Esercizi

Corso di metodologie di Analisi Dati

Academic Year 2022/2023

1 Profile likelihood ratio

The likelihood $\mathcal{L}_i$ for the i-th experiment is defined as the one already treated in class in this notebook

$$\mathcal{L}_i(s, b|n_i, \bar{b}_i) = \mathcal{P}(n_i|s+b)\mathcal{G}(\bar{b}_i|b, \sigma_b) \tag{1}$$

where $\mathcal{P}$ ($\mathcal{G}$) indicates the Poissonian (Gaussian) distribution, s (b) is the true value of the signal (background) yield, n_i is the observed total yields, $\bar{b}_i$ is the given expected background yields from an external measurement with σ_b its uncertainty.

1. for each observation n_i and $\bar{b}_i$, the maximum-likelihood estimators $\hat{s}$ and $\hat{b}$ for s and b, respectively, are defined as

$$(\hat{s}, \hat{b}) = \begin{cases} (n_i, 0) & \text{for } \bar{b}_i <= 0 \\ (n_i - \bar{b}_i, \bar{b}_i) & \text{for } 0 < \bar{b}_i <= n_i \\ (0, \hat{\hat{b}}) & \text{for } \bar{b}_i > n_i \end{cases} \tag{2}$$

   where $\hat{\hat{b}} = \frac{\bar{b}-\sigma_b^2}{2} + \sqrt{\left(\frac{\bar{b}-\sigma_b^2}{2}\right)^2 + n\sigma_b^2}$ is the estimator of b given $s = 0$.

   i) demonstrate that Eq. (2) describes the maximum-likelihood estimators for the three different cases considered.

2. Using 10000 pseudo experiments, estimate the coverage of the upper limit at 95% CL calculated with the profile likelihood method assuming asymptotic limit for the test statistic $t = -2\log\frac{\mathcal{L}\left(s=s', \hat{\hat{b}}'\right)}{\mathcal{L}\left(\hat{s}, \hat{b}\right)}$ where $\hat{\hat{b}}'$ is the fit value for b when $s = s'$. Verify the coverage in the following two conditions

   i) $s = 50$; $b = 150$; $\sigma_b = 30$

   ii) $s = 5$; $b = 15$; $\sigma_b = 9.5$

   and comment the results.

2 Look-elsewhere effect

Invariant mass measurements of a set of particles are reported in the linked dataset, which is the same dataset as the one analysed in class in this notebook. The theoretical model is described by the distribution

$$f(m; a, m_0) = \left|1 + \frac{ae^{i\theta}}{(m^2 - m_0^2) - im_0\Gamma}\right|^2 \tag{3}$$

where a (θ) is the relative magnitude (phase) of the resonant amplitude with respect to the non-resonant, m_0 (Γ) is the mass (width) of the resonance. The unknown parameters of the decay model are a and m_0, while $\theta = \pi/2 - 0.3$ is known.

1. Assuming the asymptotic limit of the test statistic $t = -2\log\frac{\mathcal{L}(a=0)}{\mathcal{L}(\hat{\hat{a}},\overline{m}_0)}$, where $\overline{m}_0$ is the assigned value to m_0 and $\hat{\hat{a}}$ is the fit value of a for $m_0 = \overline{m}_0$:

   i) determine the p-value for the null hypothesis ($a = 0$) for $m_0 = \overline{m}_0$, where $\overline{m}_0$ varies in the interval [5-18] [a.u.] in steps of 0.1. Produce a plot of p-value vs $\overline{m}_0$.

   ii) Convert the observed p-value in standard deviations. Plot the observed standard deviation vs $\overline{m}_0$.

2. Using the test statistic $t' = -2\log\frac{\mathcal{L}(a=0)}{\mathcal{L}(\hat{m},\hat{a})}$, where $\hat{m}_0$ and $\hat{a}$ represents the estimated values for m_0 and a parameters from the fit to pseudoexperiments:

   i) estimate the p-value of the null hypothesis ($a = 0$) by means of 10000 pseudo experiments. Commment on the results.

   ii) plot the distribution of t' obtained from pseudoexperiments and fit the mean using a χ^2 function. Comment on the results.