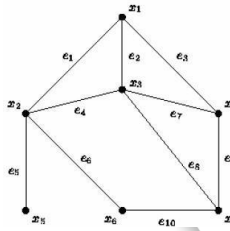


BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1

LUIS JARAMILLO MSC

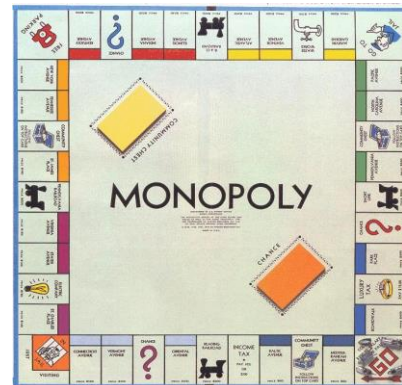
Las Cadenas de Markov

- Un conjunto de estados
- Un conjunto de cambios posibles de un estado a otro (c.u con su respectiva probabilidad)
- El estado en el estado en el tiempo n solo depende del estado en el tiempo $n-1$ (Propiedad de Markov)



BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1

LUIS JARAMILLO MSC



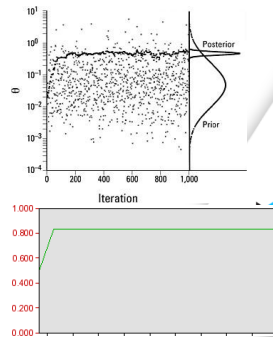
UTEC

BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1

LUIS JARAMILLO MSC

Distribución estacionaria

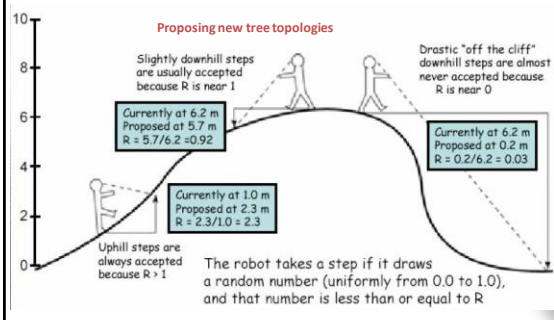
- Una cadena de Markov con estados s_1, \dots, s_n tiene al menos una distribución estacionaria π_1, \dots, π_n
- Cuando la cadena es corrida por varias generaciones la distribución s_k **tiende** a π_k
- Entonces s_1, \dots, s_n serán los estados del modelo (árboles generados) y π_1, \dots, π_n será la probabilidad posterior



BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1

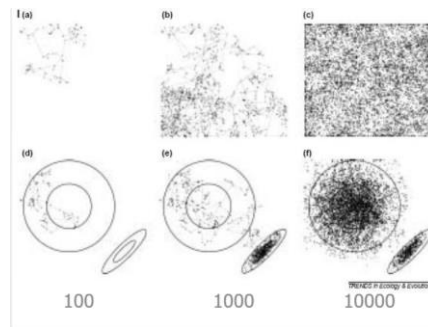
LUIS JARAMILLO MSC

Monte Carlo Markov Chain MCMC or (MC)²



BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1

LUIS JARAMILLO MSC



UTEC

BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1 LUIS JARAMILLO MSC

Atrapado en las cadenas...

- El algoritmo suele estabilizarse en regiones con alta probabilidad
- Difícilmente se moverá a otras regiones del espacio
- No producirá una muestra representativa
- Getting stuck*

© Original Artid
Reproduction rights obtainable from
www.CartoonStock.com

UTEC

BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1 LUIS JARAMILLO MSC

Metropolis Coupled Monte Carlo Markov Chain MCMCMC or (MC)³

- Frío frío, caliente caliente
- Todas menos una están *calientes*
- Las cadenas *hot* permiten cambios más drásticos que la *cold*

BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1 LUIS JARAMILLO MSC

Y como se obtiene el árbol?

- Usar los árboles que tienen mayor probabilidad (*higher hills*)
- Normalmente se eligen los árboles presentes cuando las cadenas atingen a la estabilidad

BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1 LUIS JARAMILLO MSC

peaks

Cold chain

landscape

Heat chains

UTEC

BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1 LUIS JARAMILLO MSC

2 UN POCO DE PRÁCTICA

BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1 LUIS JARAMILLO MSC

BIOINFORMÁTICA Y BIOESTADÍSTICA 2020-1 LUIS JARAMILLO MSC

1. Descargar el programa PAUP* en <http://claytonholland.com/paup4>. Sigue la opción: Win32 version: paup4-4.0a.10.

El formato NEXUS

2. Para cargar un archivo en el programa, este debe tener un formato especial conocido como NEXUS.

3. En el archivo de ejemplo Click con la estructura del formato NEXUS.

Comienza con un análisis de máxima verosimilitud y la palabra Nexus (NEXUS), observar que todos los caracteres que se encuentran entre corchetes son datos recolectados por el programa. Luego observamos que el comando *begin* nos proporciona el inicio del bloque de datos, observar que luego de cada comando aparece el símbolo: *;*. En este bloque se especifica el comando *data* (datos recolectados) y los nombres de los taxa (*taxidb1*), los que deben corresponder a los descritos en el alineamiento. El final del bloque se describe con el comando *end*.

Método FastTree

4. En el análisis más próximo desarrollado por el PAUP*, se presenta por defecto, y corresponde al comando *set criterion=j*.

5. El archivo con formato NEXUS se carga en el PAUP*, observamos la ventana, si el programa encuentra algún error en la entrada de datos con los caracteres no reconocidos.

6. Podemos realizar una búsqueda heurística utilizando el comando *hsearch*, el programa nos muestra una ventana. Desde esa ventana un resumen de los tipos de caracteres recolectados (constantes, variables y posiblemente algunos caracteres), además de algunos datos de cómo se realizó la búsqueda y reconstrucción de los árboles.

7. Para realizar el análisis *bootstrap*, se utiliza el comando *bootstrap* y el comando *stop* para especificar el número de repeticiones a realizar. Así obtenemos un árbol de cada repetición, como se representan los resultados de *bootstrap*. El valor de *bootstrap* se da como frecuencia de la aparición de los apogonismos dentro del árbol. Este número puede ser estimado también observando la columna inferior desde donde vemos los apogonismos de forma automática.

8. El árbol obtenido puede ser salvado y observado con el TreeView con los comandos *save* y *display* del archivo.

UTEC

McBayes

1. Entra al programa `mlayes`. Si estás en `Windows`, recuerda que el archivo `primates.nex` debe estar en la carpeta donde se encuentre el ejecutable de `MLayes`.
2. Tipa el comando `execute primates.nex` para que el programa lea el archivo. Revisa el output y responde: ¿Cuántos organismos (taxón) tiene el archivo?, ¿Cuánto mide el alineamiento?, ¿Son secuencias de DNA?
3. Para obtener ayuda de qué comando utilizar, tipa `help` y el nombre del comando. Por ejemplo, `help lset`, aparecerá este output:

[illegible]

4. Puede modificar el código con el siguiente comando `!set codevertext`.
5. Si el comando anterior funciona, lea este resultado:
`Setting Code to TeXart`
`Successfully net likelihood model parameters`
6. Repetir los pasos 4 y 5 para obtener la modificación de los parámetros.
7. Es posible cambiar los parámetros a priori con el comando `!prior`. ¿Cómo vería los resultados?
8. Los parámetros de la cadena de Markov puede ser vistos con `!help mcmc`.
9. Modificar los parámetros de muestra de generaciones y frecuencia de muestreo. Usamos el comando `mcmc burnin=10000 samplefreq=100`.
10. Utilizaremos el comando `!print` para imprimir, y simulamos en pasos siguientes, y preguntamos si quiere seguir corriendo. Aparecerá la información de la temperatura de las cadenas de Markov.



11. Ya que tenemos 10 000 generaciones las cuales fueron muestreadas cada 100, al final, nuestra muestra será de $10\,000/100 = 100$. De estas 100 generaciones, descartaremos las 25 primeras para generar el árbol usando los comandos `sump burnin=25` y `sumt burnin=25`
12. Salga de MrBayes con el comando `quit`



Conclusiones

- ✓ Un conjunto de cambios posibles de un estado a otro (c.u con su respectiva probabilidad)
- ✓ Normalmente se eligen los árboles presentes cuando las cadenas atingen a la estabilidad
- ✓ Uso de PAUP y MrBayes



Gracias

**HAZ DE CADA DÍA
UNA OBRA MAESTRA**

