Teilchenphysik II (Higgs-Physik) (SS 2016)

Institut für Experimentelle Teilchenphysik

Dr. R. Wolf, Dr. S. Wayand

Exercises Sheet 11

http://www-ekp.physik.uni-karlsruhe.de/~quast/ Release: Thu, 16.06.2016

Exercise 31: Lognormal Probability Density Function (presence)

The probability of a random variable x to lie in the interval of $[x, x+\mathrm{d}x]$ if the variable is normal distributed can be obtained from

$$P(x \in [x, x+\mathrm{d}x]) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}} \mathrm{d}x \tag{1}$$

where μ is the expectation value and σ the standard deviation of the *Gaussian* probability density function. You can obtain the *Lognormal* probability function by the varibale tansformation

$$x \to x'(x) = \ln(x)\,.$$

Apply the variable transformation and determine the form of the *Lognormal* distribution expressed by x, from the probability element $P(x \in [x, x+\mathrm{d}x])$ as given in Equation (1). What is the meaning of the *Lognormal* distribution compared to the *Gaussian* distribution?

Exercise 32: Confidence Intervals (presence)

In the lecture you have seen the LHC test statistic for the test of two simple hypotheses H_0 and H_1. In our case H_0 is the "background-only" hypothesis and H_1 is the "signal plus background" hypothesis. The actually observed value of the test statistic, q_{obs}, is indicated by the black line in Figure 1. It corresponds to the outcome of the experiment as a whole (e.g. of several hundred thousand events, that enter the statistical inference in the case of the search for the SM Higgs boson). To judge, which hypothesis is more compatible with the observed result q_{obs} is compared to the expected outcome of the experiment under the assumption that H_0 (H_1) were true. This comparison is usually based on a long series of simulated outcomes of the whole experiment, for which the expected outcome of each individual measurement that enters the statistical inference is randomly varied within the uncertainty of the statistics model and the nuisance parameters of the physics model are varied within their estimated systematic uncertainties. [1]

Qualitatively q_{obs} in Figure 1 is more compatible with H_0, since it is located closer to the expectation for this hypothesis than to the expectation for H_1. The confidence intervals relevant to quantify this situation in particle physics are:

---

[1] For Figure 1 $\mathcal{O}(50'000)$ simulated outcomes of the experiment have been created for each hypothesis. This has only been done for illustrative purposes. For quantitative statements many more of these simulated outcomes are needed, e.g. for the LEP results for each mass hypothesis that has been tested at order of hundred million of such outcomes of the experiment had been simulated.

$$
\begin{aligned}
(1-CL_b) &= \int_{-\infty}^{q_{\rm obs}} \mathcal{P}_b \qquad (p\text{-value}) \\
CL_{s+b} &= \int_{q_{\rm obs}}^{+\infty} \mathcal{P}_{s+b} \qquad (CL_{s+b}\text{ confidence}) \\
CL_b &= \int_{q_{\rm obs}}^{+\infty} \mathcal{P}_b \qquad (CL_b \quad \text{confidence}) \\
CL_s &= \frac{CL_{s+b}}{CL_b} \qquad (CL_s \quad \text{confidence})
\end{aligned}
$$

a) Visualize the integration paths and integral values of these confidence intervals in Figure 1. Note that in the figure not q, but $-q$ has been plotted.

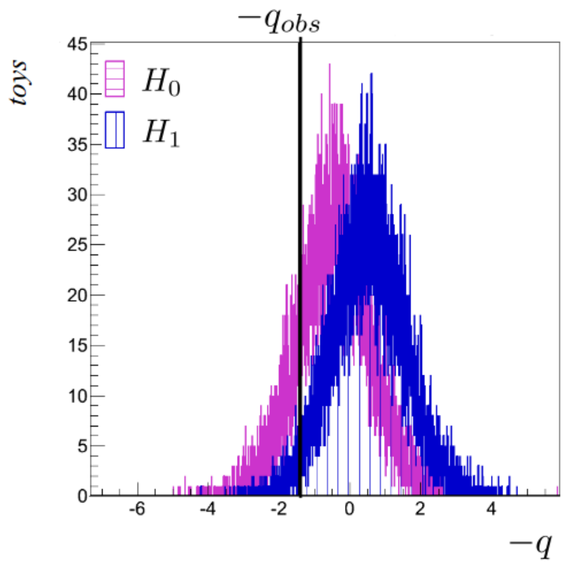

Abbildung 1: The LHC test statistic for a typical test of a “signal plus background” hypothesis (H_1, shown as a blue histogram) versus a “background-only” hypothesis (H_0, shown as a magenta histogram). The observed value of the test statistic (q_{obs}) is indicated by the black line.

b) Assume the parameter of interest to be the signal strength μ of a new signal. Does the 90% CL upper limit on μ correspond to a higher value or to a lower value of μ than the 95% CL upper limit?

Exercise 33: Fitting Exercise (computing exercise)

For this exercise we have prepared two toy datasets with $20'000$ events each in form of a binned histogram with 40 bins (of 5 GeV width) between fictive invariant mass values of 100 GeV and 300 GeV. The datasets are stored in two different root input files:

- data-scenario-A.root
- data-scenario-B.root

You can find these two input files and two further files `fit_data.C` and `FitFunctions.h` in the tarball stored under the web link:

`http://www-ekp.physik.uni-karlsruhe.de/~quast/vorlesung/TP2HiggsSS16/Uebungen/fitting-exercise.tar.gz`.

We assume that you use the same setup as for Exercise 29. You can download and extract the linked file to some local space in your virtual machine.

a) First investigate the content of the directory folder with name `fitting-exercise` that you have extracted. Open the root files that you find in there and inspect them in the root Browser:

```
> cd ./fitting-exercise/
> root -l data-scenario-A.root data-scenario-B.root
> TBrowser t
```

b) Next make yourself familiar with the example macro `fit_data.C` that we provide. For this purpose open it with your favorite editor and go through the code and the comments that you find there. Finally you can execute the macro running the following commands:

```
> root -l
> .x fit_data.C++(120)
```

The argument "120" corresponds to the mass point that you want to test. You can find an explanation of the syntax used for the last command in the head of the file.
The macro loads one of the two root files (to be defined in `line 81` of the macro), reads the data histogram from it and applies a maximum likelihood fit of a simple physics model to it. We have defined this physics model in the header file `FitFunctions.h`. It is made up from a falling exponential function with a potential signal in form of a *Gaussian* distribution on top. For simplicity reasons the width of the signal is restricted to be 20% of the tested mass value. The macro will prompt a plot of the data histogram and the fitted physics model and a significance value that has been estimated exactly as discussed in the lecture. In one file we have hidden a signal. Where is it and what significance does it have? The one who gets closest to the true answers wins a bottle of wine (or an equivalent alternative) that will be handed over during our last nominal lecture. For the first price we want three correct statements: (i) the file; (ii) the mass and; (iii) the significance of the signal.

Note 1:
The signal that you fit has a width. It does not make too much sense to choose a mass value at the very borders of the histogram for your fit. Make up your mind and decide from where on fitting makes sense.

Note 2:
We do not care how you obtain the solution. You can modify the example macro how ever you like or use it just as is. The latter requires a bit more hand work that the computer can do for you in a reproducible way if you modify the macro correspondingly.

Note 3:
If you want to learn more about the root classes that have been used throughout the example the easiest way to obtain this information is just to use your favourite search engine. Practical

key words are the class name (e.g. `TF1`) plus root.