

We use the fractional Poisson process (Laskin 2003) to introduce a new model of coalescent, the fractional coalescent or f-coalescent. This f-coalescent is then implemented in a Bayesian estimator of effective population size and immigration rate; we discuss the implementation and runtime characteristics. We explore the quality of the inference for simulated datasets and also apply the new method to existing real data sets of humpback whales, the malaria parasite *Plasmodium falclparum*, and the H1N1 influenza virus stain.

Copyright © 2017 by the Genetics Society of America
doi: 10.1534/genetics.XXX.XXXXXX
Manuscript compiled: Thursday 28th September, 2017
[1]Department of Scientific Computing, Florida State University, Tallahassee, FL 32306, smashayekhi@fsu.edu

Fractional Coalescent

Looking backward to the time, in a Wright-Fisher population of size N (haploid), the probability that two randomly selected individuals have the same parent in the previous generation is $\frac{1}{N}$, and the probability that they have different parents is $1-\frac{1}{N}$. The probability that they do share a common parent after t generations is $(1-\frac{1}{N})^t$. By using a suitable timescale τ such that one unit of scaled time corresponds to N generations, the probability that the two lineages remain distinct for more than one unit of scaled time is

$$(1-\frac{1}{N})^{N\tau} \to e^{-\tau}, \tag{1}$$

as N goes to infinity. Thus, in the limit, the coalescence time for a pair of lineages is exponentially distributed with mean 1 (Nordborg 2001).

If we assume not only the probability of having same parents for two individuals is not $\frac{1}{N}$, but there is also m different scenarios changing the coalescence probability then the neutral Kingman's n-coalescent becomes incorrect. For each scenario, we can introduce a coalescence probability as $\kappa_i \frac{1}{N}$ for all $i = 1, ..., m$. If we assume that the scenario i is not changing among generation then the probability that the two lineages do not coalescent is

$$(1-\kappa_i\frac{1}{N})^{N\tau}, \tag{2}$$

To make a generalization of Eq. (2), having different scenarios in each generation, we introduce mapping function ω_i such that the probability that the two lineages do not coalescent is

$$\omega_i(1-\kappa_i\frac{1}{N})^{N\tau}. \tag{3}$$

As a conclusion, the total probability that the two lineages remain distinct for more than one unit of scaled time is

$$\sum_{i=1}^{m}\omega_i(1-\kappa_i\frac{1}{N})^{N\tau} \to \mathcal{E}_\alpha(-\tau^\alpha) \tag{4}$$

as N goes to infinity, where $\mathcal{E}_\alpha(-\tau^\alpha)$ is the Mittag-Leffler function (see Appendix The Mittag-Leffler Functions) and α is dependent on ω_i and κ_i. Thus, in the limit, the coalescence time for a pair of lineages is distributed as the fractional generalization of the exponential distribution with rate 1 (MacNamara *et al.* 2017). This result allowed us to descibe a new framework of the coalescent which we call *fractional coalescent* or f-coalescent for short. The f-coalescent is based on the fractional Poisson Process which is a generalization of the classical Poisson process. The n-coalescent generalizes the two-lineages framework to k lineages; we similarly expand the f-coalescent from two to k lineages. The details of the proof for Eq. (4) are presented in the Appendix (The Mittag-Leffler Functions).

The waiting times in Kingman's n-coalescent have an exponential distribution and we express the probability density function of a time interval u_k with k lineages as

$$g_k(u_k|\Theta) = \lambda_k e^{-\lambda_k u_k}, \tag{5}$$

where

$$\lambda_k = \frac{k(k-1)}{\Theta}$$

with the mutation-scaled population size $\Theta = 4N_e\mu$ where N_e is effective population size and μ is mutation rate per generation. For the f-coalescent we express the waiting time with the fractional generalization of the exponential probability distribution. The exponential function is replaced by the Mittag-Leffler function $\mathcal{E}_{\alpha,\alpha}$ which has an additional parameter α (The Mittag-Leffler Functions) as

$$f_k(u_k|\Theta) = u_k^{\alpha-1}\lambda_k\mathcal{E}_{\alpha,\alpha}(-\lambda_k u_k^\alpha). \tag{6}$$

We consider values for α in the interval $(0, 1]$. The Mittag-Leffler function reduces to the exponential function with $\alpha = 1$. Details of the derivation of $\mathcal{E}_{\alpha,\alpha}$ are described in the Appendix. The probability density of all observed u_k with $k = 2, ..., K$ given the mutation-scaled population size Θ is the joint probability distribution function as

$$f(u_2, u_3, ..., u_K|\Theta) = \prod_{k=2}^{K} f_k(u_k|\Theta) = \prod_{k=2}^{K} u_k^{\alpha-1}\lambda_k\mathcal{E}_{\alpha,\alpha}(-\lambda_k u_k^\alpha) \tag{7}$$

where K is the number of samples. To extract a particular genealogy G out of the many possible topologies defined by the interval times $u_2, u_3, ..., u_K$ we need to take into account the number of possible configurations at any time u_k; using the Kingman's n-coalescent for any u_k and k lineages there are $\binom{k}{2}$ potential configurations and we pick one particular one. If we use the same assumption for our f-coalescent that only two lineages per generation can coalesce then we get

$$\begin{aligned} f(G|\Theta) &= \prod_{k=2}^{K} u_k^{\alpha-1}\frac{k(k-1)}{\Theta}\mathcal{E}_{\alpha,\alpha}(-\lambda_k u_k^\alpha) \times \frac{1}{\binom{k}{2}} \\ &= \prod_{k=2}^{K}\left(u_k^{\alpha-1}\mathcal{E}_{\alpha,\alpha}(-\lambda_k u_k^\alpha) \times \frac{2}{\Theta}\right). \end{aligned} \tag{8}$$

Extending to two populations

With two populations, we have two population size Θ_1 and Θ_2, two different number of lineages k_1 and k_2, and two coalescent merging rates

$$\lambda_1 = \frac{k_1(k_1-1)}{\theta_1}, \qquad \lambda_2 = \frac{k_2(k_2-1)}{\theta_2}.$$

Also, suppose the mutation-scaled immigration rate from population one to population two is $\lambda_{12} = k_2 M_{1\to2}$, and the immigration rate from population two to population one is $\lambda_{21} = k_1 M_{2\to1}$. By these assumptions, the total rate of waiting time is $\lambda = \lambda_1 + \lambda_2 + \lambda_{12} + \lambda_{21}$. The probability distribution function for every time interval x on the genealogy G using f-coalescent is

$$f_x(G|k_1, k_2, \theta_1, \theta_2, M_{1\to2}, M_{2\to1}) = u^{\alpha-1}\lambda\mathcal{E}_{\alpha,\alpha}(-\lambda u^\alpha)$$

$$\times \begin{cases} \frac{\lambda_1}{\lambda} \times \frac{1}{\binom{k_1}{2}} & \text{If a coalescent happens in population 1} \\ \frac{\lambda_2}{\lambda} \times \frac{1}{\binom{k_2}{2}} & \text{If a coalescent happens in population 2} \\ \frac{\lambda_{21}}{\lambda} \times \frac{1}{k_1} & \text{If a migration happens from population 2 to 1} \\ \frac{\lambda_{12}}{\lambda} \times \frac{1}{k_2} & \text{If a migration happens from population 1 to 2} \end{cases} \tag{9}$$

where after simplification we have

$$f_x(G|k_1,k_2,\theta_1,\theta_2,M_{1\to2},M_{2\to1}) = u^{\alpha-1}\mathcal{E}_{\alpha,\alpha}(-\lambda u^\alpha) \times \begin{cases} \frac{2}{\theta_1} \\ \frac{2}{\theta_2} \\ M_{2\to1} \\ M_{1\to2} \end{cases} \tag{10}$$

where u is the time of interval x and $\mathcal{E}_{\alpha,\alpha}$ is the Mittag-Leffler function.

Implementation

We implemented our method in the existing framework of our software MIGRATE (Beerli 2006). In this framework we approximate the Bayesian posterior density $p(\rho|X,\alpha) = p(\rho)p(X|\rho,\alpha)/p(X|\alpha)$ where X is the data, ρ is the parameter set for a particular population model, and α is a fixed parameter for the Mittag-Leffler function. The software uses MCMC to approximate the posterior density, calculating $f(G|\rho)$, $p(\rho)$, and $p(X|\rho,\alpha)$. To draw the time, we choose the random number which shows the probability of the event so we have

$$r = P(t > t_0) = 1 - \int_0^{t_0} u^{\alpha-1}\lambda\mathcal{E}_{\alpha,\alpha}(-\lambda u^\alpha)du = \mathcal{E}_\alpha(-\lambda t_0^\alpha), \tag{11}$$

where r is a random numbers between $(0,1)$, so for finding t_0 we solve Eq. (11) by using the bisection method. Also, for drawing the time we suppose $\lambda = (k(k-1)/(4\mu N_e)$ where k is the number of lineages at the beginning of the interval, N_e is the effective population size and μ is the rate of mutation. It should be noted that for calculating the integral in Eq. (11), we use the properties of Mittag-Leffler function (see Appendix The Fractional Poisson Process).

A particular problem for a useful implementation was the calculation time of the generalized Mittag-Leffler function. We programmed the function in the C-language using the algorithm described by (Gorenflo *et al.* 2002) that also was used to create a Matlab-function by I. Podlubny (https://www.mathworks.com/matlabcentral/fileexchange/8738-mittag-leffler-function). We compared our C Mittag-Leffler function with the one implemented in Mathematica 11.0.1.0 (Wolfram Research 2017) for correctness. During these tests we realized that our implementation will be very slow in a MCMC run, because the function will be called millions of times. We implemented a lookup table for values of the Mittag-Leffler functions $\mathcal{E}_\alpha(x)$ and $\mathcal{E}_{\alpha,\alpha}(x)$ for $\alpha \in [0.01, 0.02, 0.03, ...0.99]$. The lookup-table values were created by finding all values for a given α that have $d\mathcal{E}_{\alpha,\alpha}(x)/dx \leq 0.01$ and $d\mathcal{E}_{\alpha,1}(x)/dx \leq 0.01$. This lookup-table is then used with a Hermite-cubic-interpolation and a linear interpolation at the table end points to return values for the Mittag-Leffler functions. This interpolation is roughly 100x faster than our original C-function. Our lookup table and interpolation function can be downloaded from http://popgen.sc.fsu.edu/codes.

Methods

Time to the most recent common ancestor for the f-coalescent: In a coalescent framework the root of a genealogy is also called the most recent common ancestor of the sampled individuals. The time of the most recent common ancestor (TMRCA) is directly related to the effective population size of the population the sample was taken from. We assume that each time interval j is independent from the time interval $j-1$ or $j+1$. With a single parameter, we only need to consider the time intervals and do not need to know details of the topology of the genealogy. We compared empirical distributions of time intervals for a sample of 5 individuals for the f-coalescent, n-coalescent, and Bolthausen-Sznitman-coalescent (BS-coalescent; Bolthausen and Sznitman 1998). BS-coalescent allows merging of more than two lineages and is used to describe the waiting times in a genealogy that is affected by selection (Neher and Hallatschek 2012). All calculations were executed in Mathematica 11.0.1 (Wolfram Research 2017). For the n-coalescent, BS-coalescent, and the f-coalescent we summed independently drawn time intervals for 5, 4, 3, and 2 lineages to get the TMRCA. We drew 10,000 replicated TMRCAs for each of the three n-coalescent settings (strong population shrinkage, fixed population size, strong population growth), two BS-coalescent settings ($T_c = 0.005$ and 0.01), and the two f-coalescent settings ($\alpha = 0.9$ and 0.8). For the fixed population size n-coalescent time intervals we solved Eq. (3) in Beerli and Felsenstein (1999); for time intervals of the growing and shrinking population we used Eq. (4.6) in Hein *et al.* (2005); for the f-coalescent we used Eq. (11).

Simulated data: We evaluated the algorithms using simulations. We updated our simulator package (available at peterbeerli.com/software and bitbucket.com) to allow generating genealogies from the f-coalescent. We generated 100-locus datasets with a population size $N_e = 2500$ and mutation rate per site per generation $\mu = 0.000001$ for α in $[0.4, 0.5, ..1.0]$. Each locus had 10,000 base pairs. Datasets generated with $\alpha < 0.4$ most commonly did not have any variable sites. The datasets were then analyzed with MIGRATE with a fixed α of 0.4 or 0.5, ..., or 1.0. Each dataset was thus analyzed 7 times with different α values allowing us to compare the different runs considering each α value as a different model. We use the framework of marginal likelihoods (cf. Kass and Raftery 1995) to compare the different models. The method implemented in MIGRATE uses thermodynamic integration to approximate the marginal likelihoods (Beerli and Palczewski 2010; Palczewski and Beerli 2014). We then used these marginal likelihood as model weights (Burnham and Anderson 2002) to decide wether models could be considered to be appropriate for the data or not.

Real data: We used three biological datasets to explore whether the f-coalescent could be a better fit than Kingman's n-coalescent. The first dataset is a dataset of short control region sequences of two humpback whale populations (Roman and Palumbi 2003) (H data), the second is a small dataset of the malaria parasite *Plasmodium falciparum* (Joy *et al.* 2003) (P data), and the third dataset is a small dataset of the H1N1 influenza strain the scared the world in 2014 with an outbreak in Mexico (I data) (downloaded on August 7 2017 from www.fludb.org). We used a simple mutation model, F84 + G (Felsenstein and Churchill 1996; Yang 1993), optimized the transition/transversion ratio (ttr) and the site rate parameters in paup* (Swofford 2003). For the whale and Plasmodium data the ttr were 20.07 and 1.56, respectively. The site variation parameter estimates were 0.00456 and 0.00982. The influenza data were heterogenous: segment 1-4 had a site rate variation parameter of ~ 0.02, and the segments 5-8 had a value of > 200.0. We analyzed the total influenza data as 8 independent loci using the F84 model. Independently, we also analyzed the segments 1-4 using F84 + G with a site rate variation of 0.02 and the segements 5-8 using the F84 model.

Results

Time to most recent ancestor for f-coalescent: Figure 1 shows examples of empirical distributions of the TMRCA for a sample of five individuals for the n-coalescent and the f-coalescent with two different values for α. Each curve is based on 10,000 simulated TMRCA values. With $\alpha < 1$ the distribution becomes more peaked with more shorter time intervals and rare, but large time intervals, that lead to heavier tails than with the n-coalescent; maximal values for times in our simulations were: for $\alpha = 1.0$ it was 0.0477, for $\alpha = 0.9$ it was 181.68, and for $\alpha = 0.8$ it was 264.959. The expectation of the TMRCA for the n-coalescent for 5 samples simulated with a $\Theta = 0.01$ is 0.008. The expected TMRCA for the f-coalescent is infinite because of the heavy tails (cf. Repin and Saichev 2000). Using averages of the 10,000 samples matches the expectation for the n-coalescent ($\hat{\Theta}_{\alpha=1.0} = 0.00807681$), but did not match well for our simulation of the f-coalescent ($\hat{\Theta}_{\alpha=0.8} = 0.0294863$, $\hat{\Theta}_{\alpha=0.9} = 0.0570966$), because with $\alpha < 1.0$ we will need many more replicates to sample the heavy tail of the distribution accurately. Comparison with the BS-coalescent are more difficult because the parametrization, using Neher and Hallatschek (2012) of the coalescent rates of the BS-coalescent and the n-coalescent are not equivalent, the parameter T_c in BS-coalescent defines time time to the selectively positive innovation whereas in the f-coalescent and th n-coalescent the rate is defined by the scaled population size. There is certainly a direct relationship between T_c and Θ but under selection that relationship changes. We recognize that distributions of the f-coalescent TMRCA and the BS-coalescent TMRCA look rather similar, but the mapping of the parameter T_c and Θ will need further investigation.

Simulation: Figure 2 shows that it will be difficult to recover the parameter α that was used to simulate the data; data simulated with a particular α has a considerable range when estimated. It is particularly difficult to accurately estimate α when the data was simulated with low α. A problem with these simulations is that the number of variable sites is dependent on the mutation-scaled population size Θ and the simulated branch length. If $\Theta = 0.01$ among all simulations, the number of variables sites varies considerably over the range of α: with an $\alpha = 0.4$ about $\frac{126}{10{,}000}$ sites are variable compared to $\alpha = 1.0$ with $\frac{253}{10{,}000}$ sites. With $\alpha < 0.4$ it is common to see no variable sites in the dataset. Low variability data sets are difficult to use in any coalescent framework.

Real data results: The analysis of the humpback whale datasets (H data) showed that the gene flow between the South-Atlantic and the North-Atlantic is uneven and that the North-Atlantic receives about 5 individuals per generation but also has a much larger population size than the South-Atlantic. our marginal likelihood model selection suggests that the most appropriate model is either the n-coalescent or then a f-coalescent with a very high α, models with an $\alpha < 0.85$ have a model probability of 0.0000. The best model with $\alpha = 0.95$ had modes of $\Theta_{NA} = 0.04283, \Theta_{SA} = 0.01243, M_{SA\to NA} = 133.7,$ and $M_{NA\to SA} = 30.3$ which translates to the number if immigrants $Nm_{NA\to SA} = 0.28$ and $Nm_{SA\to NA} = 5.72$ where NA is North-Atlantic and SA is South-Atlantic. The analysis of the Plasmodium dataset revealed that the Mittag-Leffler tuning parameter α may have a wide range (Fig. 3). Model selection using marginal likelihood suggests that models within the range of $0.55 < \alpha \leq 1.0$ are preferred; models with an $\alpha < 0.55$ had a model probability of 0.0000. The maximum marginal likelihood was at $\alpha = 0.9$. The estimated mutation-scaled population size Θ varied considerably in this range. At an $\alpha = 0.55$, Θ was 0.04817 and at $\alpha = 1.0$ Θ was 0.00197. The best model had a mode and 95%-credibility for Θ of 0.00470 and from 0.00133 to 0.00860, respectively. Both, the P and the H dataset suggest that the n-coalescent model was a good fit. In stark contrast, the 8-segment dataset of the H1N1 strain of influenza from Mexico in 2014 had a well-defined maximum marginal likelihood at $\alpha = 0.8$. The influenza dataset was the only dataset among the tree tested dataset that excluded the n-coalescent as a reasonable model with confidence. Model probability for $\alpha = 0.9$ was 0.0017 and for n-coalescent ($\alpha = 1.0$) was 0.0000. The independent analysis of segments 1-4 and segments 5-8 did not change the result of the full dataset that the best α is not near 1.0 (The electronic data supplement contains the results of these analyses).

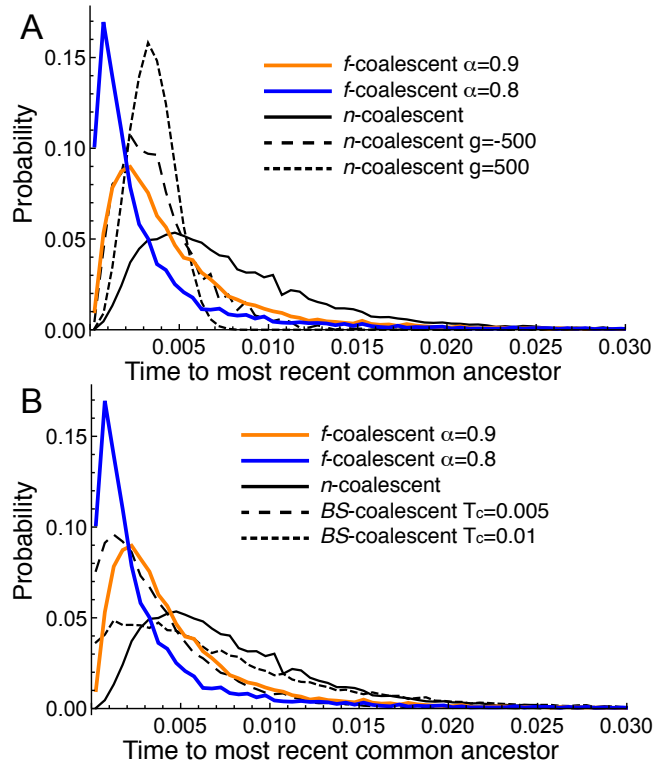

Figure 1 Empirical distribution of the time of the most recent common ancestor for various coalescents: (A) strictly bifurcating; (B) f-coalescent versus n-coalescent and multifurcating Bolthausen-Sznitman coalescent. The x-axis is truncated at 0.03. Each curve represent a histogram of 10,000 draws of the TMRCA.

Discussion

A feature of the f-coalescent is the ability to accommodate very variable time intervals. Mixtures of very short branch lengths with very large branch lengths are possible, whereas the n-coalescent forces a more even distribution of these time intervals. Extensions of the n-coalescent to allow for population growth do not match the variability of time intervals of the f-coalescent. With exponential growth, time interval near the sampling date are enlarged and near the root of the genealogy the time intervals are shortened; the n-coalescent with exponentially shrinking populations also have heavy tails, but seem to have more longer branches than the f-coalescent. In the f-coalescent time inter-

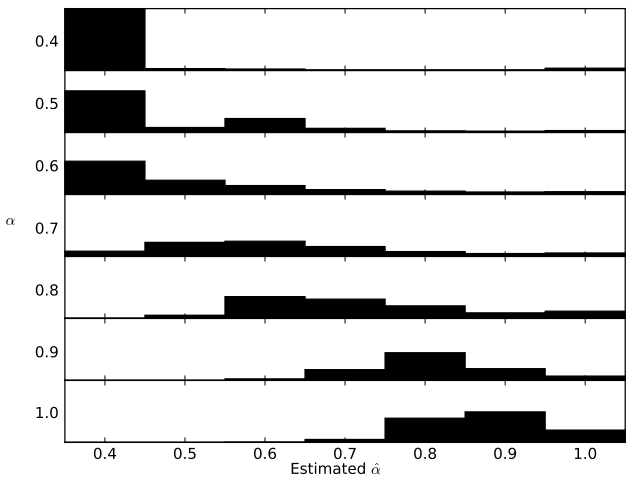

Figure 2 Comparison of recovery of the model under which the data was simulated. α was used to simulate the data; The histograms of $\hat{\alpha}$ show the number of wins in a marginal likelihood comparison. For each α, 500 simulations with 10 individuals and 10,000 base pairs were used.

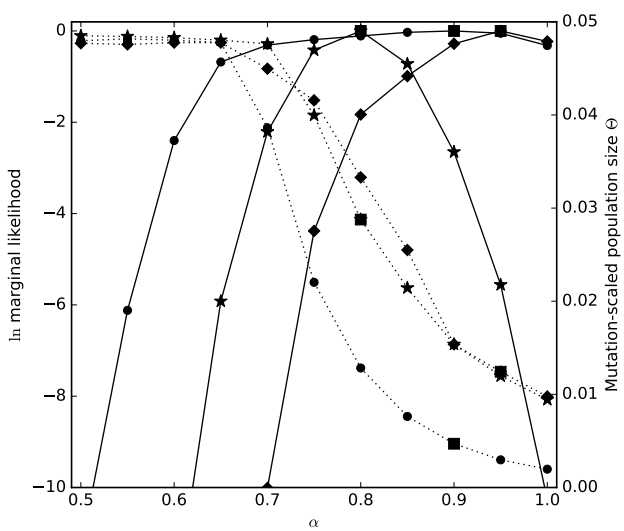

Figure 3 Model selection using relative marginal likelihoods of a *Plasmodium falciparum (circle)*, a H1N1 influenza (star), and a Humpback whale mtDNA (diamond) dataset. The square mark the value of α and the estimate for Θ of the best model for each dataset.

val near the tips are shortened and time intervals near the root are lengthened. The three real data examples suggest that the f-coalescent is a often a better fit to the data (in all of the test cases). Although, the marginal likelihood comparison of different α for the Humpback whale and the Plasmodium data did not exclude the n-coalescent. Only the influenza dataset rejected the n-coalescent and may indicate that the f-coalescent may improve our understanding of fast-evolving organisms that are under strong selection. We believe that the f-coalescent could serve as a model how to approach loci under selection, where the positively selected locus will sweep quickly through the population and result in many short coalescent intervals.

The f-coalescent is a promising extension of the coalescent because it allows considering a wider range of waiting times between events, but certainly will need more study.

Acknowledgements

This work has been supported by the National Science Foundation award DBI 1564822, and the IT resources at the research computing center of Florida State University.

Appendix

The Fractional Poisson Process

The classical Poisson process, is a common random process in probability theory. Several probability distributions arise from the classical Poisson process, for example the Poisson distribution and the exponential distribution. In this process, the number of events, $N(u)$, which happen during the time interval from 0 to u has a Poisson distribution; and, the waiting times between these events are exponentially distributed. Kingman's n-coalescent uses the classical Poisson process to model the waiting time between two coalescent events. For the f-coalescent, we change the distribution of waiting times from an exponential distribution to the fractional generalization of the exponential probability distribution.

Before we introduce the fractional generalization of the exponential probability distribution, we first introduce the fractional Poisson Process.

Fractional Poisson processes have been developed by using the classical Poisson process in two different cases, the time-fractional Poisson process and space-fractional Poisson process Orsingher and Polito (2012). The combination of these two different cases has been introduced as the space-time-fractional Poisson process Orsingher and Polito (2012).

For the classical homogeneous Poisson process we know that

$$p_k(u) = Pr\{N(u) = k\} = e^{-\lambda u}\frac{(\lambda u)^k}{k!}, \quad u \geq 0,\ \lambda > 0,\ k \geq 0, \tag{A1}$$

it is also known that $p_k(u)$ with $k \geq 0$ solves the following Cauchy problem

$$\frac{d}{du}p_k(u) = -\lambda p_k(u) + \lambda p_{k-1}(u), \tag{A2}$$

$$p_k(0) = \delta_{k0}, \tag{A3}$$

where $\delta_{n,0}$ is the Kronecker symbol. Time-fractional generalizations of the Poisson process are based on the substitution of the integer-order derivative operator appearing in Eq. (A2) with a fractional-order derivative operator such as the Riemann-Liouville fractional derivative (Laskin 2003) or the Caputo fractional derivative (Beghin *et al.* 2009) (see Definitions of Fractional

Derivatives). Properties of the time-fractional Poisson process have been studied in (Meerschaert *et al.* 2011) and (Politi *et al.* 2011). Space-fractional generalizations of the Poisson process are based on the substitution of the backward shift operator space in Eq. (A2) with the fractional backward shift operator (Orsingher and Polito 2012).

We use the time-fractional generalization of the homogeneous Poisson process which has been introduced in (Laskin 2003). In this generalization, the fractional Poisson process has been introduced as the counting process with probability $P_\alpha(n,u)$ of having n items $(n = 0,1,2,...)$ during time interval u by the following special form of the fractional Kolmogorov-Feller equation

$$D_R^\alpha P_\alpha(n,u) = \lambda(P_\alpha(n-1,u) - P_\alpha(n,u)) + \frac{u^{-\alpha}}{\Gamma(1-\alpha)}\delta_{n,0}, \quad 0 < \alpha \leq 1, \tag{A4}$$

where $\delta_{n,0}$ is the Kronecker symbol with normalization condition

$$\sum_{n=0}^{\infty} P_\alpha(n,u) = 1,$$

By solving Eq. (A4) we have Laskin (2003)

$$P_\alpha(n,u) = \frac{(\lambda u^\alpha)^n}{n!}\sum_{k=0}^{\infty}\frac{(k+n)!}{k!}\frac{(-\lambda u^\alpha)^k}{\Gamma(\alpha(k+n)+1)}, \quad 0 < \alpha \leq 1. \tag{A5}$$

In this generalization, the probability distribution function of waiting time for the fractional Poisson process can be expressed as

$$\psi_\alpha(u) = -\frac{d}{du}P_\alpha(u), \tag{A6}$$

where $P_\alpha(u)$ is the probability that a given inter arrival time is greater or equal to u

$$P_\alpha(u) = 1 - \sum_{n=1}^{\infty} P_\alpha(n,u) = \mathcal{E}_\alpha(-\lambda u^\alpha), \tag{A7}$$

where $\mathcal{E}_\alpha$ is the Mittag-Leffler function see (The Mittag-Leffler Functions). By using Eqs. (A6) and (A7) the waiting time for fractional Poisson process has the following probability distribution function

$$\psi_\alpha(u) = \lambda u^{\alpha-1}\mathcal{E}_{\alpha,\alpha}(-\lambda u^\alpha), \quad u \geq 0, \quad 0 < \alpha \leq 1, \tag{A8}$$

where $\mathcal{E}_{\alpha,\alpha}$ is the generalization of the Mittag-Leffler function (see The Mittag-Leffler Functions). The $\psi_\alpha(u)$ defined by Eq. (A8) is fractional generalization of the well known exponential probability distribution.

It should be noted that the generalization of the Mittag-Leffler function has two different parameters α and β (see The Mittag-Leffler Functions), and as it has been mentioned in Eq. (A8) the fractional Poisson process, has been derived by using the Mittag-Leffler function using ($\alpha = \beta$). Figure 4 shows the time of 100 independent events which happen in three cases of the Poisson process with various α (1.0, 0.9, and 0.8). For calculating these times we have supposed that the probability each event occurs is a random number while λ has been chosen to be fixed and same in all three different cases. With an $\alpha < 1$ more times became shorter but a few times become longer; and the variance of the times becomes larger.

Figure 4 Times between events drawn using the classical Poisson process and the fractional Poisson process with $\alpha = 0.9$, and 0.8; from top to bottom..

Definitions of Fractional Derivatives

There are various definitions for fractional derivatives. The widely used definitions include the Riemann-Liouville fractional derivatives (Podlubny 1998).

Definition : Riemann-Liouville's fractional derivative of order α is defined as

$$(D_R^\alpha f)(t) = \frac{1}{\Gamma(n-\alpha)}\frac{d^n}{dt^n}\int_0^t \frac{f(s)}{(t-s)^{\alpha+1-n}}ds, \quad n-1 < \alpha \leq n,\ n \in \mathbb{N}, \tag{A9}$$

where $\alpha > 0$ is the order of the derivative and n is the smallest integer greater than α. For the Riemann-Liouville's derivative we have

$$D_R^\alpha c = c\frac{t^{-\alpha}}{\Gamma(1-\alpha)}, \tag{A10}$$

where c is a constant. Replacing c with t^ν we get

$$D_R^\alpha t^\nu = \frac{\Gamma(\nu+1)}{\Gamma(\nu+1-\alpha)}t^{\nu-\alpha}, \quad n-1 < \alpha < n,\ \nu > -1,\ \nu \in \mathbb{R}. \tag{A11}$$

The Mittag-Leffler Functions

The Mittag-Leffler functions $\mathcal{E}_\alpha(x)$ with $\alpha > 0$ have been introduced as (Mainardi 2010)

$$\mathcal{E}_\alpha(x) = \sum_{n=0}^{\infty}\frac{x^n}{\Gamma(\alpha n+1)}, \quad \alpha > 0, x \in \mathcal{C}. \tag{A12}$$

This function provides a simple generalization of the exponential function because of the substitution in the exponential series of $n! = \Gamma(n+1)$ with $(\alpha n)! = \Gamma(\alpha n+1)$. A straightforward generalization of the Mittag-Leffler function is obtained by replacing the additive constant 1 in the argument of the *Gamma* function in Eq. (A12) by an arbitrary complex parameter β. For this function we use the following notation

$$\mathcal{E}_{\alpha,\beta}(x) = \sum_{n=0}^{\infty}\frac{x^n}{\Gamma(\alpha n+\beta)}, \quad \alpha > 0, \beta, x \in \mathcal{C} \tag{A13}$$

of course $\mathcal{E}_{\alpha,1}(x) = \mathcal{E}_\alpha(x)$. The Mittag-Leffler function can be expressed as a mixture of exponentials (MacNamara *et al.* 2017)

$$\mathcal{E}_\alpha(-\lambda x^\alpha) = \int_0^\infty \omega(\kappa)e^{-\kappa x}d\kappa, \tag{A14}$$

where $\omega(\kappa)$ is the probability density

$$\omega(\kappa) = \lambda\frac{sin(\alpha\pi)}{\pi}\frac{\kappa^{\alpha-1}}{\kappa^{2\alpha}+2\lambda\kappa^\alpha cos(\alpha\pi)+\lambda^2}.$$

The discrete form of Eq. (A14) is

$$\mathcal{E}_\alpha(-\lambda x^\alpha) = \sum_k \omega_k(1-\kappa_k\frac{x}{n})^n, \tag{A15}$$

as $k \to \infty$ where $\omega_k = \omega(\kappa_k)$. As $n \to \infty$ then $\omega_k \to \omega(\kappa)$ and $\sum_k \omega_k(1-\kappa_k\frac{x}{n})^n \to \int_0^\infty \omega(\kappa)e^{-\kappa x}d\kappa$. It should be noted since

summation in Eq. (A15) is converging as $n \to \infty$, we can rewrite Eq. (A15) as

$$\mathcal{E}_\alpha(-\lambda x^\alpha) = \sum_k \omega_k (1 - \kappa_k \frac{1}{n})^{xn}. \tag{A16}$$

Literature Cited

Bagley, R. L. and P. J. Torvik, 1985 Fractional calculus in the transient analysis of viscoelastically damped structures. AIAA journal **23**: 918–925.

Beerli, P., 2006 Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. Bioinformatics **22**: 341–345.

Beerli, P. and J. Felsenstein, 1999 Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics **152**: 763–773.

Beerli, P. and M. Palczewski, 2010 Unified framework to evaluate panmixia and migration direction among multiple sampling locations. Genetics **185**: 313–326.

Beghin, L., E. Orsingher, *et al.*, 2009 Fractional Poisson processes and related planar random motions. Electron. J. Probab **14**: 1790–1826.

Bolthausen, E. and A.-S. Sznitman, 1998 On Ruelle's probability cascades and abstract cavity method. Communications in Mathematical Physics **197**: 247–276.

Burnham, K. and D. Anderson, 2002 *Model Selection and Multimodel Inference: A Practical Information-theoretic Approach*. Springer, New York.

Carpinteri, A. and F. Mainardi, 2014 *Fractals and fractional calculus in continuum mechanics*, volume 378. Springer.

Felsenstein, J. and G. A. Churchill, 1996 A hidden markov model approach to variation among sites in rate of evolution. Molecular Biology and Evolution **13**: 93–104.

Gorenflo, R., J. Loutchko, and Y. Luchko, 2002 Computation of the Mittag-Leffler function $e_{\alpha,\beta}(z)$ and its derivative. Fractional Calculus and Applied Analysis **5**: 491–518.

Griffiths, R. and P. Marjoram, 1996 Ancestral inference from samples of DNA sequences with recombination. Journal of Computational Biology **3**: 479–502.

Hein, J., M. H. Schierup, and C. Wiuf, 2005 *Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory*. Oxford University Press, Oxford.

Hey, J. and R. Nielsen, 2004 Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. Genetics **167**: 747–760.

Joy, D. A., X. Feng, J. Mu, T. Furuya, K. Chotivanich, *et al.*, 2003 Early origin and recent expansion of Plasmodium falciparum. Science (New York, N.Y.) **300**: 318–321.

Kass, R. E. and A. E. Raftery, 1995 Bayes factors. Journal of the American Statistical Association **90**: 773–795.

Kuhner, M., 2006 LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics **22**: 768–70.

Kuhner, M. K., J. Yamato, and J. Felsenstein, 1998 Maximum likelihood estimation of population growth rates based on the coalescent. Genetics **149**: 429–434.

Laskin, N., 2003 Fractional Poisson process. Communications in Nonlinear Science and Numerical Simulation **8**: 201–213.

MacNamara, S., B. Henry, and W. McLean, 2017 Fractional euler limits and their applications. SIAM Journal on Applied Mathematics **77**: 447–469.

Mainardi, F., 2010 *Fractional calculus and waves in linear viscoelasticity: an introduction to mathematical models*. World Scientific.

Meerschaert, M. M., E. Nane, P. Vellaisamy, *et al.*, 2011 The fractional Poisson process and the inverse stable subordinator. Electronic Journal of Probability **16**: 1600–1620.

Neher, R. A. and O. Hallatschek, 2012 Genealogies of rapidly adapting populations. Proceedings of the National Academy of Sciences **110**: 437–442.

Neuhauser, C. and S. M. Krone, 1997 The genealogy of samples in models with selection. Genetics **145**: 519–34.

Nordborg, M., 2001 Coalescent theory. Handbook of statistical genetics .

Notohara, M., 1990 The coalescent and the genealogical process in geographically structured population. Journal of Mathematical Biology **29**: 59–75.

Orsingher, E. and F. Polito, 2012 The space-fractional Poisson process. Statistics & Probability Letters **82**: 852–858.

Palczewski, M. and P. Beerli, 2014 Population model comparison using multi-locus datasets. In *Bayesian Phylogenetics: Methods, Algorithms, and Applications*, edited by M.-H. Chen, L. Kuo, and P. O. Lewis, pp. 187–200, CRC Press.

Podlubny, I., 1998 *Fractional differential equations: an introduction to fractional derivatives, fractional differential equations, to methods of their solution and some of their applications*, volume 198. Academic press.

Politi, M., T. Kaizoji, and E. Scalas, 2011 Full characterization of the fractional Poisson process. EPL (Europhysics Letters) **96**: 20004.

Repin, O. N. and A. I. Saichev, 2000 Fractional Poisson law. Radiophysics and Quantum Electronics **43**: 738–741.

Roman, J. and S. Palumbi, 2003 Whales before whaling in the North Atlantic. Science **301**: 508–510.

Rossikhin, Y. A. and M. V. Shitikova, 1997 Applications of fractional calculus to dynamic problems of linear and nonlinear hereditary mechanics of solids. Applied Mechanics Reviews **50**: 15–67.

Swofford, D., 2003 PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.

Takahata, N., 1988 The coalescent in two partially isolated diffusion populations. Genetical Research **52**: 213–22.

Wilkinson-Herbots, H. M., 1998 Genealogy and subpopulation differentiation under various models of population structure. Journal of Mathematical Biology **37**: 535–585.

Wolfram Research, 2017 Mathematica, Version 11.1. Champaign, IL.

Yang, Z., 1993 Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Molecular Biology and Evolution **10**: 1396–1401.