

Questions for Discussion

1.1

What are we plotting in the above two plots? What does an individual entry in the blue histogram represent? What does an individual entry in the red histogram represent?

We're plotting histograms of the size of S2 flashes for background-like (a β source) and signal-like events (a neutron source). An entry in the blue histogram represents the number of flashes for a neutron source with a particular flash size, and an entry in the red histogram represents the number of flashes for a β source with a particular flash size.

1.2

The peaks of the two histograms are separated in $\text{Log}(S_2)$ space. What does this separation mean for our ability to distinguish background-like (β) events from signal-like (neutron) events?

This separation means it's theoretically possible to distinguish between background-like and signal-like events.

1.3

What do the widths of the two histograms mean? What does it mean that the histograms overlap a little bit? Hint: this overlap is what was previously referred to as "bleed" in the opening paragraph.

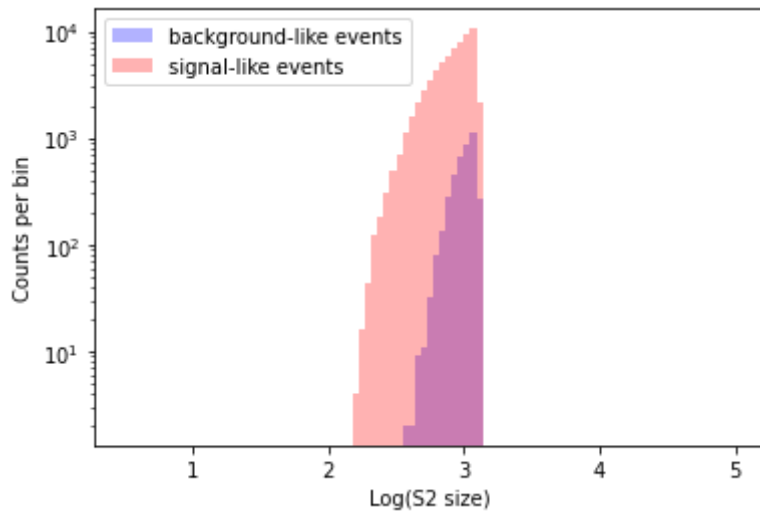
The width of the histograms represents the range of S2 flash sizes. The fact that the histograms overlap means that it'll be difficult to discern whether some events are signal or background.

2.1

Play around with the cut value, X_c , to get a sense of what events remain if you change X_c . Include in your write-up two screenshots of this plot, each

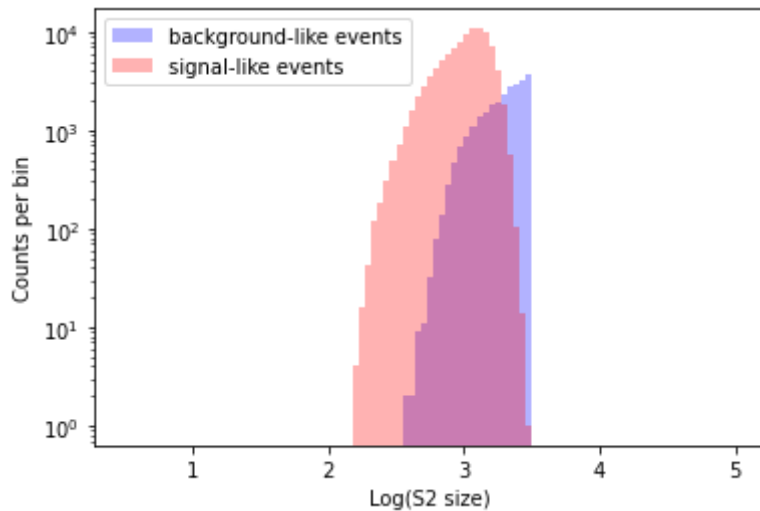
corresponding to a different cut value. Include with both plots the number of signal-like and background-like events remaining for that cut value.

$$X_c = 3.1$$



Number of signal-like event remaining after cut: 65579
Number of background-like events remaining after cut: 3925

$$X_c = 3.5$$



Number of signal-like event remaining after cut: 97713
Number of background-like events remaining after cut: 25120

2.2

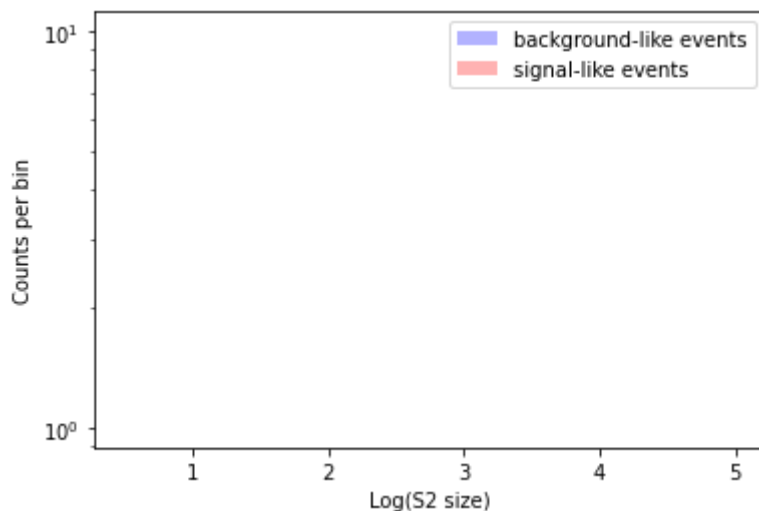
In the end, we will want a cut that removes background events but keeps signal (dark matter) events. If we are expecting our dark matter events to follow a distribution with the same shape as the red histogram, where do you think we might want to set our cut value?

We might want to set it to be around 3.5.

3.1

In the code above, set $X_c=2.0$. How many true signal events do we accept with our cut? How many background events leak into our selection? What is the disadvantage of such a strict cut?

The advantage of this cut is that we leak 0 background events, but the disadvantage is that we also accept 0 true signal events.

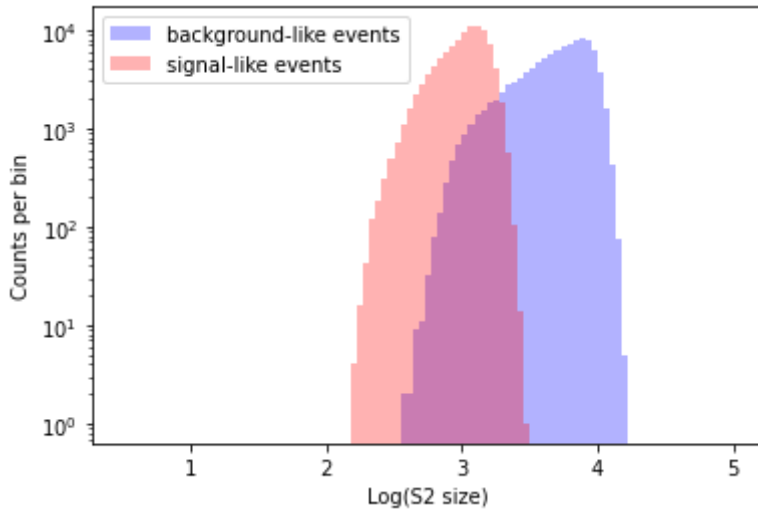


```
Number of signal-like event remaining after cut: 0
Number of background-like events remaining after cut: 0
```

3.2

In the code above, now set $X_c=4.5$. How many true signal events do we accept? How many background events leak into our selection? What is the disadvantage of such a relaxed cut? (Ryan's side note: due to a slight bug in my code this neutron dataset actually has 97,713 events, not 100,000 as originally claimed.)

The advantage of this cut is that we accept all 97713 true signal events, but the disadvantage is that we also leak all 100000 background events.



Number of signal-like event remaining after cut: 97713

Number of background-like events remaining after cut: 100000

4.1

Suppose we set our cut at $X_c=3.5$. Based on these curves, what is the fraction of background events that will survive the cut? What is the fraction of signal events that will survive the cut?

All of the signal events will survive the cut, but 10% of the background events also survive.

4.2

If my dark matter signal has the same distribution shape as the neutron calibration data, but we're only expecting 10 dark matter events total, on average how many dark matter events should survive a cut at $X_c=3.5$?

On average, all 10 dark matter events should survive.

4.3

If my background has the same distribution shape as the background calibration data, but we're expecting 200 background events total, on average how many background events should survive a cut at $X_c=3.5$?

On average, 20 background events should survive.

4.4

Assume a Poisson uncertainty in the surviving background events (i.e. the uncertainty is the square root of the number of surviving events). Given the number you found in 4.3, will a cut at $X_c=3.5$ allow us to confidently say that we've discovered dark matter in this scenario? Why or why not? Is there a way to better improve the sensitivity of our search?

Disclaimer: I'm not sure about my answer to this question.

No, because the number of background events is 20 ± 4.472 so the uncertainty, 4.472, is large in size compared to the number of signal events: 10. I'm not sure if there's a mathematically better way to improve the sensitivity of our search; I think we have to consider measuring something other than the size of S2 flashes.

4.5

Using the answers to the last few questions, I want you to propose a cut value for X_c that you think is best for confidently identifying whether we see dark matter events. Keep in mind that dark matter interactions are rare enough that none have been conclusively detected yet.

I still believe that the 3.5 cutoff is the best idea because it accepts all the signal events, which are exceedingly rare so we don't want to risk missing them.