I. PRINCIPLES OF THE CHI-SQUARED TEST

A. Poisson distribution

In addition to the normal (or Gauss) distribution discussed earlier, the *Poisson distribution* is another important limiting distribution often encountered in nature. Usually we deal with the Poisson distribution when we count events that occur randomly but at a well defined average rate. For example, consider the number of car accidents happening on a busy road in a two-month period. Obviously, these are fairly random events. Therefore, the number of accidents in two consecutive two-month periods might not be the same. However, if such factors as the road condition, the busyness of traffic, the weather etc. are fairly constant in time, we can expect that the *average* number of accidents in the two-month period will be also constant. In other words, the *average rate* of accidents will be well defined. In this case, we expect that the probability to have n accidents in a given two-month period is given by the distribution function

$$P_\nu(n) = \frac{\nu^n e^{-\nu}}{n!}, \tag{1}$$

where ν is a parameter which physical meaning will be clarified in a minute. This is called the Poisson distribution. Note that, unlike the Gauss distribution, it is completely defined by only one parameter ν (in case of Gauss distribution we need to know two parameters X and σ). To find its physical meaning, let's find the average value of n. First note that the Poisson distribution is defined for discrete (integer) values, and has to be normalized according to

$$\sum_{n=0}^{\infty} P_\nu(n) = 1. \tag{2}$$

The average value of counts per period is

$$\langle n \rangle = \sum_{n=0}^{\infty} n P_\nu(n) = \sum_{n=1}^{\infty} \frac{\nu^n e^{-\nu}}{(n-1)!} = \nu e^{-\nu} \sum_{n=1}^{\infty} \frac{\nu^{n-1}}{(n-1)!} = \nu. \tag{3}$$

Above, we dropped term for $n = 0$, which is identically zero, and used a well known expansion series $e^x = \sum_{n=0}^{\infty} x^{-n}/n!$. Thus, the parameter ν in the distribution function in Eq. (1) is just the mean value of distribution. Note that this justifies our assumption of the well defined rate of events. Indeed, if we count events during a time period of duration T (e.g. two

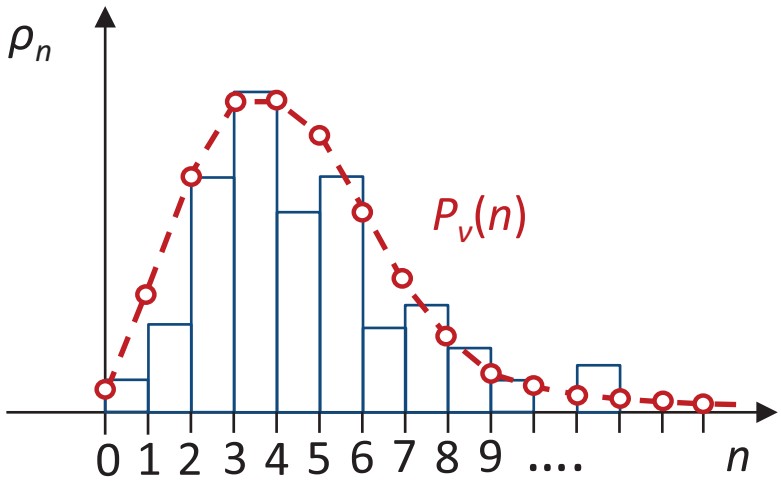

FIG. 1: Histogram of fractional occurrences of events and their continuous limit, the Poisson distribution $P_\nu(n)$ given by Eq. 1.

months in the above example), the rate of events r is obviously given by $r = \nu/T$, which is well defined number.

In reality, we do not know distribution parameter ν in advance, but we can obtain an estimate for ν by counting events in a series of N time intervals of duration T, e.g. $[0, T]$, $[T, 2T]$, ..., $[(N-1)T, NT]$. Suppose, we obtained a series of N results n_1, n_2, .., n_N. Obviously, the mean value if $\langle n \rangle_N = \sum_{i=1}^{N} n_i$. Again, according to the *principle of maximum likelihood*, this number will be our best estimate of the parameter ν (and, therefore, the rate of the events) providing that the random occurrence of events is governed by the Poisson distribution.

Once again, I'd like to emphasize that distribution $P_\nu(n)$ is the *limiting distribution*. In the example of measurements given above, you can calculate the *fractional occurrences* ρ_n of n events in an interval for every integer number $n = 0$, 1, .. by counting how many times number n occurs in your experimental set n_1, n_2, .., n_N, and divide it by N. Obviously, $\sum_{n=0}^{\infty} \rho_n = 1$. If you plot ρ_n as a function of n in the form of a histogram, you might obtain something like depicted in Fig. 1. If you do counting for a very large number of time intervals, ideally $N \to \infty$, you expect to become closer and closer to distribution $P_\nu(n)$, see dashed segmented line in Fig. 1. Note that this line is a schematic depiction of distribution, which is defined only for discrete (integer) values!

Next, lets find the dispersion σ^2 of the distribution (1). By definition, we have

$$\sigma^2 = \sum_{n=0}^{\infty}(n-\nu)^2 P_\nu(n) = \sum_{n=0}^{\infty} n^2 P_\nu(n) - \nu^2 = \langle n^2\rangle - \nu^2. \tag{4}$$

To find the mean value of n^2, let's instead consider mean value of $n(n-1)$. We have

$$\langle n(n-1)\rangle = \sum_{n=0}^{\infty} n(n-1)\frac{\nu^n e^{-\nu}}{n!} = e^{-\nu}\nu^2 \sum_{n=2}^{\infty}\frac{\nu^{(n-2)}}{(n-2)!} = \nu^2. \tag{5}$$

Thus, we have $\langle n^2\rangle = \nu^2 + \nu$, therefore $\sigma^2 = \nu$. In other words, the Standard Deviation of the Poisson distribution is equal to $\sqrt{\nu}$. This is very famous result known as a "square-root rule", and it will be used in the following discussions. Also, note that the Poisson distribution is *completely* defined by its mean value ν, which also gives the dispersion of the distribution.

Note that although the distribution $P_\nu(n)$ is defined for the integer values n, the mean value is not necessary an integer number. It can as well be much smaller than 1, that is $\langle n\rangle << 1$. Indeed, an event can have such a slow rate that it's rarely observed during let's say a ten-minute interval. Sometimes we'll be lucky to observe one (or may be even two!) event, but most of the time we will not see it if we repeat observations many times. Of cause, the mean value of events per ten-minute interval will be much less than 1.

Finally, I'd like to warn against some common misuses of the above results. The first example is concerned with the estimation of a physical quantity that contains the number of events in its definition. For example, suppose we would like to estimate experimentally the rate of a certain process, for example the emission of α-particles by a radioactive sample. Then, we count number of events due to this process, for example number n of clicks of the Geiger counter brought close to the sample, during a time period T. As a result of this *single* measurement, we can conclude that the average number of events during this time period is n, therefore the average rate is $r = n/T$. Now, how do we estimate the error of our result? We can assume that the events are governed by the Poisson distribution, therefore the Standard Deviation of out result for n is given by $\sqrt{n}$, and from the equation for propagation of errors we conclude that the error of our result for r is $\sqrt{n}/T$. Note that it would be completely wrong to apply the "rule of square root" directly to r, that is say that our error is $\sqrt{r}$. *Remember that the Poisson distribution is defined only for integer numbers, while the physical quantity r can be any number.*

As a second example, consider that we would like to improve the above result and carried out counting of clicks during N consecutive time intervals, each of duration T. Obviously, our best estimate for the radioactive decay rate r becomes $\langle n \rangle / T$. What is the estimated error of our result? From the discussion given in the Handout 1 it is clear that the error in $\langle n \rangle$ is given by the Standard Deviation of the Mean, thus the estimated error of our result is given by SDM/T. Not that it will be completely wrong to estimate our error in $\langle n \rangle$ as $\sqrt{\langle n \rangle}$. Indeed, as mentioned above the mean value of n can as well be less than 1, in which case $\sqrt{\langle n \rangle}$ will be larger than 1. And it would be a nonsense to state your result as $\langle n \rangle / T \pm \sqrt{\langle n \rangle}/T$ because your result for the decay rate becomes negative within the error bars! It is easy to see that the SDM in the above example will be always less than or equal to $\langle n \rangle$.

B. Basic idea of the Chi-Squared test

As has been repeated many times, we would be never able to determine the parameters of the limiting distribution (for example X and σ for the normal distribution, or ν for the Poisson distribution) exactly because that would require to repeat our measurements infinite number of times, $N \to \infty$. Instead, we assume that our measured random quantity obeys a particular distribution, and use the principle of maximum likelihood to find our best estimates (and corresponding errors) for these parameters. However, we might not be sure that our quantity indeed obeys this distribution. In this case, it is desirable to establish some procedure and criteria to judge if the measured quantity indeed obeys this distribution, and this procedure should be based on the available experimental data. This procedure is called the χ^2-test (Chi-Squared).

Suppose we performed N measurements of a physical quantity x (either continuous or discrete) and obtained a series of results x_1, x_2, .., x_N. We suspect that spread of these results is due to some random processes and obey the Gaussian distribution $\rho_G(x)$. We can plot the histogram of fractional occurrence ρ_k, like we discussed in the Handout 1, and visually compare it with $\rho_G(x)$. However, we desire some *quantitative* comparison between the two. Well, let's divide the whole range of all possible x into a number of "bins" such that each bin contains a significant number of measured results x_i. Next, let's calculate the number of measured values x_i that fall into k-th bin and call this number O_k (*observed*

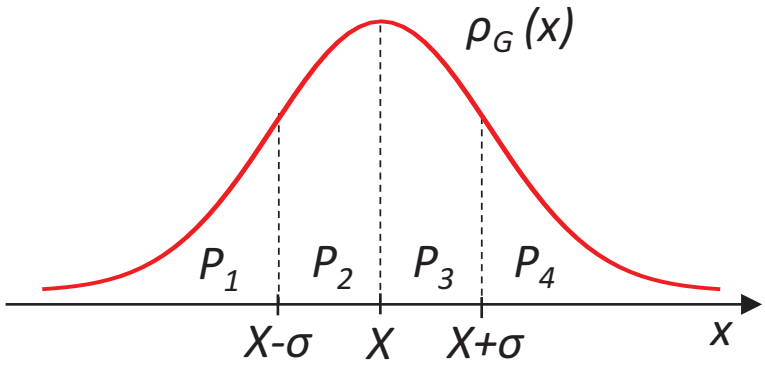

FIG. 2: Example of division into bins for χ^2-test of the Gauss distribution.

number). Typically, for the χ^2-test the number of bins should not be very larger, certainly much less than we needed in Handout 1 to visualize the distribution . For example, in case of the Gauss distribution, it is reasonable to use $K = 4$ bins, which divide the range of all possible x into four intervals $[-\infty, X-\sigma]$, $[X-\sigma, X]$, $[X, X+\sigma]$, and $[X+\sigma, \infty]$, see Fig. 2. For each interval, we can calculate the probability P_k that the result of a single measured will fall into this interval. Thus, if we are dealing with N measurements, we can calculate the *expected* number E_k of measurements that fall into the k-th interval as $N \times P_k$. Obviously, if our quantity indeed obeys the Gauss distribution, the observed numbers O_k should be close to the expected numbers E_k. How close?

To answer this question, let's ask ourselves what do we expect if we repeat the procedure of making N measurements many times, and count O_k for each bin for all these repeated procedures. Naturally, we expect that different 0_k for the same bin will fluctuate around the "true" value E_k with a standard deviation given by the "square-root rule" as $\sqrt{E_k}$. Thus, we expect that the absolute value of the difference $O_k - E_k$ should not be very different from 1, providing that distribution of x is indeed described by $\rho_G(x)$. Thus, we introduce the quantity χ^2 as

$$\chi^2 = \sum_{k=1}^{K} \frac{(O_k - E_k)^2}{E_k}, \tag{6}$$

and state that our assumption that the distribution of x obeys $\rho_G(x)$ is valid if $\chi^2 \approx K$, where K is the number of chosen bins. This statement formulates the χ^2-test.

C. Degrees of freedom, reduced χ^2 and probabilities of χ^2

Now, let us refine the above definition. Note that in order to calculate the probabilities P_k, where $k = 1, 2, .., K$, we need to know the parameters of the distribution, which actually are never known *a priori*. What can we do? Well, again we can find our best estimates of these parameters from our experimental data using the *principle of maximum likelihood*. For example, we can approximate the parameters X and σ for the Gauss distribution by the mean value $\langle x \rangle_N$ and standard deviation SD calculated from our N measurements. Then, we can calculate probabilities P_k and, therefore, χ^2. However, note that by using our measurement results in calculating E_k, we introduce some *constraints* in our test for the "true" distribution. In the above example, we introduce $C = 3$ constrains: the number of measurements N, the estimation of X as the mean value $\langle x \rangle$, and the estimation of σ as the SD. Correspondingly, we define the number of *degrees of freedom* as $D = K - C$, and the reduced χ^2 as

$$\tilde{\chi}^2 = \frac{\chi^2}{D}, \tag{7}$$

and formulate our criteria for validity of the distribution as $\tilde{\chi}^2 \approx 1$.

From the above it is clear that the number of beans has to be at least larger than the number of constrains. For example, in the above example of Fig. 2 we have $D = 4 - 3 = 1$. Suppose now that we calculated the reduced Chi-Squared and obtained $\tilde{\chi}^2 \approx 1.5$. Is it sufficiently close to 1 to consider this validity test as positive? What we can do is to use the Gauss distribution and calculate the probability $P_{\tilde{\chi}^2 \geq 1.5}$ that the value of $\tilde{\chi}^2$ is larger than or equal to 1.5. If this probability is small, lets say below 1%, we can say that based on our data it is unlikely that distribution of our quantity is goverened by the Gauss distribution. If it is let's say about 25%, we can assume that it is quite reasonable that our distribution is indeed the Gauss distribution. Usually, we define a *significance level* (e.g. 5% significance level) and compare the result for probablity of $\tilde{\chi}^2$ with this number. If it is larger than the significance level, we say that we accept the assumption of validity at this significance level. Otherwise, we reject the asumption at this significance level.

The calculated probabilities $P_{\tilde{\chi}^2 \geq A}$ depend on D and are usually stated in tables. For example, a typical table states numbers for $P_{\tilde{\chi}^2 \geq A}$ for $D = 1, 2, ...,$ and some discrete values of A. Usually these tables are calculated for the Gauss distribution, but also are

very reasonably account for other distributions, which can be approximated by the Gauss distribution. For example, for values of $\nu > 8$ the Poisson distribution is very closely described by Gauss distribution with $X = \nu$ and $\sigma = \sqrt{\nu}$, as you can check numerically yourself.