The distribution of genome shared identical by descent for a pair of full sibs by means of the continuous time Markov chain
Hongki Julie, Udjianna S. Pasaribu, and Adi Pancoro

Citation: AIP Conference Proceedings 1692, 020007 (2015); doi: 10.1063/1.4936435
View online: http://dx.doi.org/10.1063/1.4936435
View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/1692?ver=pdfcov
Published by the AIP Publishing

Articles you may be interested in
Efficient maximum likelihood parameterization of continuous-time Markov processes

The evolution of the market share of competing brands modeled with Markov chains

Calculating Mean First Passage Times from Markov Models of Proteins

Graph transformation method for calculating waiting times in Markov chains

An adjustable aperiodic model class of genomic interactions using continuous time Boolean networks (Boolean delay equations)
Chaos 13, 1167 (2003); 10.1063/1.1608671
The Distribution of Genome Shared Identical By Descent For A Pair Of Full Sibs By Means Of The Continuous Time Markov Chain

Hongki Julie\(^1\), Udjianna S. Pasaribu\(^2\), and Adi Pancoro\(^3\)

\(^1\)Mathematic Education Department, Sanata Dharma University.
\(^2\)Mathematics Department, Institute Technology of Bandung.
\(^3\)Biology Department, Institute Technology of Bandung.
* Email: hongkijulie@yahoo.co.id

Abstract. This paper will allow Markov Chain’s application in genome shared identical by descent by two individual at full sibs model. The full sibs model was a continuous time Markov Chain with three state. In the full sibs model, we look for the cumulative distribution function of the number of sub segment which have 2 IBD haplotypes from a segment of the chromosome which the length is \(t\) Morgan and the cumulative distribution function of the number of sub segment which have at least 1 IBD haplotypes from a segment of the chromosome which the length is \(t\) Morgan. This cumulative distribution function will be developed by the moment generating function.

Keywords: Full Sibs, the continuous time of Markov Chain, the moment generating function, and the cumulative distribution function.

INTRODUCTION

The genes at a given locus of two related individuals are said to be identical by descent (IBD) if one is a physical copy of the other or both are physical copies of the same gene in a common ancestor, i.e. Both are copies that arose by DNA replication from the same ancestral sequence without any intervening mutation.

In 1983, Donnelly (Stefanov, 2002) calculates the probability that individuals in a given relationship share any part of the genome IBD by means of the continuous time of Markov Chain, whose states are the vertices of a hypercube. This research was considered by Stefanov (2000) as the first concerning IBD calculation, in the framework of the continuum model. Stefanov examine more deep than Donnelly and give to us a methodology to find of the cumulative distribution function of the proportion of genome shared identical by descent on chromosome segments by two individuals in a grandparent-type relationship (2000) and a pair of full or half sibs (2002). Stefanov’s methodology could be applied because we had data about distance between two locus for each chromosome. This data was produced by Human Genome Project (HGP). One of the HGP’s goals was mapped and sequenced human’s genome. This mapping is very useful to know what kind of gene in human’s chromosome and the expression from that gene. This paper adapted Markov Chain model from Stefanov (2002). Stefanov found the cumulative distribution function from the infimum of the length chromosome segment by the characteristic function, but in this paper we found cumulative distribution function by the moment generating function.

FULL SIBS

According to Stefanov (2002), for a pair of full sibs the relevant hypercube is four dimensional and the coordinates are either 0 or 1. 0 denote a grand paternal DNA and 1 denote a grand maternal DNA. The first two coordinates indicate the parental transmissions for sib one and the other two do the same for sib two. So, first and third coordinates denote DNA which were generated by their mother and second and fourth coordinates denote DNA which were generated by their father. The DNA at a locus or a segment from a homologous chromosome is called haplotype. From that coding process we will get 16 states of the hypercube. The sixteen states of the hypercube can be divided into three groups of vertices indicating whether 0, 1, or 2 haplotypes are
shared IBD at a locus. So, this case can be analyzed by a continuous time Markov Chain with three states and \( i,j = 0, 1, 2 \). The states are denoted by 0, 1, and 2 corresponding to the number of shared IBD haplotype. Cross over was occurred for this case can be looked at Figure 1.

<table>
<thead>
<tr>
<th>vertices in hypercube</th>
<th>number of shared IBD haplotype</th>
<th>vertices in hypercube</th>
<th>number of shared IBD haplotype</th>
<th>vertices in hypercube</th>
<th>number of shared IBD haplotype</th>
<th>vertices in hypercube</th>
<th>number of shared IBD haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 1, 1</td>
<td>2</td>
<td>0 1, 1 1</td>
<td>1</td>
<td>1 0, 0 1</td>
<td>0</td>
<td>0 1, 0 0</td>
<td>1</td>
</tr>
<tr>
<td>1 1, 1</td>
<td>1</td>
<td>1 1, 0 0</td>
<td>0</td>
<td>0 1, 0 1</td>
<td>2</td>
<td>0 0, 1 0</td>
<td>1</td>
</tr>
<tr>
<td>1 1, 0</td>
<td>1</td>
<td>1 0, 1 0</td>
<td>2</td>
<td>0 0, 1 1</td>
<td>0</td>
<td>0 0, 0 1</td>
<td>1</td>
</tr>
<tr>
<td>1 0, 1</td>
<td>1</td>
<td>0 1, 1 0</td>
<td>0</td>
<td>1 0, 0 0</td>
<td>1</td>
<td>0 0, 0 0</td>
<td>2</td>
</tr>
</tbody>
</table>

**Figure 1.** Sixteen possibility a copy gene at one locus for a pair of full sibs.

In full sibs case, the continuous data on a chromosomal segment consists of the lengths of the consecutive pieces (subsegments) characterized by the number of haplotype shared IBD. Let \( t \) be the length (in Morgan) of a chromosome segment of interest. The sojourn time in state \( i \) (\( i = 0, 1, 2 \)), within time interval of length \( t \), is the length of genome whose each location has \( i \) haplotype shared IBD by two sibs on that segment. Such a genome will be called briefly a genome with \( i \) haplotype IBD. The sojourn time in state 2, within time interval of length \( t \), is the length of genome whose each location has 2 haplotype shared IBD by two sibs on that segment.

The number of entries to state \( i \), within time interval of length \( t \), is related to the number of distinct pieces (sub segment), of a chromosomal segment of length \( t \), on each of which exactly \( i \) haplotype are shared IBD.

The number of entries from state 0 to state \( i \), within time interval of length \( t \), is related to number of sub segment which have at least 1 IBD haplotypes from a chromosomal segment which the length is \( t \) Morgan. The number of entries from state \( i \) to state 2, within time interval of length \( t \), is related to number of sub
segment which have 2 IBD haplotypes from a chromosomal segment which the length is \( t \) Morgan.

\( N_0(t) \) denotes the number of one step transition from state \( i \) to state \( j \), \( i, j = 0, 1, 2, \) and \( A(t) \) denote the initial state.

a. \( N_{ij}(t) \), given the initial state is \( i \), counts the number of sub segment which have 2 IBD haplotypes from a chromosomal segment which the length is \( t \) Morgan.

b. \( N_{0ij}(t) \), given the initial state is \( 0 \), counts number of sub segment which have at least 1 IBD haplotypes from a chromosomal segment which the length is \( t \) Morgan.

The relationship between \( \tau_k, y/j \) and \( H_{2FS}(t) \) and \( H_{1,2FS}(t) \) as follow:

a. \( H_{2FS}(t) > k \iff (\tau_{k+1,12}/i \leq t, A(0)=0) \) or \((\tau_{k+1,12}/i \leq t, A(0)=1) \) or \((\tau_{k,12}/i \leq t, A(0)=2) \);

b. \( H_{1,2FS}(t) > k \iff (\tau_{k,01}/i \leq t, A(0)=0) \) or \((\tau_{k,01}/i \leq t, A(0)=1) \) or \((\tau_{k,01}/i \leq t, A(0)=2) \);

Consequence, cumulative distribution function of \( H_{2FS}(t) \) and \( H_{1,2FS}(t) \) related to cumulative distribution function of \( \tau_k, y/j \) as follows:

\[
P(H_{2FS}(t) \leq k) = 1 - \frac{1}{2} \sum_{i=0}^{k} \left[ P(A(0)=i)P(A(0)=i) - P(A(0)=2)P(A(0)=2) \right], k = 0, 1, \ldots \quad (1)
\]

\[
P(H_{1,2FS}(t) \leq k) = 1 - \frac{1}{2} \sum_{i=0}^{k} \left[ P(A(0)=i)P(A(0)=i) - P(A(0)=0)P(A(0)=0) \right], k = 0, 1, \ldots \quad (2)
\]

**Proposition 1:**

\[
M_{\tau_{12},y/j}(y) = M_{\tau_{12},j}(y) = \frac{1}{2 \left( 1 - \frac{y^2}{4} \right) - 1^k}, k = 1, 2, \ldots
\]

\[
M_{\tau_{10},y}(y) = \frac{1}{2 \left( 1 - \frac{y^2}{4} \right) - 1}, k = 1, 2, \ldots
\]

**Proposition 2:**

\[
M_{\tau_{12},1}(y) = \left\{ \frac{1}{2 \left( 1 - \frac{y^2}{4} \right) - 1^k}, k = 1, 2, \ldots \right\}
\]

\[
M_{\tau_{10},1}(y) = \frac{1}{2 \left( 1 - \frac{y^2}{4} \right) - 1^k}, k = 1, 2, \ldots
\]

Proof: (look Hongki, 2005).

So, if we want to find cumulative distribution function of \( H_{2FS}(t) \) and \( H_{1,2FS}(t) \), we can be used inversion formula for Fourier Transformation. We can get the value of cumulative distribution function of \( \tau_k, y/j \) for \( t = 0.5 \) M, \( t = 1.75 \) M, \( t = 1.80 \) M, \( t = 1.85 \) M, and \( t = 1.90 \) M with Maple programs.
Table 2. $F_{H_t^{(t)}(k)}$, for $t = 0.5 \text{ M}$, $t = 1.75 \text{ M}$, $t = 1.80 \text{ M}$, $t = 1.85 \text{ M}$, and $t = 1.90 \text{ M}$

<table>
<thead>
<tr>
<th>$k$</th>
<th>$F_{H_t^{(0.5)}(k)}$</th>
<th>$F_{H_t^{(1.75)}(k)}$</th>
<th>$F_{H_t^{(1.80)}(k)}$</th>
<th>$F_{H_t^{(1.85)}(k)}$</th>
<th>$F_{H_t^{(1.90)}(k)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.4062679319</td>
<td>0.0937667072</td>
<td>0.0884317309</td>
<td>0.0834003023</td>
<td>0.0786551480</td>
</tr>
<tr>
<td>1</td>
<td>0.8588306426</td>
<td>0.3646658317</td>
<td>0.3501710083</td>
<td>0.3361450751</td>
<td>0.3228516053</td>
</tr>
<tr>
<td>2</td>
<td>0.9856720669</td>
<td>0.6842471426</td>
<td>0.6682925538</td>
<td>0.6523555050</td>
<td>0.6364619108</td>
</tr>
<tr>
<td>3</td>
<td>0.992529136</td>
<td>0.8902490552</td>
<td>0.8806849178</td>
<td>0.8707691154</td>
<td>0.8605206715</td>
</tr>
<tr>
<td>4</td>
<td>0.999769183</td>
<td>0.9727891255</td>
<td>0.9692220907</td>
<td>0.9653683926</td>
<td>0.9612234846</td>
</tr>
<tr>
<td>5</td>
<td>0.99995336</td>
<td>0.9950446667</td>
<td>0.9941462791</td>
<td>0.9931341749</td>
<td>0.9919917599</td>
</tr>
<tr>
<td>6</td>
<td>0.999999933</td>
<td>0.9993172633</td>
<td>0.9991553644</td>
<td>0.998637240</td>
<td>0.9987385233</td>
</tr>
<tr>
<td>7</td>
<td>0.9999999999</td>
<td>0.999268937</td>
<td>0.999050922</td>
<td>0.998780027</td>
<td>0.9998446382</td>
</tr>
<tr>
<td>8</td>
<td>0.999993716</td>
<td>0.999915021</td>
<td>0.999885383</td>
<td>0.999847074</td>
<td>0.99999753</td>
</tr>
<tr>
<td>9</td>
<td>0.999995961</td>
<td>0.999993815</td>
<td>0.999991237</td>
<td>0.999987736</td>
<td>0.99999954</td>
</tr>
<tr>
<td>10</td>
<td>0.999999975</td>
<td>0.999999628</td>
<td>0.999999445</td>
<td>0.999999185</td>
<td>0.99999998</td>
</tr>
<tr>
<td>11</td>
<td>0.999999988</td>
<td>0.999999981</td>
<td>0.999999970</td>
<td>0.9999999954</td>
<td>0.999999998</td>
</tr>
<tr>
<td>12</td>
<td>0.999999999</td>
<td>0.999999999</td>
<td>0.999999999</td>
<td>0.999999999</td>
<td>0.999999999</td>
</tr>
<tr>
<td>13</td>
<td>0.999999999</td>
<td>0.999999999</td>
<td>0.999999999</td>
<td>0.999999999</td>
<td>0.999999999</td>
</tr>
</tbody>
</table>

Table 3. $F_{H_t^{(t)}(k)}$, for $t = 0.5 \text{ M}$, $t = 1.75 \text{ M}$, $t = 1.80 \text{ M}$, $t = 1.85 \text{ M}$, and $t = 1.90 \text{ M}$

<table>
<thead>
<tr>
<th>$k$</th>
<th>$F_{H_t^{(0.5)}(k)}$</th>
<th>$F_{H_t^{(1.75)}(k)}$</th>
<th>$F_{H_t^{(1.80)}(k)}$</th>
<th>$F_{H_t^{(1.85)}(k)}$</th>
<th>$F_{H_t^{(1.90)}(k)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0338338209</td>
<td>0.0002279706</td>
<td>0.0001866465</td>
<td>0.0001528133</td>
<td>0.0001251129</td>
</tr>
<tr>
<td>1</td>
<td>0.74468221</td>
<td>0.1870747438</td>
<td>0.1764901686</td>
<td>0.1664947980</td>
<td>0.157060704</td>
</tr>
<tr>
<td>2</td>
<td>0.9727930630</td>
<td>0.5422541900</td>
<td>0.5238518480</td>
<td>0.5057951721</td>
<td>0.488103400</td>
</tr>
<tr>
<td>3</td>
<td>0.9985517070</td>
<td>0.8262409555</td>
<td>0.8127332598</td>
<td>0.7989158381</td>
<td>0.7848026818</td>
</tr>
<tr>
<td>4</td>
<td>0.9995457565</td>
<td>0.9542580148</td>
<td>0.9486365761</td>
<td>0.9362206610</td>
<td>0.9362204363</td>
</tr>
<tr>
<td>5</td>
<td>0.9999908012</td>
<td>0.9931202631</td>
<td>0.9988078433</td>
<td>0.988127924</td>
<td>0.987240363</td>
</tr>
<tr>
<td>6</td>
<td>0.9999999868</td>
<td>0.9987690196</td>
<td>0.9984874147</td>
<td>0.9981501667</td>
<td>0.997594838</td>
</tr>
<tr>
<td>7</td>
<td>0.9999999999</td>
<td>0.9999854296</td>
<td>0.9999026104</td>
<td>0.9997772814</td>
<td>0.9997175629</td>
</tr>
<tr>
<td>8</td>
<td>0.9999999999</td>
<td>0.999993716</td>
<td>0.9999814704</td>
<td>0.9999787240</td>
<td>0.9999717333</td>
</tr>
<tr>
<td>9</td>
<td>0.999999961</td>
<td>0.999988338</td>
<td>0.999983526</td>
<td>0.999977016</td>
<td>0.99999164</td>
</tr>
<tr>
<td>10</td>
<td>0.999999953</td>
<td>0.999999929</td>
<td>0.999998948</td>
<td>0.999998456</td>
<td>0.999999914</td>
</tr>
<tr>
<td>11</td>
<td>0.999999988</td>
<td>0.999999964</td>
<td>0.999999994</td>
<td>0.999999997</td>
<td>0.999999999</td>
</tr>
<tr>
<td>12</td>
<td>0.999999999</td>
<td>0.999999999</td>
<td>0.999999999</td>
<td>0.999999999</td>
<td>0.999999999</td>
</tr>
<tr>
<td>13</td>
<td>0.999999999</td>
<td>0.999999999</td>
<td>0.999999999</td>
<td>0.999999999</td>
<td>0.999999999</td>
</tr>
</tbody>
</table>

CONCLUSION

1. The sixteen states of the hypercube can be divided into three groups of vertices indicating whether 0, 1, or 2 haplotypes are shared IBD at locus. So, we can be analyzed cross over what happen at a pair of full sibs with a continuous time Markov Chain with three states.

2. The relationship between cumulative distribution function of $H_t^{FS}(t)$ and $H_t^{FS}_{1,2}(t)$ with cumulative distribution function of $S_t^{g/g}$ as follows:

$$P(H_t^{FS}(t) \leq k) = 1 - \sum_{i=0}^{k} P(\tau_{t,i}^{1,2} \leq t \mid A(0) = i) \cdot P(A(0) = i) - P(\tau_{t,i}^{1,2} \leq t \mid A(0) = 2) \cdot P(A(0) = 2), \quad k = 0, 1, ...$$

$$P(H_t^{FS}_{1,2}(t) \leq k) = 1 - \sum_{i=0}^{k} P(\tau_{t,i}^{1,2} \leq t \mid A(0) = i) \cdot P(A(0) = i) - P(\tau_{t,i}^{1,2} \leq t \mid A(0) = 0) \cdot P(A(0) = 0), \quad k = 0, 1, ...$$
3. At a pair of full sibs, for \( k_1 = k_2 \), but \( t_1 < t_2 \),
then \( F_{H_2}^{k_1}(t_1) > F_{H_2}^{k_2}(t_2) \) and
\( F_{H_1}^{k_1}(t_1) > F_{H_1}^{k_2}(t_2) \).

REFERENCES