The statistical appendix to *People v. Collins*:  
*A fallacy in the use of the Poisson distribution*

The statistical appendix to *People v. Collins* makes three arguments for overturning the conviction:
i) There was no foundation for the individual probabilities given;
   ii) Even if the individual probabilities were correct, the possible lack of independence of the events 
   listed would invalidate the calculation of their simultaneous occurrence by simply multiplying the 
   probabilities of the individual events, so one could not conclude that the probability of their 
   simultaneous occurrence was 1/12,000,000.
   iii) Even if the true probability of the simultaneous occurrence of all the events in question was 
   1/12,000,000 the fact that the Los Angeles metropolitan area had (then) a popualtion of about 
   4,000,000 created a reasonable likelihood that another couple fitting the descriptions of Janet and 
   Malcolm Collins existed in the area, which would negate the crieterion of "proof beyond a reasonable 
   doubt" necessary for a criminal conviction.

This last argument was based, as explained below, on a fallacious use of the Poisson distribution.

To simplify matters (as the writer of the appendix did), instead of considering the characteristics 
identifying Janet and Malcolm Collins as applying to a couple, let's suppose that they apply to a single 
individual. Assume, therefore, that the probability of finding an individual with the given characteristics 
is 1/12,000,000. The mean number of such individuals in a population of 4,000,000 would then be 1/3, 
that is, if we took many samples of size 4,000,000 then the average number of individuals with the given 
characteristics would be 1/3. Note that some samples may contain no such individuals and some more 
than one, but if we take many samples, add up the number of matching individuals found and divide by 
the number of samples then this should approach 1/3, in the long run. So let us consider a Poisson 
distribution with mean $\mu = 1/3$.

The probability of seeing exactly $n$ individuals matching the given characteristics in a sample of 
4,000,000 would then be 

$$p(n) = \frac{e^{-\mu} \mu^n}{n!}$$

with $\mu = 1/3$. Suppose that we have picked a sample in which we have already identified 
one individual with the given characteristics. What is the probability that there will be at least a second 
such individual in the sample? Let us write $p(\geq n)$ for the probability of finding at least $n$ matching 
individuals in the sample. Then the probability of finding at least a second matching individual in a 
sample *in which there is already one* is $p(\geq 2)/p(\geq 1) = (1-p(0) - p(1))/(1-p(0))$. (Note that the numerator is 
the probability of *not* finding precisely none or precisely one, hence of finding at least two; the 
denominator is similarly the probability of finding at least one. Why must we divide by this quantity?) 
The approximate value is .157. The writer of the appendix concludes from this that there is a substantial 
probability that there is a second couple in the Los Angeles metropolitan area matching the 
characteristics ascribed to Janet and Malcolm Collins by the witnesses.

The fallacy is that what this analysis actually shows is that if there were, somehow, many replicas of the 
Los Angeles area in the universe and one were found with a couple matching the characteristics of Janet 
and Malcolm Collins, then the probability that it contained at least a second such couple would be about 
.157. This is a statement about the collection of the replicas of Los Angeles in this strange universe; it 
says nothing about any *single* one, in particular, the one where the crime occurred.