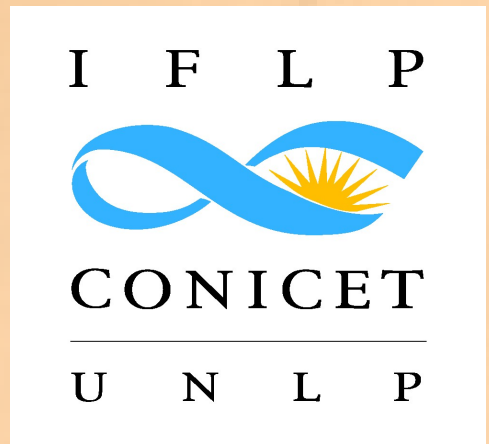




Distribución del número de ancestros en especies de reproducción biparental.

Jarne C. (1); Caruso M. (2)
(1) IFLP - Universidad Nacional de La Plata.
(2) Universidad de Granada - Dpto. Física Teórica y del Cosmos- Campus de Fuentenueva. España

El problema de calcular el número de ancestros de un individuo a la generación t , perteneciente a una especie que se reproduce en forma biparental, tiene una solución estadísticamente determinada. Pues no es posible reconstruir el árbol de sus sucesivos progenitores, que se va formando siguiendo un orden cronológico, partiendo solo del supuesto de que la especie se reproduce de esa manera. Sabemos que no es válida la sencilla progresión del número de ancestros a la generación t dada por 2^{t+1} , ya que al retroceder cronológicamente la probabilidad de que antepasados nuestros sean parientes entre sí aumenta. Esto implica una restricción en el número medio de ancestros respecto de 2^{t+1} . Sin embargo, sabemos que 2^{t+1} corresponde al valor máximo del número de ancestros posible para la t -ésima generación.

Introducción

Considerando la posibilidad de que exista consanguinidad entre los ancestros mostramos de qué manera podemos reconstruir el árbol de progenitores a través de una **Cadena de Markov** en tiempo discreto, embebida en una cadena en tiempo continuo.

Construyendo el modelo:

Definimos $Y(t)$ como la variable aleatoria, de un proceso en tiempo continuo según

$$Y(t) = 2^{t+1} - X(t)$$

de manera que $X(t)$ de cuenta del número de ancestros que son parientes entre sí. El parámetro t da cuenta del lapso de tiempo entre eventos. El proceso real, que es en tiempo discreto, queda embebido en una cadena de tiempo continuo.

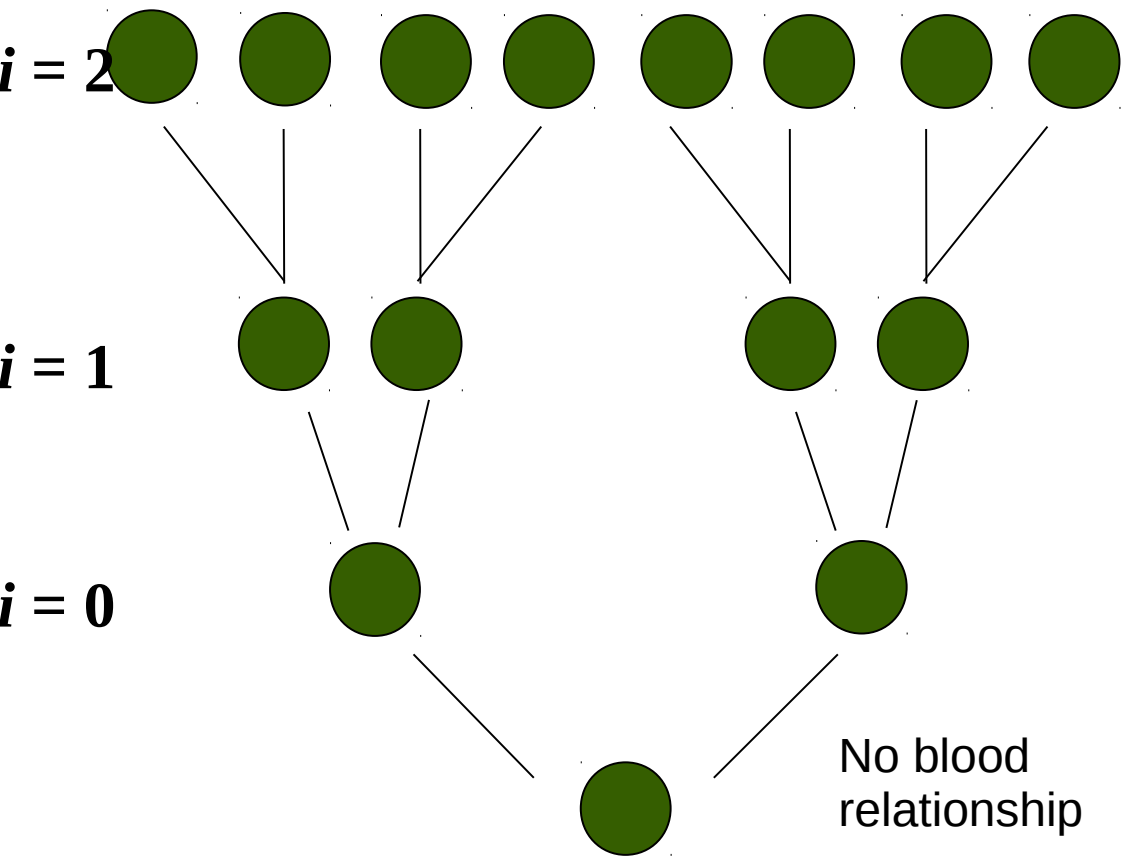
No consideraremos una distribución para cada generación que nos permita dictaminar cuál es el origen de la correlación entre generaciones.

Daremos cuenta de la restricción en el número de ancestros, pero no consideraremos la causa que originó la disminución en $Y(t)$. Esto nos lleva a considerar **procesos homogéneos**.

Formalizaremos el problema, de manera que el valor medio de ancestros tenga la forma:

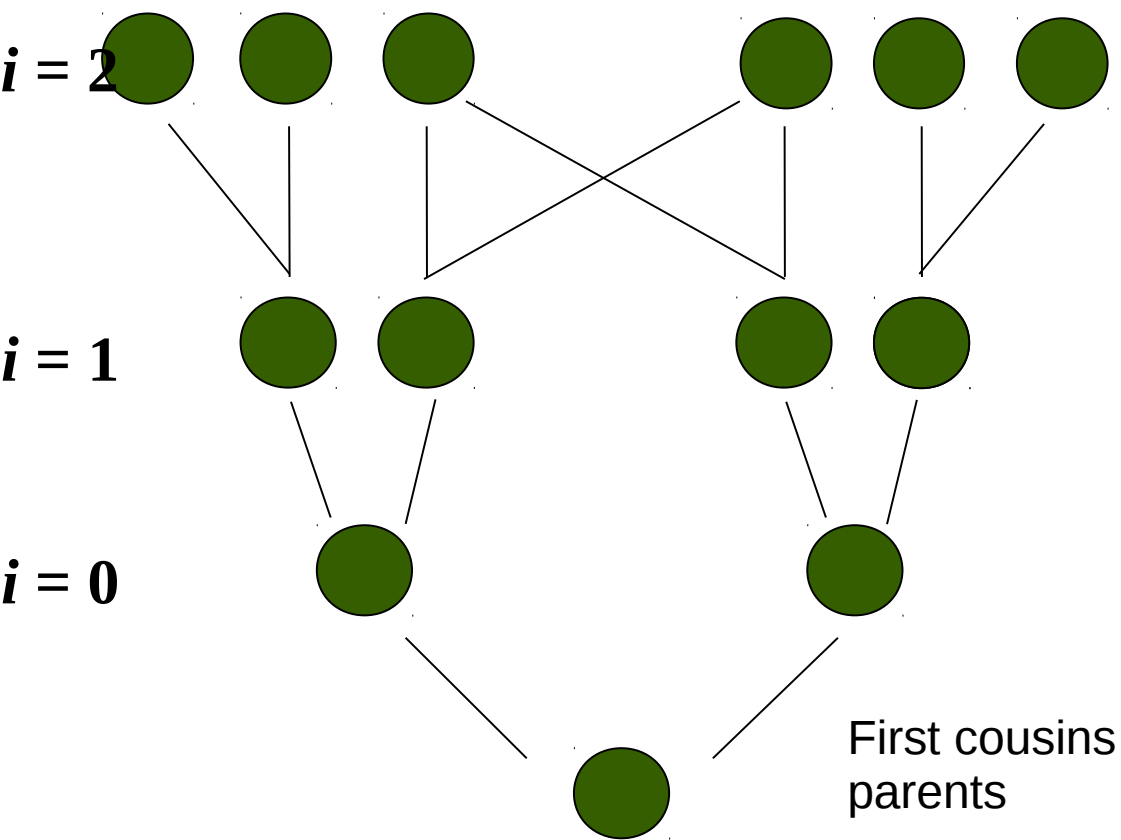
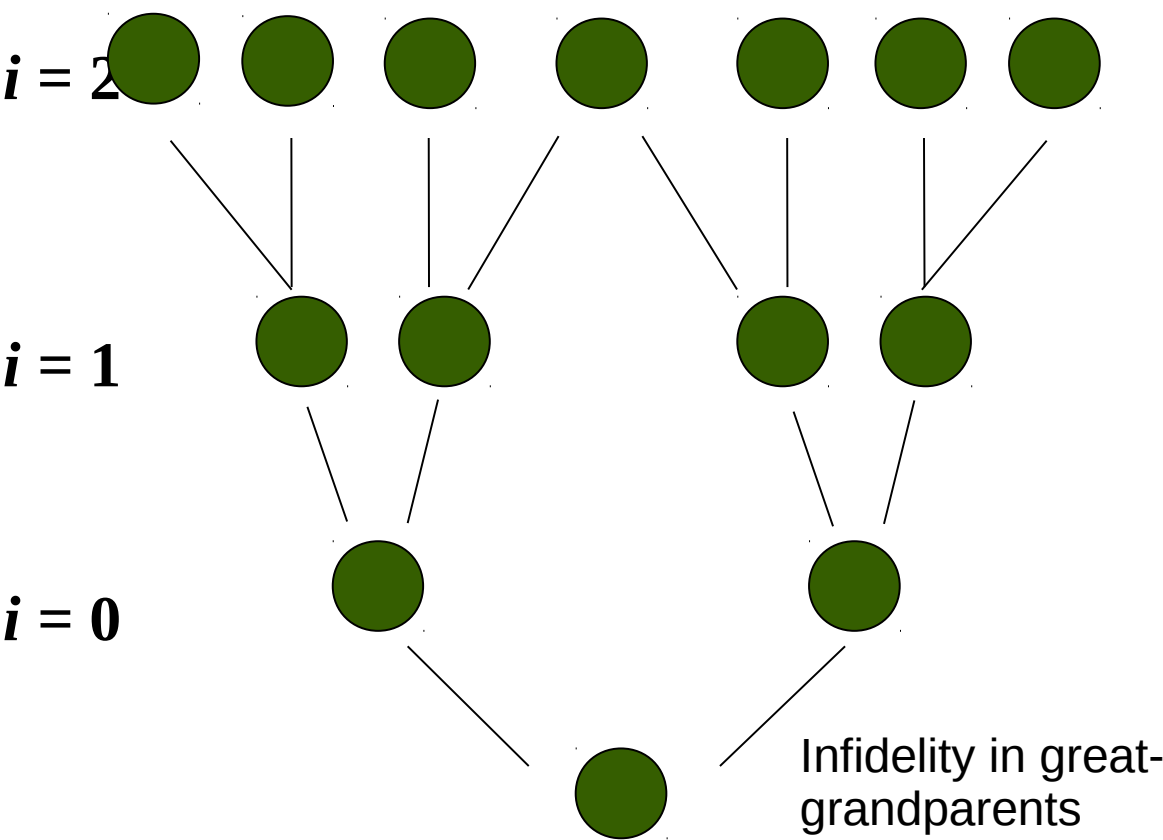
$$\alpha(t) = \langle 2^{t+1} - X(t) \rangle = 2^{t+1} - c(t)$$

donde $c(t)$ es el término que restringe, por consanguinidad, al número de ancestros, originalmente concebido por la pirámide ancestral.



Izquierda: Número máximo posible de ancestros para la tercer generación.

Abajo: Algunos ejemplos de formas posibles de restricción en el número de ancestros.



Hemos considerado el espacio muestral como todo el conjunto natural. Sabiendo que dicho espacio estaba acotado en cada generación y a fin de hallar la escala correcta de la interacción, debemos renormalizar la distribución mediante la transformación:

$$p_n(t) \longrightarrow p_n^*(t) = \Lambda(t)p_n(t)$$

de modo que $p_n(t)$ y $p_n^*(t)$ correspondan al mismo **proceso de Markov**.

El requisito de covarianza queda satisfecho si la transformación es de gauge, esto implica que debemos definir una derivada covariante:

$$\partial_t = d_t - \varpi(t) \quad \text{con:} \quad \varpi(t) = \frac{d_t \Lambda(t)}{\Lambda(t)}$$

Resultados y Conclusiones

A partir de considerar que la restricción principal en el número máximo de individuos se debe a la endogamia, se propuso un modelo estocástico que reduce el número de ancestros en cada generación considerando a cada una relacionada con la siguiente a través de un **proceso de Markov**. Este modelo nos llevo a obtener una distribución para el número de ancestros de la cual pudimos obtener sus cumulantes.

Encontramos un parámetro libre en la teoría que permitiría dar cuenta del tipo de especie que se trate, según sea el número máximo de generaciones que haya vivido de dicha especie.

La distribución renormalizada será:

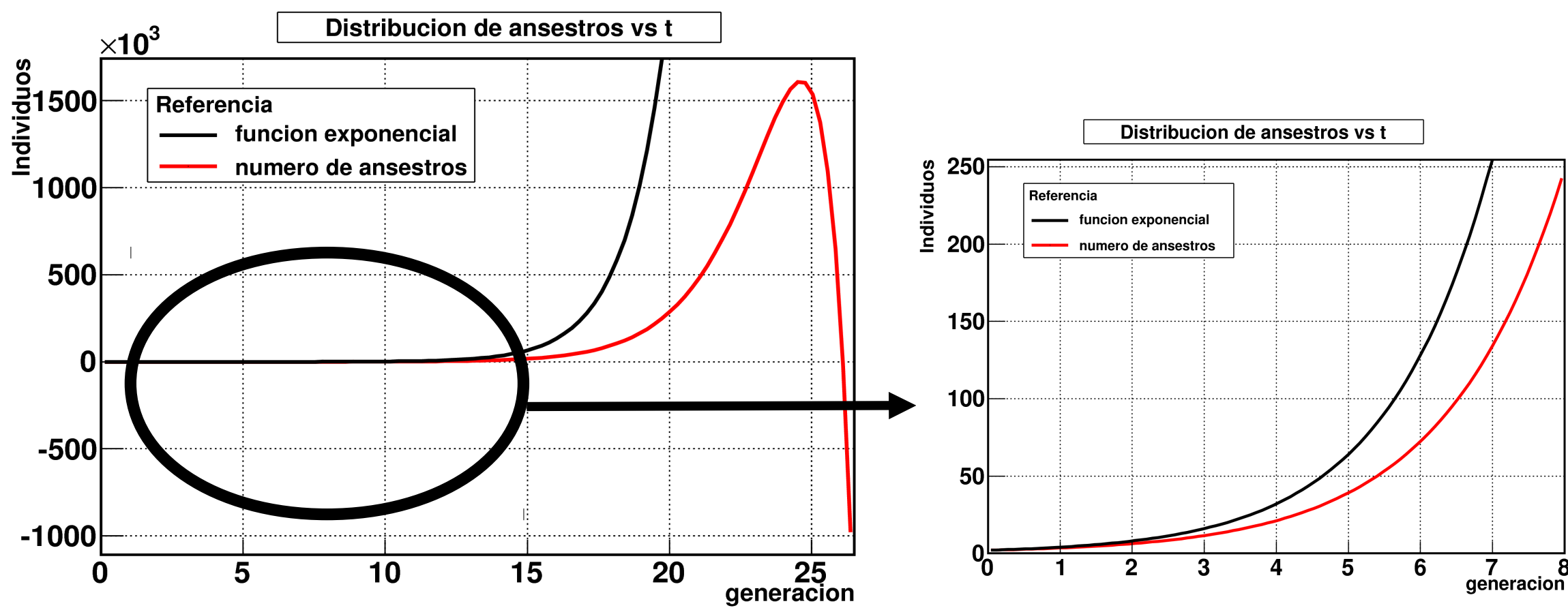
$$p_n^*(t) = \lambda^{t+a} e^{-2\mu t} [I_n(2\mu t) + I_{n+1}(2\mu t)]$$

El valor medio del número de ancestros:

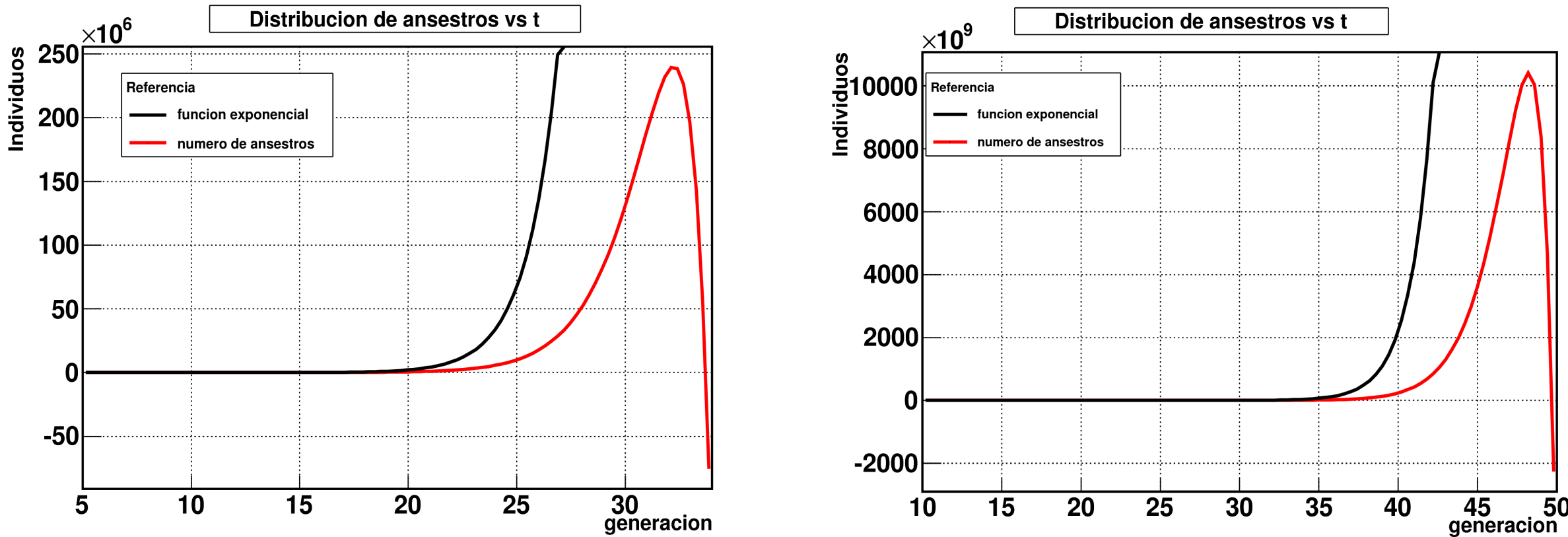
$$\alpha(t) = 2^{t+1} - \lambda^{t+a} \left\{ e^{-2t} [2t I_1(2t) + (2t + \frac{1}{2}) I_0(2t)] - \frac{1}{2} \right\}$$

Con $\lambda=2$ y a un parámetro que cambia el alcance de la distribución.

Valor medio de la distribución obtenida en función de la generación



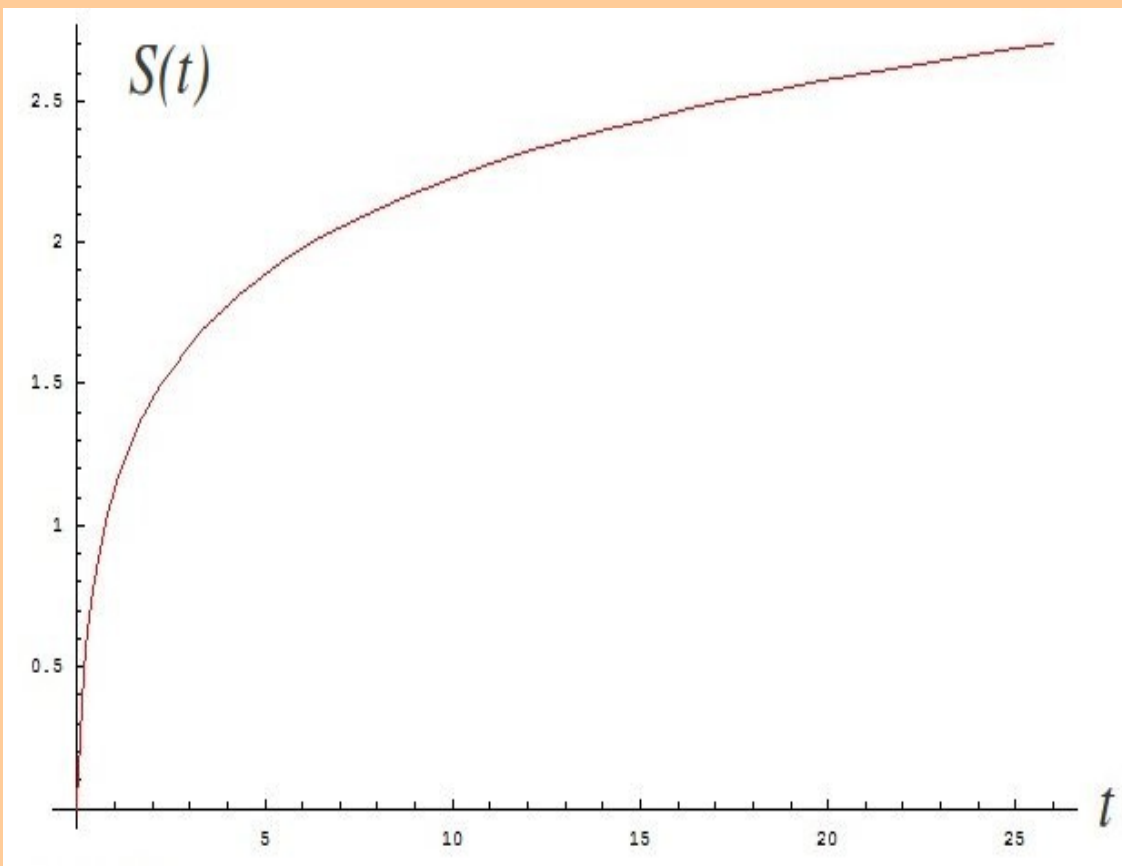
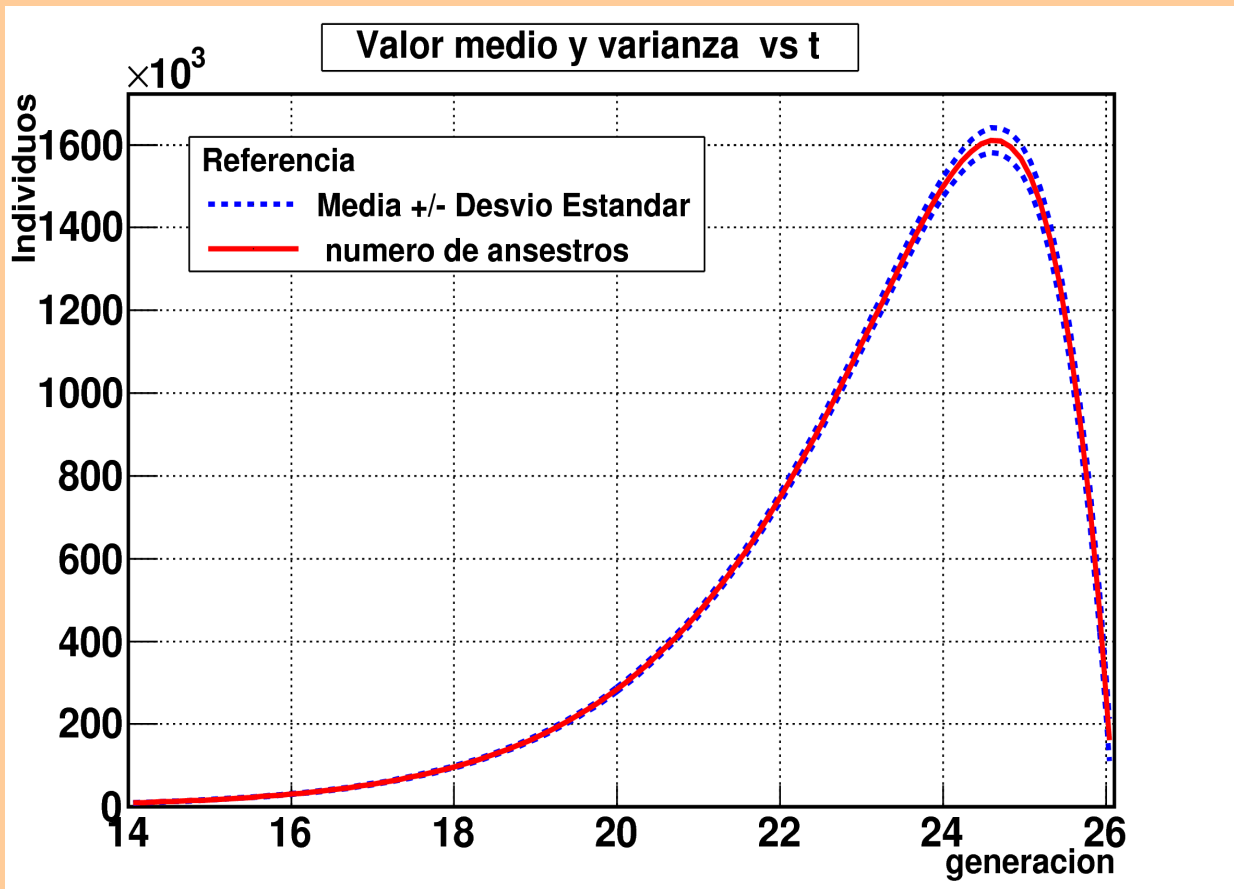
Variando el parámetro a se obtienen las distintas curvas:



Varianza y Entropía de la Distribución

$$\sigma(t) = \lambda^{t+a} \left\{ 2t + \frac{1}{2} - e^{-2t} [2t I_1(2t) + (2t + \frac{1}{2}) I_0(2t)] - c_n^2(t) \right\}$$

$$\mathcal{S}(t) = \sum_n p_n(t) \ln[p_n(t)]$$



Izq.: Se muestra aquí el valor medio +/- su desviación estándar.

Der.: La entropía asociada a la dinámica de ancestros presenta un comportamiento creciente hacia generaciones pasadas, asociado al crecimiento monótono del valor medio de $X(t)$.

Referencias:

- [1] Masaaki Kijima, Markov Processes for Stochastic Modeling, Chapman & Hall (1997)
- [2] Leonard Kleinrock, Queueing Theory, Volume I, Wiley-Interscience Publication (1975)
- [3] M. Abramowitz, I. A. Stegun, Handbook of mathematical functions, Dover books on mathematics(1972)
- [4] T. M. Apostol, Mathematical analysis, Addison-Wesley Pub. Co., (1974)
- [5] William Feller, An introduction to probability theory and its applications, Volume I, Willey (1971)

Agradecimientos:

En especial deseamos agradecer a nuestros directores quienes nos apoyaron en este trabajo especialmente porque no se relaciona directamente con nuestras áreas de trabajo de Doctorado. Un agradecimiento especial a Gabriel y Micaela quienes apoyan cada día, a Santiago Caruso por prestarnos la imagen, y finalmente a todos los que participaron discutiendo con nosotros y enriqueciendo este trabajo.