Assume that preservation within the history of a species occurs at a stochastically constant rate $r$ per species per million years. Over a span of time $T$, the probability of exactly $k$ instances of preservation follows a Poisson distribution:

$$Pr\{k; r, T\} = e^{-rT}(rT)^k / k!.$$ 

Now consider the spacing (gaps) between events, where the exact number of events is not specified.

1. What is the probability, $Pr\{\geq x\}$, that a gap will be at least as long as $x$?

2. What, therefore, are the distribution and density functions that describe gap size?

3. What is the expected (mean) gap size?

4. What is the median gap size?

5. What is the variance in gap size?

6. How might you test whether the fossilization process in a particular empirical example is indeed stochastically constant as assumed?
GEOS 33000/EVOL 33000

**Exercise 2**  (due January 19th)

Write a computer program to simulate a random walk of $n$ steps, in which a single “morphological” variable increases (with probability $p_i$) or decreases (with probability $p_d$) at each step.

Let $p_i = 0.6$ and let $n = 100$. Run the simulation 10,000 times. For each simulation, perform the number-of-positive increments test and determine whether you can reject the null hypothesis that $p_i = p_d = 0.5$.

For $n = 100$, how large must $p_i$ be for you to be reasonably certain of rejecting the null hypothesis?

For $p_i = 0.6$, how large must $n$ be for you to be reasonably certain of rejecting the null hypothesis?

Discuss the significance of your results.

Exercise 3  (due January 26th)

1. Write a computer program to simulate a time-homogeneous birth-and-death process with discrete time steps. Start with a single lineage. At each time step allow the lineage to branch (with probability $p$), become extinct (with probability $q$) or persist unchanged (with probability $1 - p - q$).

2. Set $p = q = 0.1$ and allow the simulation to run for $t = 50$ time steps. Consider the number of lineages, $n_t$, extant at time $t$, where $n_t$ can be equal to zero. Run the simulation enough times to determine the mean, median, and variance of $n_t$ for $t = 50$. Compare this with the exact probability distribution (see Raup 1985).

3. Repeat (2), but this time consider only those runs in which the clade is still extant at time $t = 50$ (i.e. in which $n_t > 0$). In other words, develop the probability distribution conditioned on the survival of the clade to time $t$. Compare your simulation results with the exact distribution.

4. Write a program to simulate the homogeneous branching process in continuous time. Start with a single lineage and draw its duration at random from the appropriate distribution. Having drawn its duration, draw the number of direct descendants from the appropriate distribution. Having drawn the number of descendants, draw their times of origin from the appropriate distribution. Repeat until all lineages—the initial one as well as its direct and indirect progeny—are accounted for.

Exercise 4  (due February 2nd)

Consider the following genus-level cohort survivorship table:

<table>
<thead>
<tr>
<th>Interval of origin</th>
<th>Cohort size</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100</td>
<td>73</td>
<td>56</td>
<td>42</td>
<td>32</td>
<td>27</td>
<td>21</td>
<td>18</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>60</td>
<td>47</td>
<td>39</td>
<td>32</td>
<td>27</td>
<td>24</td>
<td>22</td>
<td>20</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>74</td>
<td>60</td>
<td>47</td>
<td>38</td>
<td>27</td>
<td>26</td>
<td>24</td>
<td>22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>71</td>
<td>55</td>
<td>43</td>
<td>37</td>
<td>32</td>
<td>27</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td>71</td>
<td>64</td>
<td>49</td>
<td>40</td>
<td>35</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>68</td>
<td>53</td>
<td>43</td>
<td>37</td>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Write a program to obtain (numerically) the maximum-likelihood estimate of the species-level origination and extinction rates, \( p \) and \( q \). Apply to these data. Graphically depict the likelihood surface and comment on the significance of its shape. (Assume all intervals have the same length, and express rates per lineage-interval. Assume origins are at lower boundaries of intervals.)