To: Chunglee Kim, c-kim1@northwestern.edu

From: Angie Hugeback, hugeback@galton.uchicago.edu

Goal:

To estimate the true number of pulsars that exist within the volume where we are able to observe them.

Definition:

$X \sim \text{Binomial}(N, p)$ means that X is a Binomial random variable with N trials and probability of success p. In particular, X is the number of successes obtained out of N trials, where each trial is independent of the others, and each trial has probability p of a success.

Approach:

In your simulation, N is the number of pulsars that you create, and X is the number of those pulsars that are observed by your simulated telescope. You can think of each simulation of a single pulsar as an independent trial, and if that pulsar is observed then you count it as a success. Your data, X, is the number of pulsars that were observed in the simulation. The initial quantity that you are interested in estimating is p, which is the overall probability that an individual pulsar will be observed.

To estimate $\hat{p}$, simply run your simulation for a long time, and then your estimate is $\hat{p} = X/N$. This estimate should be very precise, and you can proceed as if this quantity p is now known. To further justify this argument, a very conservative estimate of the variance for your estimator $\hat{p}$ would be $\frac{\hat{p}(1-\hat{p})}{N}$, which becomes negligible as N gets very large.

Now, the final step is to obtain a confidence interval for the true number of pulsars that exist within the volume where you have the ability to observe them. Call this number M.

Let Y be the actual count of pulsars that have been discovered so far. You can think of Y as binomial, where to be $Y \sim \text{Binomial}(M, \hat{p}$. So you have Y and $\hat{p}$, and you are interested in estimating the actual number of pulsars, M.

You can calculate an "exact binomial confidence interval" for the true count of pulsars, M. Consider the hypothesis test for $H_0 : p = \hat{p}$ versus $H_a : p \neq \hat{p}$, where $Y \sim \text{Binomial}(M, \hat{p})$. A 95% confidence interval for M is (M_{min}, M_{max}) where

- M_{min} = smallest number M such that we will not reject the null hypothesis.
- M_{max} = largest number M such that we will not reject the null hypothesis.

I have written some code in R for you to obtain this confidence interval if your are interested. You can adjust the values for your data:

- the true number of pulsars that were observed (observedCount)
- the observational probability that you obtain from your simulation (pHat).