Frequentist statistics in a nutshell

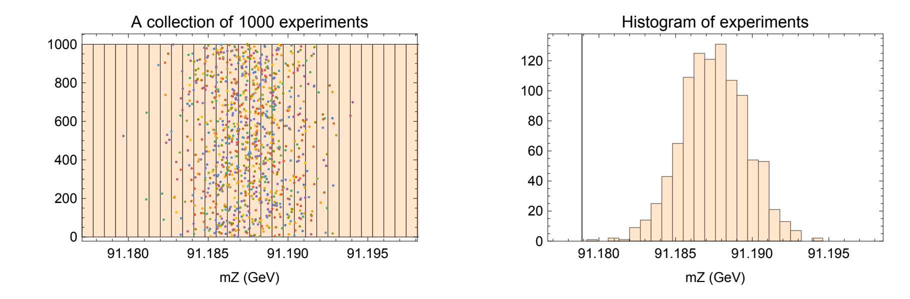
From measured (random) data, frequentist statistics answers the following question:

assuming some hypothesis $\mathcal{H}$ is true (the null hypothesis), are the observed data likely ?

Example: assuming the Standard Model is true, is my best fit value for m_{Z} likely ?

m_Z can be measured in e^+e^- collisions in the relevant invariant mass window. One can use the best fit value $\widehat{m}_Z$ of the resonance peak location as an *estimator* of the true value of m_Z. Estimators are functions of the data and thus are random variables. The estimator is said to be *consistent* if it converges to the true value when data statistics tends to infinity (*e.g.* maximum likelihood estimators are consistent). Another useful concept is the *bias*, which is defined as the difference between the average of the estimator among a large number of finite statistics experiments with the true value. Consistency implies that the bias vanishes asymptotically.

Assuming one can repeat many times the same experiment, one gets a collection of $\hat{m}_Z$ values. The histogram of this random sample brings information on the most likely value of m_Z and the average accuracy of the experiments.

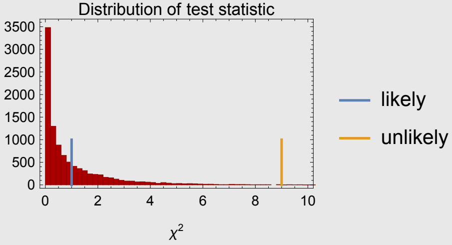


However in practice one only performs one (or a few) experiment(s). Thus one has to find a way to conclude whether the observation is likely from the information of a single experiment.

Repeated experiments and p-value

Whether given data are likely or not is usually quantified using a *test statistics* t, which is a function of data X such that *e.g.* low values supports the null hypothesis $\mathcal{H}$ whereas large values go against it.

Then from the distribution of X one may compute the distribution of $t(X)$, as well as the probability $p(X_0)$ that the value $t(X)$ of a (often fictitious) repeated experiment is larger than the observed value $t(X_0)$: if $p(X_0)$ is large (small) it means that $t(X_0)$ is small (large) with respect to 'typical' values of $t(X)$, and thus that the observed data are in good (bad) agreement with the null hypothesis.



Confidence intervals and coverage

The hypothesis $\mathcal{H}$ is said to be *simple* if it completely specifies the distribution of the data X. In this case the p-value constructed from $t(X)$ is nothing else than the CDF of t, and thus the p-value is uniformly distributed with the observed value X_0.

In case of a numeric hypothesis $\mathcal{H} : X_{\rm true} = \mu$, the p-value curve allows the construction of *confidence intervals*: the interval of μ defined by $p \geq 1 - \mathrm{CL}$ contains $X_{\rm true}$ at the frequency CL, as follows from the uniformity of p.