

Identification of graphical interaction structures from time series data of Allele frequencies

Yasin

Department of Mathematics, Makerere University
Department of Mathematics, KTH Royal Institute of
Technology

*Second Network Meeting for Sida- and ISP-funded PhD Students in
Mathematics*

Stockholm 26–27 February 2018

My Advisors

Timo Koski
Main advisor
KTH

Boualem Djehiche
Assistant advisor
KTH

Juma Kasozi
Assistant advisor
Makerere Univ.

J. Y. T Mugisha
Assistant advisor
Makerere Univ.

Identification of graphical interaction structures from time series data of Allele frequencies

Outline

- Coupled Wright-Fisher diffusion

Outline

- Coupled Wright-Fisher diffusion
 - Examples of Interaction networks

Outline

- Coupled Wright-Fisher diffusion
 - Examples of Interaction networks
- Model selection

Outline

- Coupled Wright-Fisher diffusion
 - Examples of Interaction networks
- Model selection
- Conclusion and Future work

Coupled Wright-Fisher diffusion

$$d\mathbf{X}_t = [\boldsymbol{\mu}(\mathbf{X}_t) + \mathbf{G}(\mathbf{X}_t, \boldsymbol{\theta})]dt + \sqrt{\mathbf{D}(\mathbf{X}_t)}d\mathbf{W}_t, \quad (1)$$

$\boldsymbol{\theta}$ is a vector of parameters.

Coupled Wright-Fisher diffusion

$$d\mathbf{X}_t = [\boldsymbol{\mu}(\mathbf{X}_t) + \mathbf{G}(\mathbf{X}_t, \boldsymbol{\theta})]dt + \sqrt{\mathbf{D}(\mathbf{X}_t)}d\mathbf{W}_t, \quad (1)$$

$\boldsymbol{\theta}$ is a vector of parameters.

Application: Modelling changes in Allele frequencies considering possible interactions between different loci. (see Aurell et. al 2017)

Coupled Wright-Fisher diffusion

For instance adopting (Aurell et al. 2017),

1) **Two alleles and two loci**,

$$\begin{aligned} dx_t^{(1)} &= [u_1^{(1)} - (u_1^{(1)} + u_2^{(1)})x_t + h(1 - x_t)x_t y_t]dt + \\ &\quad \sqrt{x_t(1 - x_t)}dW_1, \\ dy_t^{(2)} &= [u_1^{(2)} - (u_1^{(2)} + u_2^{(2)})y_t + h(1 - y_t)y_t x_t]dt + \\ &\quad \sqrt{y_t(1 - y_t)}dW_2, \end{aligned} \tag{2}$$

with initial conditions

$$x_0 = X_0, y_0 = Y_0, 0 \leq x_t \leq 1, 0 \leq y_t \leq 1.$$

Coupled Wright-Fisher diffusion

2) Two alleles and Four loci,

$$dx_t^{(1)} = \mu^{(1)}(x_t^{(1)})dt + x^{(1)}(1 - x^{(1)})(h_2x^{(2)} + h_5x^{(3)} + h_7x^{(4)})dt + \sqrt{x_t^{(1)}(1 - x_t^{(1)})}dW_t^{(1)} \quad (3)$$

$$dx_t^{(2)} = \mu^{(2)}(x_t^{(2)})dt + x^{(2)}(1 - x^{(2)})(h_2x^{(1)} + h_3x^{(3)} + h_6x^{(4)})dt + \sqrt{x_t^{(2)}(1 - x_t^{(2)})}dW_t^{(2)} \quad (4)$$

$$dx_t^{(3)} = \mu^{(3)}(x_t^{(3)})dt + x^{(3)}(1 - x^{(3)})(h_5x^{(1)} + h_3x^{(2)} + h_4x^{(4)})dt + \sqrt{x_t^{(3)}(1 - x_t^{(3)})}dW_t^{(3)} \quad (5)$$

$$dx_t^{(4)} = \mu^{(4)}(x_t^{(4)})dt + x^{(4)}(1 - x^{(4)})(h_7x^{(1)} + h_6x^{(2)} + h_4x^{(3)})dt + \sqrt{x_t^{(4)}(1 - x_t^{(4)})}dW_t^{(4)} \quad (6)$$

Coupled Wright-Fisher diffusion

$$\mu^{(i)}(x_t^{(i)}) = u_1^{(i)} - (u_1^{(i)} + u_2^{(i)})x_t^{(i)} \quad i = 1, 2, 3, 4.$$

Coupled Wright-Fisher diffusion

$$\mu^{(i)}(x_t^{(i)}) = u_1^{(i)} - (u_1^{(i)} + u_2^{(i)})x_t^{(i)} \quad i = 1, 2, 3, 4.$$

Generally, different interaction configuration, different diffusion

Coupled Wright-Fisher diffusion

$$\mu^{(i)}(x_t^{(i)}) = u_1^{(i)} - (u_1^{(i)} + u_2^{(i)})x_t^{(i)} \quad i = 1, 2, 3, 4.$$

Generally, different interaction configuration, different diffusion

Suppose a diffusion process is discretely sampled (data is available). What is the interaction structure underlying the data?

Model selection

Bayes factor:

$$B_{12}(X) = \frac{P(X/m_1)}{P(X/m_2)}, \quad (7)$$

$$P(X/m_i) = \int_{\mathbb{R}^n} P(X/\theta, m_i)P(\theta/m_i)d\theta. \quad (8)$$

Model selection

By Girsanov theorem,

$$P(X/\theta, m_i) = \frac{d\mathbb{P}_\theta}{d\mathbb{P}_{\theta_0}} = \exp \left\{ \int_0^1 A^{-1}(\mathbf{C}_i(\mathbf{X}_t, \theta) - \mathbf{C}_i(\mathbf{X}_t, \theta_0)) \cdot d\mathbf{X}_t - \frac{1}{2} \int_0^1 [A^{-1}\mathbf{C}_i(\mathbf{X}_t, \theta) \cdot \mathbf{C}_i(\mathbf{X}_t, \theta) - A^{-1}\mathbf{C}_i(\mathbf{X}_t, \theta_0) \cdot \mathbf{C}_i(\mathbf{X}_t, \theta_0)] dt \right\} \quad (9)$$

$\mathbf{A} \cdot \mathbf{B}$ denotes the dot product between \mathbf{A} and \mathbf{B} .

Model selection

Observation : $P(X/m_i)$ is **intractable** !

Model selection

Observation : $P(X/m_i)$ is **intractable** !

Approximated via ABC-methods (Tina et al. 2009)

Challenge: requires exact simulation of multivariate diffusion of the form (1). (Beskos et al. 2008 and recently in Blanchet & Zhang (2017) for general multivariate diffusion but complicated to be used in practice,)

Model selection

Observation : $P(X/m_i)$ is **intractable** !

Approximated via ABC-methods (Tina et al. 2009)

Challenge: requires exact simulation of multivariate diffusion of the form (1). (Beskos et al. 2008 and recently in Blanchet & Zhang (2017) for general multivariate diffusion but complicated to be used in practice,)

Question Would the exact simulation problem be simpler if we have the stationary density to diffusion explicitly known at the same time avoiding estimation of the transition density?

Simulation Results

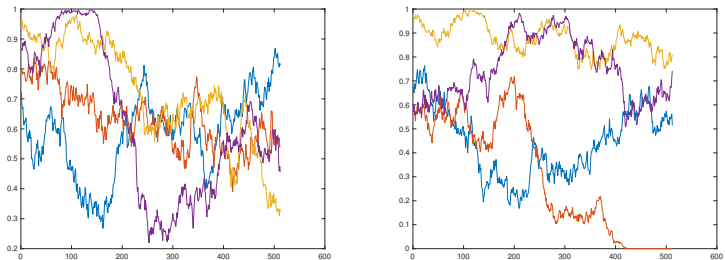


Figure: Sample paths of coupled Wright-Fisher diffusion model for four loci and two alleles using projection method of Wang & Carreira-Perpinan (2013)

Future work

In the future we intend to explore the following;

Future work

In the future we intend to explore the following;

- extension our work to more loci and more alleles

Future work

In the future we intend to explore the following;






- extension our work to more loci and more alleles
- a more general interactions other than same type interactions

Future work




In the future we intend to explore the following;

- extension our work to more loci and more alleles
- a more general interactions other than same type interactions
- develop exact simulation methods that rely on availability of stationary densities at the same time avoiding transition density approximations




Key References

-  Ait-Sahalia Y. (2008). "Closed-form likelihood expansions for multivariate diffusions " The Annals of Statistics 2008, Vol. 36, No. 2, 906-937 DOI: 10.1214/009053607000000622.
-  Aurell E., Ekeberg M., Koski T.(2017) "Wright-Fisher model with selection and mutation" submitted.
-  Beskos, A. and Roberts, G. O. (2005). Exact simulation of diffusions. The Annals of Applied Probability 2005, Vol. 15, No. 4, 2422-2444.
-  Beskos, A., Papaspiliopoulos, O., and Roberts, G. O. (2006). Retrospective exact simulation of diffusion sample paths with applications. Bernoulli, 12, 1077-1098.
-  Beskos, A., Papaspiliopoulos, O., and Roberts, G. O. (2008).

Key References

-  Dangerfield C.E., Kay D., MacNamara S., & Burrage K. (2012) "A boundary preserving numerical algorithm for the Wright-Fisher model with mutation". BIT Numerical Mathematics, 52(2), pp. 283-304.
-  Jenkins P. A. & Spano D. (2016) "Exact simulation of the Wright-Fisher diffusion " arXiv:1506.06998v2 [stat.ME] 22 Jul 2016.
-  Picchini U. (2013) "Inference for SDE models via Approximate Bayesian Computation " arXiv:1204.5459v6 [stat.ME] 8 Nov 2013.

Key References

-  Skwark M. J., Croucher N. J., Puranen S. , Chewapreecha C., Pesonen M., Xu Y. Y., Turner P. , Harris S.R., Beres S. P., Musser J. M., Parkhill J. , Bentley S.D., Aurell E. and Corander J. (2017). “ Interacting networks of resistance, virulence and core machinery genes identified by genome-wide epistasis analysis” PLoS Genetics, 13(2), e1006508.
-  Taraskin F. A. (1974) “ On the asymptotic normality of vector-valued stochastic integrals and estimates of drift parameters of a multidimensional diffusion process” Theor. probability and Math. Statist. (1974) No. 2.
-  Tina T., Welch D., Strelkova N., Ipsen A., Stumpf M. P.H. (2009) “ Approximate Bayesian Computation scheme for parameter inference and model selection in dynamical

Tack så mycket!

Thank you!