Statistical tests

Alexander Khanov

PHYS6260: Experimental Methods is HEP
Oklahoma State University

October 25, 2017
Statistics

- **Statistic (singular) $S$:** any function defined on a set of data $x = x_1, \ldots, x_N$. Examples:
  - sample mean $\langle x \rangle = \frac{1}{N} \sum_{i=1}^{N} x_i$
  - sample variance $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$
  - likelihood $L = f(x, p)$
    - reminder: in the likelihood definition, we assume that variables $x$ are described by a joint p.d.f. $f(x, p)$ which depends on $n$ parameters $p = p_1, \ldots, p_n$
    - likelihood ratio $\lambda = \frac{f(x, p)}{f(x, P)}$
      - $P$ are values of $p$ which maximize $L(p)$
  - Since measurements $x$ are random variables, $S$ is also a random variable with its own p.d.f. $f(S)$
P-value

- Let $S$ be constructed in a way that the larger $S$, the worse the data-to-model agreement.
- For a given set of data we have some observed value of the statistic $S_{obs}$.
- **p-value $\mathcal{P}$**: probability that $S > S_{obs}$ (i.e. $S$ is found in a region where the data-to-model agreement is worse than observed).

\[
\mathcal{P} = \int_{S=S_{obs}}^{\infty} f(S) \, dS
\]

- The less the p-value, the smaller the probability that the hypothesis we are testing is valid.
- What we are usually testing is a null hypothesis (the assumption that there is no signal), with corresponding p-value $\mathcal{P}_0$.
  - if $\mathcal{P}_0$ is very small then most probably the signal is there.
Higgs p-value
Pearson’s $\chi^2$ statistic

- Which statistic to use to calculate p-value?
- Suppose we have $N$ measurements $n = n_1, \ldots, n_N$ which are Poisson distributed random variables, and $N$ predicted values $\nu = \nu_1, \ldots, \nu_N$ which depend on $n$ parameters $p = p_1, \ldots, p_n$
- Pearson’s $\chi^2$ statistic is defined as

$$\chi^2 = \sum_{i=1}^{N} \frac{(n_i - \nu_i)^2}{\nu_i}$$

- If $n_i$ are large, Pearson’s $\chi^2$ statistic follows $\chi^2$ p.d.f. $f(\chi^2, NDF)$, where $NDF = N - n$
- p-value for Pearson’s $\chi^2$ statistic:

$$P = \int_{\chi^2_{obs}}^{\infty} f(\chi^2, NDF) \, d\chi^2$$
Reminder: $\chi^2$ distribution

- This is the distribution of a sum of the squares of $k$ independent standard normal random variables ($x_0 = 0, \sigma = 1$)

$$f(x, k) = \frac{1}{2^\frac{k}{2} \Gamma \left( \frac{k}{2} \right)} x^{k-1} e^{-\frac{x}{2}}$$

- If $k$ variables $x_i$ are distributed normally then $\sum_i \frac{(x_i - x_{0i})^2}{\sigma_i^2}$ is distributed as $\chi^2$

- $\chi^2$ distribution has mean value $k$ and variance $2k$
Confidence intervals

- If the parameter estimator $P$ is believed to be Gaussian distributed, one can just quote its mean value $\langle P \rangle$ and the standard deviation $\Delta P = \sqrt{\sigma_P^2}$ as the result of the measurement.
- **Confidence interval**: range of parameter values $p_{\text{min}} < p < p_{\text{max}}$ such that the probability that the true value of the parameter $p_{\text{true}}$ lies within the range is a predefined number $\alpha$ called confidence level.
Confidence intervals (2)

- What to do in the case when the parameter estimator distribution is not Gaussian, or there are physical boundaries on possible values of $p$?

- The procedure:
  - consider a test of the hypothesis that the parameter’s true value is $p$
  - exclude all values of $p$ where the hypothesis would be rejected at a significance level $\alpha$ (in other words, where the p-value is less than $\alpha$)
  - the remaining values of $p$ constitute the confidence interval at confidence level $\alpha$

- Which test to use for this procedure?
  - a popular choice is the likelihood ratio
  - the confidence intervals obtained in this way are known as Feldman and Cousins