Введение в обработку экспериментальных ускорительных данных (практический курс)

Л.В.Кардапольцев

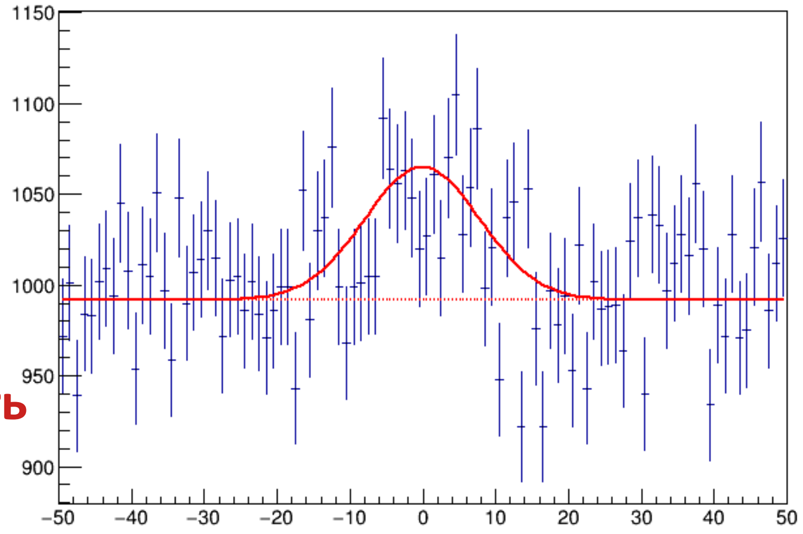
l.kardapoltsev@gmail.com

Пятое занятие

Значимость сигнала.
Оптимизация критериев отбора

Значимость сигнала

- Пусть у нас есть данные, которые мы хотим описать в рамках двух гипотез:
  - H_0: только фон (B)
  - H_1: сигнал плюс фон (S+B)
- Значмость сигнальной гипотезы H_1 это вероятность получить число сигнальный событий >S, предполагая, что в дных есть только фон (H_0)
- Эту вероятность принто выражать в стандартных отклонениях для нормального распределения
- Z функция данных, случайное число
- Ожидаемая значимость это меданное значение Z



$$\Phi = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

$$Z = \Phi^{-1}(1-p)$$

$$Z_{exp} = med(Z)$$

Ожидаемая значимость для расп. Пуассона

- В наших условия отбора прошло n событий, при этом ожидается s событий сигнала и b событий фона
- При этом предположим, что s << b и b > 10, так что распределение для n хорошо описывается распределением Гаусса
- Величину $s/\sqrt{b}$ называют «figure of merit» и используют для оптимизации условий отбора

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

Вероятность обнаружить > n событий нулевой гипотезы

$$p_0 = 1 - \Phi\left(\frac{n-b}{\sqrt{b}}\right)$$

$$Z = \Phi^{-1}(1-p) = \frac{n-b}{\sqrt{b}}$$

$$Z_{exp} = med(Z) = \frac{s}{\sqrt{b}}$$

Ожидаемая значимость

- Подгонка методом максимального правдоподобия в гипотезах b и s+b
- Тогда определим

$$q_0 = 2\left[\ln L_b - \ln L_{s+b}\right] \quad \text{для } s>0$$
$$q_0 = 0 \quad \text{для } s<0$$

- Если гипотезы b и s+b отличаются на один свободный параметр, тогда q_0 распределено согласно

$$q_0 \sim \chi^2_1 \quad \text{или} \quad \sqrt{q_0} \sim N_{0,1}$$

- Тогда вероятность получить число сигнальных событий > s

$$p_0 = P\left(|N_{0,1}| > \sqrt{q_0}\right) = 2\left(1-\Phi\left(\sqrt{q_0}\right)\right)/2$$

$$Z = \Phi^{-1}(1-p) = \sqrt{q_0} \qquad Z_{exp} = med\left(\sqrt{q_0}\right)$$

- Если разница в числе параметров между b и s+b больше 1, то нужно использовать квантиль χ^2_n

Ожидаемая значимость для расп. Пуассона

- Выведем **значимость** сигнала для **распределения Пуассона точно**
- В наших условия отбора прошло **n событий**, при этом ожидается **s событий сигнала и b событий фона**
- В случае если стоит задача не поиска нового сигнала, а **оптимизации ошибки измерения в присутствии фона**, то нужно искать **максимум параметра**

$$\frac{s}{\sqrt{s+b}}$$

$$L_{s+b}=\frac{(s+b)^n}{n!}e^{-(s+b)}$$

$$L_b=\frac{b^n}{n!}e^{-b}$$

тогда

$$q_0=2[\ln L_b-\ln L_{s+b}]=2[n\ln(1-s/b)-s]$$

$$Z_{exp}=\sqrt{2[(s+b)\ln(1+s/b)-s]}$$

$$Z_{exp}\approx s/\sqrt{b}\qquad \text{при } s\ll b$$