Analysis notebook for improved $2\nu\beta\beta$ spectra.

The **goal** of this analysis is to compare 2 (slighty) different angular distributions.

The **methodology** used in the analysis is as follows. First, the angular distribution of $2\nu\beta\beta$ is given by the equation: $\frac{d\Gamma}{dcos(\theta)} \sim N(1 + K(\psi_{31}, \psi_{51})cos\theta)$, equation (24) from paper by Odiviu. In the initial stage of the analysis 2 sets of $2\nu\beta\beta$ spectra were simulated in Falaise, $1e8$ simulated events each. First, a spectrum with $K = -1.0$ and second with $K = -0.65$ (a somewhat arbitrary value, close to what is mentioned in table (3) in Ovidiu's paper). The simulated data were passed through flreconstruct and the following data cuts were applied:

1. `Two negatively charged tracks reconstructed,`
2. `Two vertices on the source foil, within given distance from each other,`
3. `Sum of electron energies within the range: ``E_{sum} \in (0, 3500)~keV``,`
4. `Two individual Optical Module hits,`
5. `Two associated Optical Module hits.`

From this two angular distributions were obtained. First, the so-called θ angular distribution, where θ represents the **decay** angle between the two electrons. θ distribution cannot be experimentally measured. Second, the so-called ϕ angular distribution, where ϕ represents the **escape** angle - ie. the angle which can be measured with SuperNEMO, the angle between the two electrons at the moment they escape the source foil.

The two spectra are shown below.

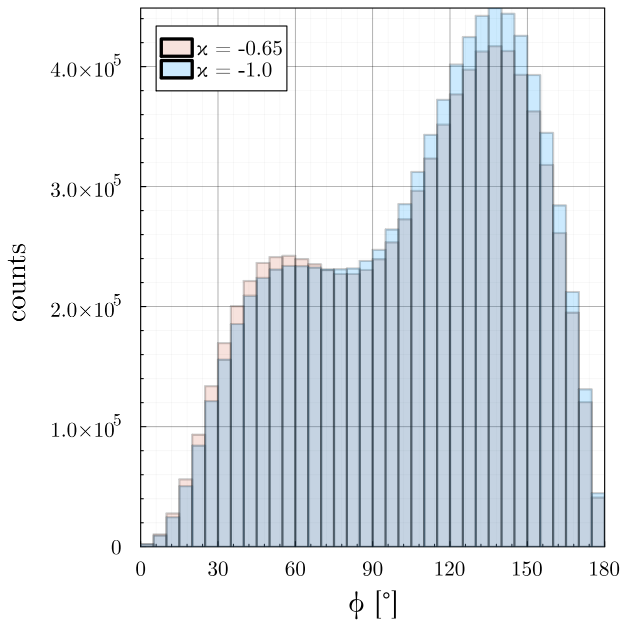
```
Activating project at `~/Work/PhD_Thesis/SNImproved2bbSpectra`
```

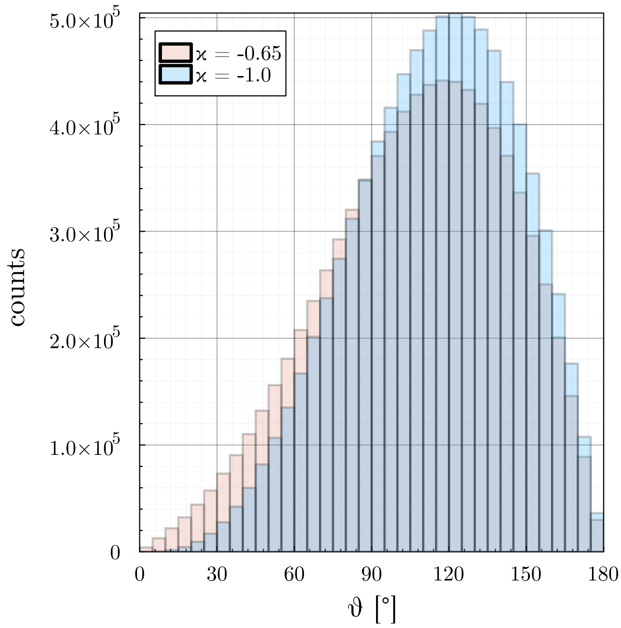
```
"/home/shoram/Work/PhD_Thesis/Job17/Figures/"
```

```
dfs = ▸[]
```

```
files =
▸["/home/shoram/Work/PhD_Thesis/Job17/Data/kappa_-0.65/MomPosEneThetaPhi_97e6E_k_m065.rc
```

```
Δϕ = 5
```



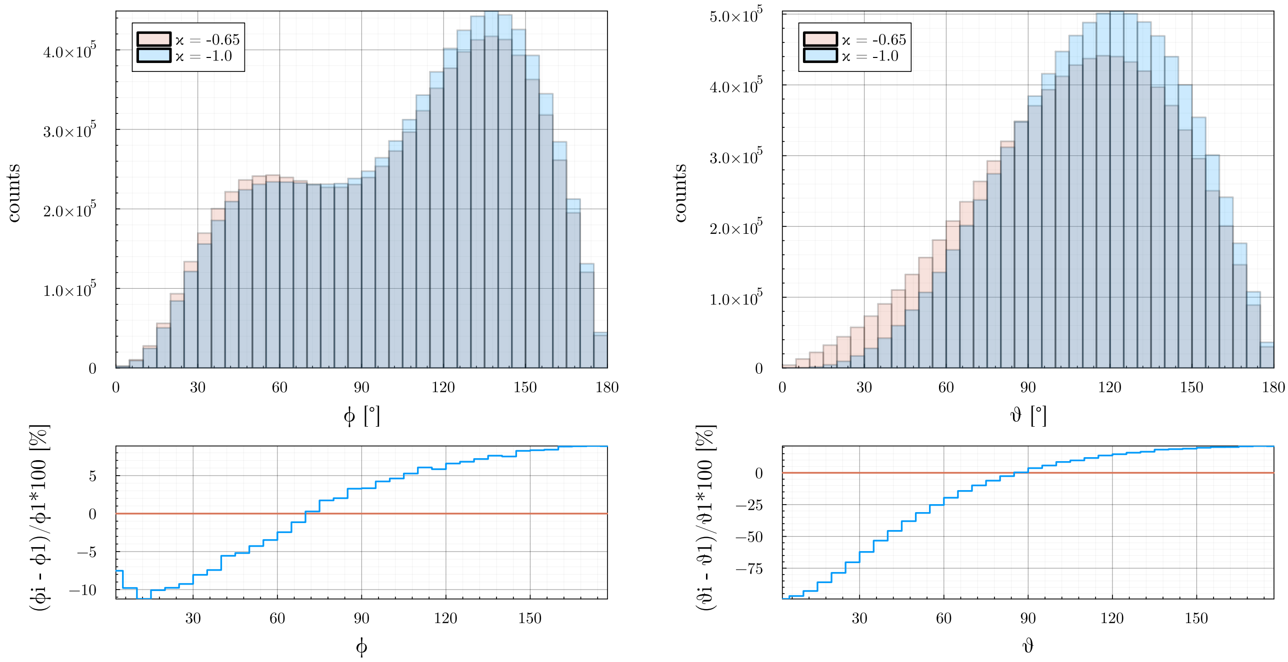


We can see the two angular distributions, ϕ (right), θ (left) for the two values of K, with $K = -0.65$ in *red* and $K = -1.00$ in *blue*.

We can see that, with θ being the input (theoretical) spectrum, there is quite a bit of a drastic change moving toward ϕ distribution. The individual features of the spectrum and hyptheses as to why the shape is as is are outlined in **Detector Effects** chapter. (https://github.com/Shoram444/SNAngularCorrelation/tree/master/DetectorEffects)

To get a better overview of the differences between the two spectra, let us look at the plots of residuals, defined as: $\frac{h_i^2 - h_i^1}{h_i^1} * 100$, where h_i^1 represents the number of events in the *i-th* bin of the *reference* distribution (arbitrarily, distribution for both ϕ and θ was chosen to be the one with $K = -0.65$), and h_i^2 is the number of events in the *compared* distribution.

```
▸ [4103, 12735, 22107, 32306, 44187, 57466, 73305, 90564, 110229, 132131, 156022, 180835, 2
```



Disclaimer: for the moment we are dealing with unnormalized distributions (though the numbers are similar so it is not a big change after normalization)

Looking at the two figures for the residuals, we can notice the following. First, for both distributions ϕ and θ the biggest difference is at low angles (thougg, we must take into account that the detector performance at such angles is lowered compared to higher angles). Second, while the difference for the θ distribution can be of the order of $\sim 75\%$, the change is nowhere near as drastic for the ϕ distribution. This provides a challenge, as θ is not experimentally measurable so we must deal with the smaller difference of the two.

From now on, we will only deal with ϕ distribution

Methodology for statistical analysis

The goal is to quantify the difference between the two angular distributions. We wish to provide the following answers:

1. ```
   How many events (how large a statistics) must be measured to be able to disti
   nguish two spectral shapes within given ``n_{\sigma}`` confidence?
   ```
2. ```
   Is it feasible to obtain such statistics within the 5 year data-taking period
   of SuperNEMO? If not, what would have to be the parameters?
   ```
3. ```
   What is the sensitivity of SuperNEMO toward the given distribution?
   ```

To fulfill the first goal, we have devised a methodology with the aim to find ROI, where the difference is statistically most significant.

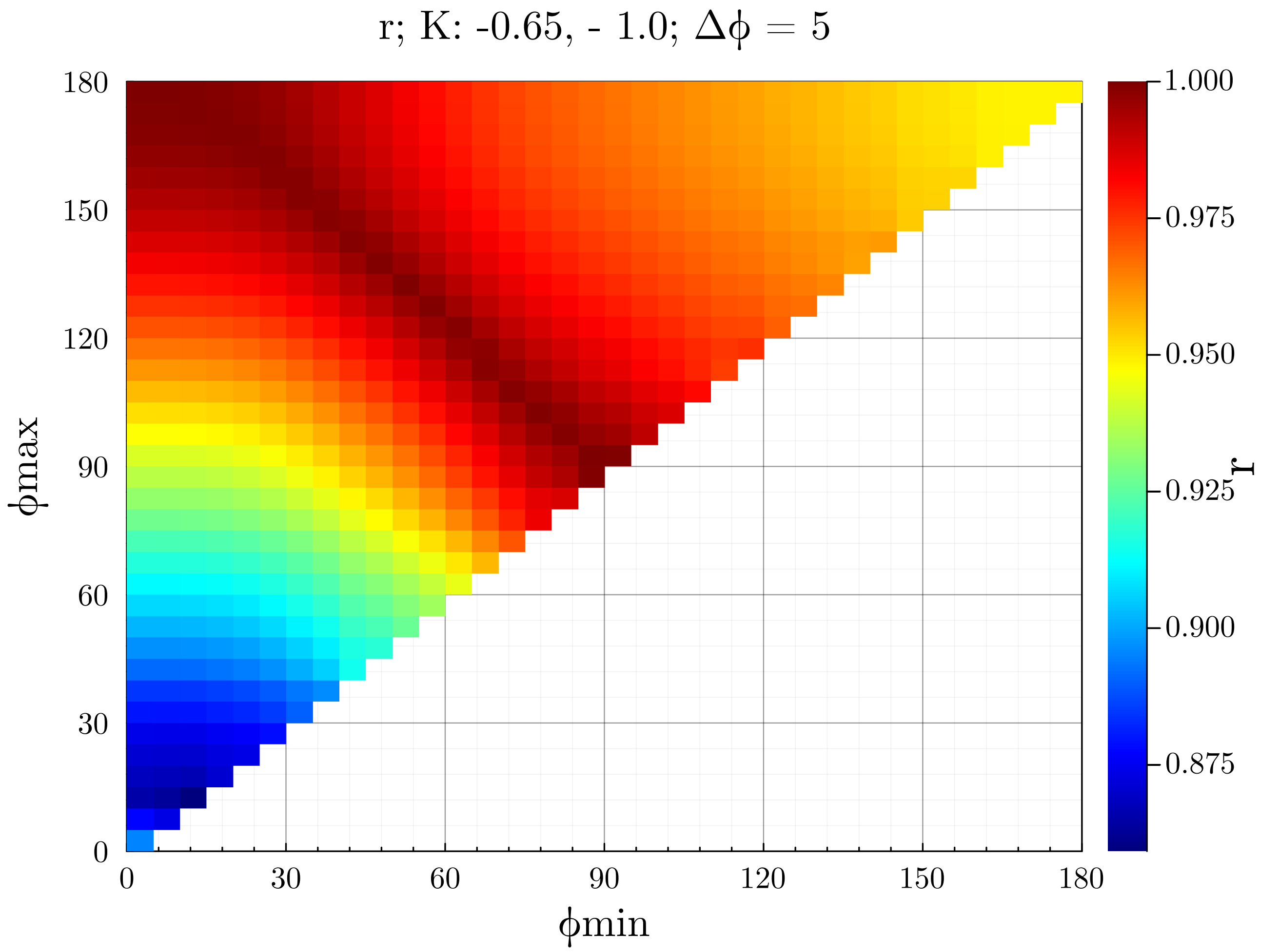
The main idea arises from comparing the difference between the two distributions in terms of bin heights ($h_i^j; k \in (1,2)$). First, we introduce the ratio $r_i \equiv \frac{h_i^{j,k}}{h_i^{k,j}}$, where whether i is in the numerator or j is determined dependent on whether $j < k$ or $j > k$, respectively. Furthermore, to be more *fair* in comparisons, we exchange $h_i^{j,k}$ with $\varepsilon_i^{j,k} \equiv \frac{h_i^{j,k}}{total\_number\_of\_events}$, ie. the normalized bin height. Thus, the ratio now becomes: $r_i \equiv \frac{\varepsilon_i^{j,k}}{\varepsilon_i^{k,j}}$

To determine ROI where the ratio r_i is most favourable for our purposes, that is the lowest number, we create *maps* of various ranges of ϕ. The maps are created by taking some range $\phi \in (\phi_{min}, \phi_{max})$, and calculating the respective r_i. The results are shown in the figure below. Each square in the figure represents a certain range, to be read out by the upper-left corner of the square. (That is, the range $\phi \in (0, \Delta\phi)$ is represented by the square in the down-left corner, the very first square.)

```
get_needed_statistics (generic function with 2 methods)
```

```
c = :jet
```
```

```
hmRatios =
```



▸ (0.86, (10, 15))

In the figure we can see the calculated r_i for each range of ϕ's. The most ideal value is the lowest, ie. r = 0.86 for $\phi \in (10, 15)$ °.

This however, does not tell the whole story yet. We must consider the uncertainty of the result as well as the uncertainty of the simulation. Furthermore, since the aim was to answer **how many events are required** to distinguish the two spectra, we still have some steps to take.

We will therefore produce a few more figures and calculations.

First, to calculate the number of events required, we begin with the following:

Assume M represents the number of events in the *smaller* bin i and N represents the number of events in the *larger* bin. Then in order to distinguish the two bins from each other at the level of n_σ, their difference must be at least equal (or greater) than the sum of their respective uncertainties.

$$M - N = n_\sigma(\Delta M + \Delta N); \Delta M = \sqrt{M}, \Delta N = \sqrt{N}$$

Dividing the equation by M, substituting $r = \frac{M}{N}$ and rearranging yields:

$\tilde{M}(r) = n_\sigma^2(\frac{r+\sqrt{r}}{1-r})^2$, where $\tilde{M}$ simply represents the minimum number of events in M required to distinguish M from N by n_σ.

Now, using uncertainty propagation, we can find the uncertainty on $\tilde{M}$ as:

$$\Delta\tilde{M} = n_\sigma^2(\frac{r+\sqrt{r}}{1-r})^3 r\sqrt{1/M + 1/N}.$$

The uncertainty in $\tilde{M}$ can be improved by obtaining higher statistics.

However, due to the fact that we do not precisely understand the detector angular correlations, we do not know precisely the analytical value for r either. (If we knew perfectly the correlation $\theta \to \phi$, we could obtain r analytically from the input angular distributions). We must, therefore, take into account the uncertainty on r.

$$\Delta r = r\sqrt{1/M + 1/N};$$

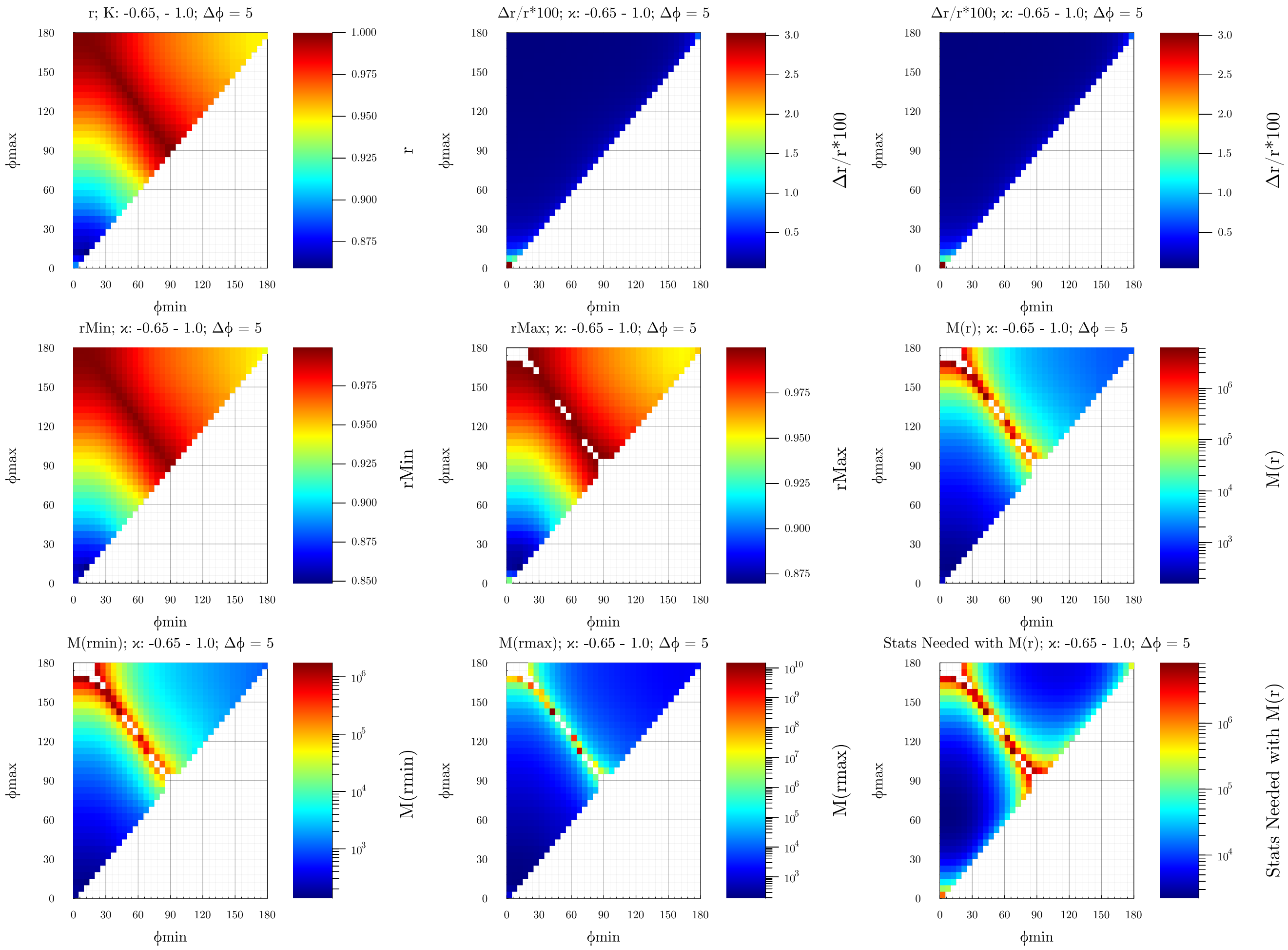
Since the behavior of r is very much non-linear ($\tilde{M}$ rapidly explodes for r close to 1.0) we consider the maximum (and minimum) uncertainties:

$$r_{max} = \frac{M+\Delta M}{N-\Delta N}\ r_{min} = \frac{M-\Delta M}{N-\Delta N}.$$

Here, the worst case scenario (r_{max}) can in some cases exceed 1.0, (ie. $r > 1.0$). Such regions will be removed from the analysis and will show in the maps as empty squares.

To answer the question *how many events SN needs to measure to distinguish* we can convert the $\tilde{M}(r)$ into the number of events needed $Stats$ as follows: $Stats = \tilde{M}(r)/\varepsilon$. That is, we scale the number of needed events in the bin by the proportion of the total events that bin represents. $Stats$ then gives the total statistics needed to obtain desired $\tilde{M}(r)$.

Finally, let us look at all of the mentioned values. In the figure below we show maps for: 1. r, 2.Δr, 3. $\Delta r/r * 100$, 4. r_{min}, 5. r_{max}, 6. $\tilde{M}(r)$, 7. $\tilde{M}(r_{min})$, and 8. $\tilde{M}(r_{max})$, 9. $Stats(r)$.




```
cartesianIdx_to_range (generic function with 2 methods)
```

```
▶ (0, 65)
```

The resultant minimal values are:

- r = 0.86 @ (10, 15)°
- δr = 0.05 @ (0, 180)°
- $M(r)$ = 160.9 @ (10, 15)°
- $M(r_{min})$ = 136.69 @ (10, 15)°
- $M(r_{max})$ = 191.93 @ (10, 15)°
- $Stats(r)$ = 2207.51 @ (0, 65)°

This is as far as we got so far. Conclusions are not yet made. Further analysis will be done later.

Question is what combinatiation is best for $Stats$ and the uncertainties, combined with how well the detector performs for given angular region.