Université de Genève — Autumn term 2022

# Statistical Methods in Physics (14P058)

Prof. Federico Sánchez (federico.sancheznieto@unige.ch)
Dr. Hepeng Yao (hepeng.yao@unige.ch)
Dr. Knut Zoch (knut.zoch@unige.ch)

# Exercise V – Basics of parameter estimation

24 November & 1 December 2022, 9:15, room: SCI–202

In this exercise, we are going to look at some basics of parameter estimation. Let's first set the foundation and implement random sampling from a Gaussian distribution, before looking at the $\chi^2$ distribution and the $t$-distribution in more detail.

**Task 1:** *Gaussian sampling & calculation of the $\chi^2$ value*

a) Implement a Gaussian distribution as you did for Exercise sheet 3. Plot a Gaussian distribution for $\mu = 2$, $\sigma = 0.5$ to test your implementation.

   *Hint:* choose the signature of the function to have the following form:

```
def Gauss(x, mu=0, sigma=1):
```

b) Implement a function that samples $x, y$ pairs from a Gaussian distribution. Ideally your function should have a signature such as this one:

```
def sample_from_Gaussian(n_samples, mu=0, sigma=1):
```

   *Hint:* initialise a random number generator (`np.random.default_rng()`), which you can use to draw random numbers from a uniform distribution, $r \in [0, 1)$. Based on the inversion trick, you can then use the percent point function (i.e. the inverse of the c.d.f.!) of the Gaussian distribution as implemented in `scipy.stats.norm.ppf(r, loc=0, scale=1)` to project $r$ on the Gaussian envelope. The location parameter corresponds to the population mean $\mu$, the scale to the standard deviation $\sigma$ of the population. For sampling in $y$, you can simply use your previously defined `Gauss(x, mu, sigma)` function.

c) Now sample a low number of $x, y$ pairs, e.g. $n = 6$, from the Gaussian distribution. Plot them together with the Gaussian envelope as a scatter plot. You can play around with the sample size to get a better idea if your random sampling is working.

d) Implement a function to calculate the $\chi^2$ value of the sample according to:

$$\chi^2 = \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^2} \tag{1}$$

**Task 2:** *Comparison with the $\chi^2$ distribution*

Now we fix $n = 6$, but perform the same experiment a total of $n_{\text{exp}} = 5000$ times.

a) Calculate the $\chi^2$ values of each of the experiments and fill them into a histogram. Compare the histogram with the p.d.f. of the corresponding $\chi^2$ distribution.

b) What does it mean if a variable, such as the one defined in eq. (1), is distributed according to a $\chi^2$ distribution?

*Hint:* the p.d.f. of the $\chi^2$ distribution is accessible via `stats.chi2.pdf(x, df)`.

**Task 3:** *Student's t-distribution*

Last but not least we want to calculate the $t$-test variable according to

$$t = \sqrt{n}\,\frac{\bar{x} - \mu}{\hat{\sigma}} \tag{2}$$

for a set of $n_{\text{exp}} = 5000$ experiments. Here, $\hat{\sigma} = s$ is the (Bessel-corrected) estimate of the standard deviation of the population, i.e., the sample standard deviation.

a) Implement a function to calculate the $t$-test variable according to eq. (2) for a list of sampled values. The function could for example have the following signature:

```
def calculate_t(x, mu):
```

b) Again, use $n = 6$ and $n_{\text{exp}} = 5000$ and calculate the $t$-test variable for each of the experiments. Fill them into a histogram. Compare the histogram with the p.d.f. of the corresponding Student's $t$-distribution. What is the number of degrees of freedom to choose here? Why?

c) What does it mean if a variable, such as the one defined in eq. (2), is distributed according to a $t$-distribution?

*Hint:* the p.d.f. of the $t$-distribution is accessible via `stats.t.pdf(x, df)`.