Fellow: Zoë Bilodeau
Home Institution: Skidmore College

Mentors: Tina Peters (University of Delaware) and Chris Tunnell (Rice University)

Dates: May 16th - August 5th

Fast Integration of Poisson Distributions for Dead Sensor Marginalization

My research will focus on analyzing data collected by the XENONnT dual-phase time projection chamber, which is a type of particle detector. A particle detector is a device used to learn about particles by recording how they interact with other particles, in this case liquid and gaseous xenon. The particle detector used in this research records the position of particles passively traveling through the xenon. The detector uses many different sensors to track photons and their locations in relation to one another by recording the electrical signals the particles trigger, which are their primary scintillation signal and their ionization signal.

With the data we get from the particle detector, we are interested in understanding what interactions are occuring between the different particles to determine what kind of particles flew into the detector with the intention of identifying dark matter. One way we are doing this, and the one I will be working on, is done by looking at the photons generated during collisions between the xenon and particles that are passing through the detector. Figuring out what correlations exist between the number of photons detected by each sensor in the particle detector will help us discern where the collision happened and what kind of interaction occurred between the particles, with the goal that this will allow us to identify dark matter particles.

To find these correlations between particles we will need to perform statistical tests and use machine learning to predict how the photons will act, and the way we are going about doing so is by using a probabilistic graphical model. These models can represent probability distributions over a multiple-dimensional space as we need to be able to do with the data from the particle detector to work on understanding correlations between sensors. Out of the different types of distributions, the one that fits our data best is the Poisson distribution. Poisson distributions are used when the data is the number of times a certain event occurred, or something is being counted in a specific amount of time or space. We are counting the number of photons detected by each sensor in the particle detector, and the distribution of the number of photons detected are well described by a Poisson distribution. Once the data is in the graphical model, we will represent the correlations as multidimensional Poisson distributions.

We are particularly interested in the case where one or more sensors in the detector are turned off or malfunctioning. In order to make inferences from the graph without data from these sensors, we must marginalize over these variables in the graphical model. This requires calculating a numerical integral over a Poisson distribution in one or more dimensions. These computations can be very complex, and as the graph may consider correlations with (at least) the seven sensors adjacent to each sensor, it can be a high dimensional problem. Computing this is unfortunately very costly and would take a long time to do with the amount of data being collected within this experiment. What I will be working on is an algorithm that optimizes this process so that these calculations can be done quickly and not require an incredibly sophisticated computer.

I'll begin writing the algorithm with pseudocode. This is essentially writing out the steps in an algorithm in language that is easy for people to read, so that the focus at this time is designing the algorithm as a whole, which is easier and takes less time than trying to design it while writing out each individual line. More complex algorithms designed to minimize runtime and overall cost, like the one I will work on, tend to require careful planning to pick the best possible strategy and to avoid using unnecessary looping and therefore brainstorming using pseudocode can make the process much easier. After making a design I'm confident in, I'll begin writing and testing code.

The Python packages numba and cython (among others), as well as profiling tools such as snakeviz and cProfile will be used in this project. The final deliverable will be a Python package that can both perform numerical integration of multidimensional Poisson distributions and be integrated into an existing Probabilistic Graphical Model package.

Timeline

- Weeks 1-2: Review related literature of related works to gain a better background on these topics, and begin writing down approaches to the problem.
- Weeks 3-5: Continue writing ideas. Focus on algorithm design and write it out in pseudocode.
- Weeks 6-8: Being implementing and testing final algorithm design.
- Weeks 9-10: Fine-tune program and begin preparing presentation.