Statistical Data Analysis
Problem sheet 6
Due Monday 18 November, 2024

Exercise 1: A number n of observed events is modeled as a Poisson distributed variable with mean $s + b$, where s and b are the expected numbers of events from the signal and background processes, respectively. Suppose $b = 3.7$ and $n_{\rm obs} = 15$ are observed.

1(a) [4 marks] Compute the p-value of the hypothesis that $s = 0$. To sum Poisson probabilities, you can use the relation

$$\sum_{n=0}^{m} P(n;\nu) \;=\; 1 - F_{\chi^2}(2\nu; n_{\rm dof}) \,,$$

where $P(n;\nu)$ is the Poisson probability for n given a mean value ν, and F_{χ^2} is the cumulative χ^2 distribution for $n_{\rm dof} = 2(m+1)$ degrees of freedom. This can be computed in python using `scipy.stats.chi2.cdf` or with the ROOT routine `TMath::Prob`.

1(b) [1 mark] Find the corresponding equivalent Gaussian significance Z and evaluate numerically (see lecture notes).

Exercise 2: Suppose x follows a Gaussian distribution with mean μ and variance σ^2 and consider an independently and identically distributed sample of N values $x_1, \ldots, x_N$.

2(a) [1 mark] Write down the log-likelihood function for the mean μ and variance σ^2.

2(b) [4 marks] Show that the maximum-likelihood estimators for μ and σ^2 are

$$\hat{\mu} \;=\; \frac{1}{N}\sum_{i=1}^{N} x_i \,,$$

$$\widehat{\sigma^2} \;=\; \frac{1}{N}\sum (x_i - \hat{\mu})^2 \;.$$

2(c) [5 marks] Find the Fisher information matrix for μ and σ^2 (use $\theta_1 = \mu$, $\theta_2 = \sigma^2$).

2(d) [3 marks] By using this matrix derive the following expressions for the variances of the estimators and their covariance:

$$V[\hat{\mu}] \;=\; \frac{\sigma^2}{N} \,,$$

$$V[\widehat{\sigma^2}] \;=\; \frac{2\sigma^4}{N} \,,$$

$$\mathrm{cov}[\hat{\mu}, \widehat{\sigma^2}] \;=\; 0 \,.$$

State any assumptions needed to arrive at your answer.

2(e) [2 marks] Show with a sketch how one can estimate the standard deviations of $\hat{\mu}$ and $\widehat{\sigma^2}$ using the log-likelihood function. State how the sketch reflects in this case the appropriate value of the covariance of the two estimators.