

Propuesta de Tesis en Ingeniería Matemática

Identificación del proponente

- Nombre: María Inés Fariello Rico
- Último título obtenido: Doctorado
- Lugar de trabajo: Facultad de Ingeniería e Instituto Pasteur de Montevideo
- Área de trabajo: Probabilidad y Estadística/Genética de Poblaciones
- Información de contacto: fariello@fing.edu.uy

Identificación de la propuesta de proyecto de tesis

- Título del proyecto: Extensión del test haplotípico para detectar selección hapflk a poblaciones con eventos de migración
- Área temática del conocimiento de la propuesta: Probabilidad y Estadística/Genética de poblaciones
- Resumen: (entre media y una carilla)

La genética de poblaciones es un área donde se encuentran la estadística y la genética. Esta disciplina describe las poblaciones genéticamente, determinando por ejemplo, su estructura, la influencia de determinadas poblaciones sobre otras, tamaños poblacionales miles de años atrás o buscando trazas de selección. En los últimos años, motivados por la abundancia de datos genómicos producidos, esta disciplina ha crecido a

pasos agigantados ya que es necesario desarrollar técnicas estadísticas para poder explorar y entender este tipo de datos.

La detección de selección es una sub-área de la genética de poblaciones, que busca detectar genes que presenten patrones que indiquen que las poblaciones no han evolucionado de manera neutra. Bajo las hipótesis adecuadas, se puede decir que los patrones detectados surgen debido a presiones selectivas. Éstas pueden ser de carácter natural (como los genes responsables de la digestión de leche en humanos), o artificial, como lo son las características productivas en animales.

Diversos métodos han surgido en los últimos años. Luego del 2000 y la explosión de datos genómicos, los tests se han enfocado en tomar en cuenta las dependencias entre genes, ya que éstos se transmiten juntos de una generación a la otra. Por otro lado, las estructuras poblacionales definen distintas relaciones de covarianza entre las poblaciones. *Fariello et al.* desarrollaron un test que toma en cuenta ambos tipos de dependencias, pero está restringido a grupos de poblaciones con pocos o ningún evento migratorio. Esta condición se debe a que el test se basa en una matriz de parentesco entre poblaciones, y el método utilizado para calcularla, tanto como el modelo, asumen que no hay migraciones entre las poblaciones. *Pickrell and Pritchard* proponen un modelo similar al usado por *Fariello et al.* (propuesto por *Bonhomme et al.*) que toma en cuenta las migraciones y estima una matriz de parentesco entre poblaciones, pero estimando eventuales migraciones entre las poblaciones.

El estudiante deberá evaluar si el modelo utilizado en *Pickrell and Pritchard* puede ser integrado al test de *Fariello et al.* En el caso que sea posible, deberá realizar simulaciones para evaluar el funcionamiento de dicho test y compararlo con su versión anterior. Si el test funciona sobre simulaciones, se aplicará a datos reales.

- Posibles aplicaciones científicas: Este método servirá para detectar trazas de selección casi en cualquier grupo de poblaciones, ya que levanta la restricción de utilizar solamente poblaciones que no hayan tenido migraciones que necesita el actual método hapflk. Cuando se desarrolló este método se aplicó principalmente a poblaciones animales, para las cuales su estructura poblacional estaba bien definida. Esta extensión permitirá su aplicación a tanto a humanos como a otras es-

pecies, que provengan de poblaciones mezcladas. El estudiante tendrá la oportunidad de aplicar el método a datos de genomas uruguayos.

- Posibles aplicaciones productivas y/o sociales:

La detección de selección puede ayudar a detectar genes relevantes para la producción agrícola nacional. Identificar estos genes, puede ser muy útil en programas posteriores de selección genómica.

- Metodología:

El estudiante deberá primero familiarizarse con el modelo probabilístico asociado a los tests desarrollados por [1–3]. Luego deberá proponer cómo integrar el modelo planteado por [3] al test haplotípico hapflk. Si el estudiante encuentra otra manera de incluir los eventos de migración en el modelo de [1], también podrá hacerlo. Para validar el método el estudiante deberá realizar simulaciones de evolución de poblaciones con eventos de migración entre ellas. Algunos scripts en python están disponibles, pero deberán ser adaptados. También si lo considera pertinente podrá utilizar otros softwares de simulación, como por ejemplo MS de [4].

- Perfil esperado del estudiante: Indique si la tesis está asignada a un estudiante, o que conocimientos previos son deseables para la concreción.

Son deseables conocimientos de R y conocimientos básicos de python. Los conocimientos de genética de poblaciones son deseados pero no excluyentes.

Bibliografía relevante (indique entre 4 y 8 referencias relevantes)

References

- [1] Bonhomme, M. et al. *Detecting selection in population trees: The Lewontin and Krakauer test extended* Genetics, 2010
- [2] Fariello, M. I. et al. *Detecting signatures of selection through haplotype differentiation among hierarchically structured populations* Genetics, 2013

[3] Pickrell, J. K. and Pritchard, J. K. *Inference of population splits and mixtures from genome-wide allele frequency data*. Nature Precedings, 2012

[4] Hudson, R. *Generating samples under the Wright-Fisher neutral model of genetic variation* Bioinformatics, 2002

- Comentarios adicionales:
- Lugar y Fecha de la propuesta:
Montevideo, 4 de agosto de 2014