Data Analysis and Interpretation

Report

Assignment Three

Anish Kulkarni, Parth Jatakia, Toshi Parmar, Vedant Basu

Our Webpage

3 November 2016

---

1 Assigned Responsibilities

The following Responsibilities were designated to each team member:

Week	Role Assigned	Member
Week Three	Leader	Parth Jatkia
	Coder	Vedant Basu
	Website Manager	Toshi Parmar
	Report Writer	Anish Kulkarni

Table 1: Week By Week Summary of Assigned Roles

2 Summary of solutions to the assigned problems

1. **Problem 1: Error regions**

   (a) For each temperature T_i assume the measurement to be a random variable that is Gaussian distributed about the mean $\overline{l_i} = mT_i + c$ and standard deviation σ.

   (b) Assuming each measurement to be independent, the log likelihood function is given by:

   $$\log(L(m, c, \sigma)) = \sum_{i=1}^{N} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(l_i - \overline{l_i})^2}{2\sigma^2}\right)\right)$$

   where N is the number of measurements.

   (c) We then minimize the negative of the log likelihood function using the function *minimize* in python, to find the best fit parameters: $\hat{m}, \hat{c}$ and $\hat{\sigma}$

   $$\hat{m} = 22.90 \ mm/Kelvin$$

   $$\hat{c} = 993.51 \ mm$$

   $$\hat{\sigma} = 26.17 \ mm$$

   (d) The equations for the n-σ error region for m and c is given by:

   $$\log(L(m, c, \hat{\sigma})) = \log(L(\hat{m}, \hat{c}, \hat{\sigma})) - n * \frac{2}{2} \qquad 1\text{-}\sigma \ \textit{error region}$$

   Which can be expanded to get : (all summations have index $i = 1$ to N)

   $$m^2 \sum T_i^2 + 2mc \sum T_i + Nc^2 - 2m \sum T_i l_i - 2c \sum l_i + \sum l_i^2 = \sum (l_i - \overline{l_i})^2 + 2n\sigma^2$$

Following is a plot of 1-σ and 2-σ error regions:

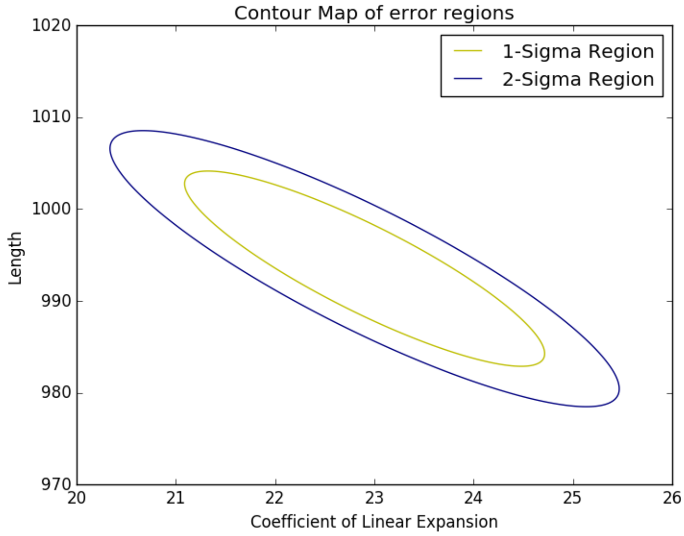

Figure 1: error regions for m and c

(e) Fixing c to it's maximum likelihood estimate,

$$the\ error\ interval\ for\ m = [22.78\ mm/Kelvin, 23.03\ mm/Kelvin]$$

Fixing m to it's maximum likelihood estimate,

$$the\ error\ interval\ for\ c = [992.79\ mm, 994.24\ mm]$$

2. **Problem 2: Discover dark matter**

(a) Histograms of given data and background-only process were plotted:

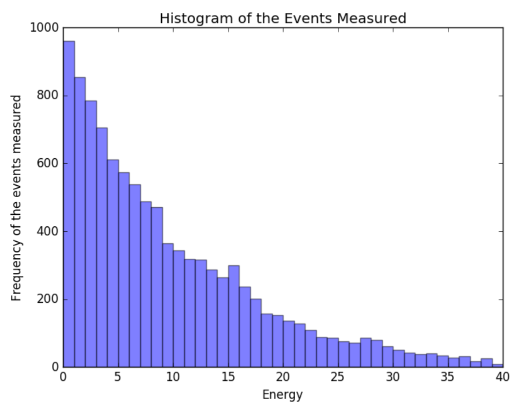

Figure 2: Recoil Energy Data

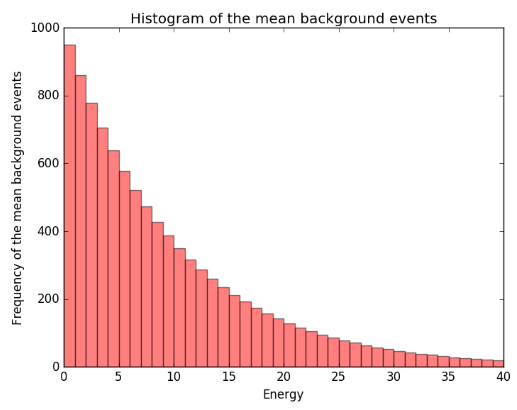

Figure 3: Background

(b) Histogram of the mean number of expected events, considering signal + background was plotted for various cross-section values: For cross-section = 1, 10 and 100 one can

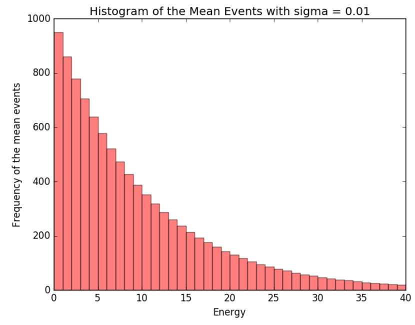

Figure 4: cross-section=0.01

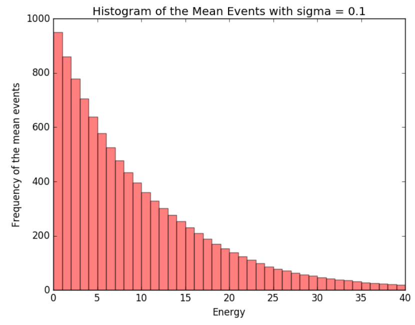

Figure 5: cross-section=0.1

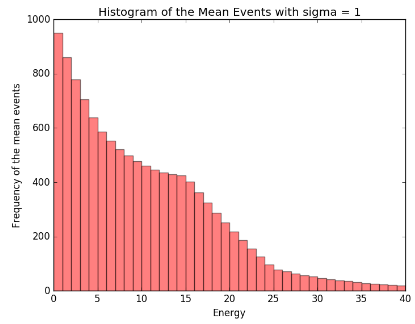

Figure 6: cross-section=1

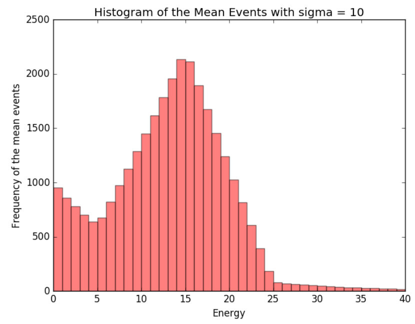

Figure 7: cross-section=10

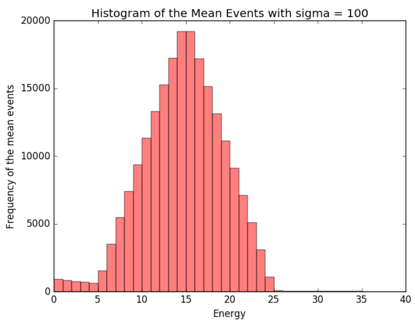

Figure 8: cross-section=100

tell by eye whether or not a dark matter signal is present.

(c) For a large number of scattering events and sufficiently small bin size, probability of measured recoil energy being in a particular bin is small. Hence, we can assume that the number of events in a particular bin is Poisson distributed with mean:

$$\lambda_i = \left(\left(\frac{dN}{dE_R}\right)_{signal} + \left(\frac{dN}{dE_R}\right)_{background}\right) * 1KeV$$

Further, for large number of bins, number of measurements in each bin can be considered to be independent random variables. Hence, the log likelihood function is the sum of individual log likelihood functions for each bin. Let B be the number of bins and let d_i and λ_i be the number of events observed and expected, respectively, in the i^{th} bin. Then,

$$\log(L(\sigma)) = \sum_{i=1}^{B} \log\left(\frac{\lambda_i(\sigma)^{d_i}}{d_i!} e^{-\lambda_i(\sigma)}\right)$$

which on simplification and after ignoring constants gives:

$$\log(L(\sigma)) = \sum_{i=1}^{B} d_i \log(\lambda_i(\sigma)) - \lambda_i(\sigma)$$

Following is a plot of the log likelihood function:

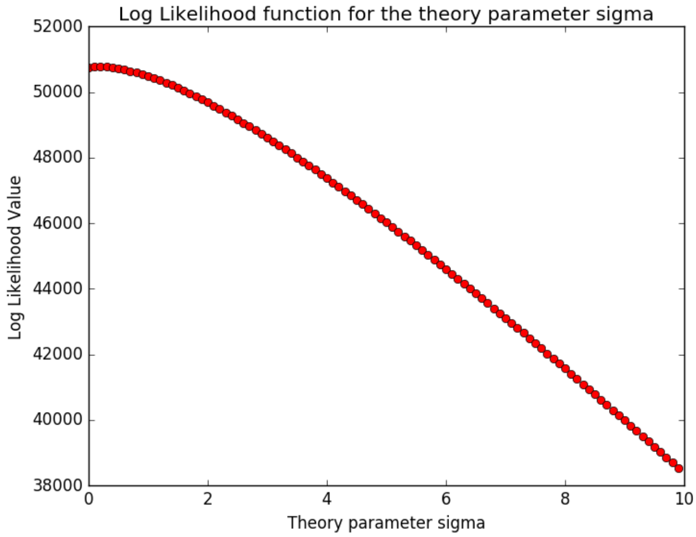

Figure 9: Log likelihood

(d) Maximizing the above calculated log likelihood function gives the maximum likelihood estimate (MLE) for the cross-section:

$$\hat{\sigma} = 0.177 \; fb$$

The 1-σ interval for cross-section is given by

$$\log(L(\sigma)) = \log(L(\hat{\sigma})) - \frac{1}{2}$$

$$1\text{-}\sigma \; \textit{interval} \; = [0.147 \; fb, 0.207 \; fb]$$

(e) Suppose H_0 is the background only hypothesis and H_1 is the hypothesis that dark matter exists. To compare the two hypothesis, we have calculated the log of the likelihood ratio:

$$\log\left(\frac{L(H_0)}{L(H_1)}\right) = \sum_{i=1}^{N} d_i \log\left(\frac{\lambda_i(H_0)}{\lambda_i(H_1)}\right) - (\lambda_i(H_0) - \lambda_i(H_1))$$

$$\log\left(\frac{L(H_0)}{L(H_1)}\right) = -18.98$$

$$\frac{L(H_0)}{L(H_1)} = 5.7 \times 10^{-9}$$

Thus, based on likelihood ratio, we can conclude that the data supports existence of dark matter.