Goodness of Fit Tests

Given measurements, $x_1,\ldots,x_n$, following an unknown distribution $f(x)$ and if $f_0(x)$ is a specified distribution, we may want to test the hypothesis,

$$H_0 : \quad f(x) = f_0(x)$$

As usual we form a test statistic of known distribution and define rejection and acceptance regions with probabilities α and $1-\alpha$, assuming H_0 is true.

Pearson's χ^2 test

- exact for large samples only
- data are binned into N exclusive bins
- the hypothesis under test:

$$H_0 : \quad p_1=p^0_1, \quad p_2=p^0_2, \quad \ldots. \quad p_N=p^0_N$$

where $\sum_{i=1}^{N} p^0_i = 1$

To test whether the observed number of entries in each bin is compatible with the predicted number, form the variable,

$$X^2 = \sum_{i=1}^{N} (n_i - np^0_i)^2/np^0_i = \left(\sum_{i=1}^{N} n_i^2/np^0_i\right) - n$$

when H_0 is true, X^2 approximately follows the χ^2 distribution with N-1 degrees of freedom. (As long as np^0_i is large enough so Poisson is approximately Gaussian.)

- If H_0 is false, then X^2_{obs} will take on larger values, so define the rejection region to be at the largest values of X^2.

- Remarks about the choice of binning in section on Least squares fitting apply here.

- If the data were used to determine L linear parameters of the model, the X^2 would follow χ^2 distribution with N-1-L degrees of freedom (if the determination was done with the same binning and found using LS or ML).

- If unbinned ML used to determine parameters, X^2 no longer strictly χ^2 (N-1-L), but it is bounded by χ^2(N-1) and χ^2(N-1-L). If N>>L, there is little difference.

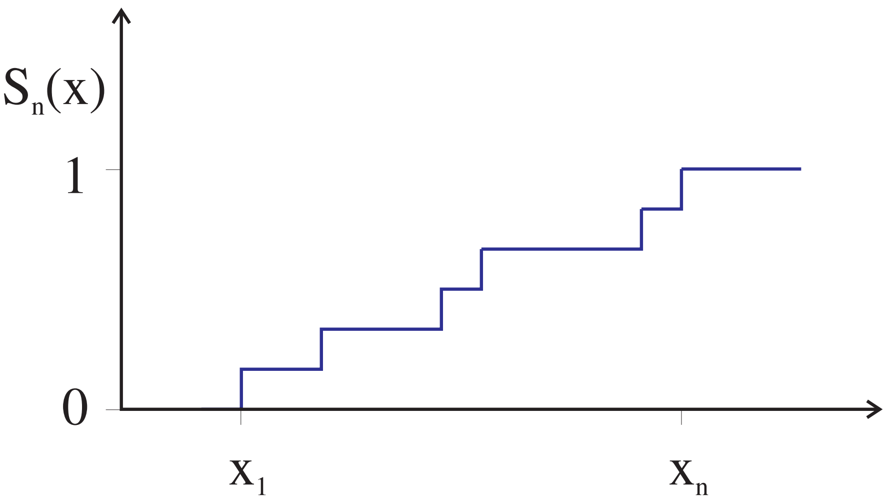
Kolmogorov - Smirnov test

- avoids binning of data
- superior to Pearson's χ^2 for small samples

Given n observations of x, form an ordered sample, ascending in magnitude: $x_1, x_2, \ldots, x_n$

The cumulative distribution is defined by

$$S_n(x) = \begin{cases} 0 & x < x_1 \\ i/n & x_i \leq x < x_i{+}1 \\ 1 & x \geq x_n \end{cases}$$



Compare this to the expected cumulative distribution, $F_0(x)$:

Form the quantity, $D_n = \max |S_n(x) - F_0(x)|$

If $F_0(x)$ is completely specified (ie. no parameters deduced from the data), then D_n is independent of $F_0(x) \Rightarrow D_n$ is distribution free.

Large n limit:

$$P(D_n \leq z/\sqrt{n}) = 1 - 2 \sum_{r=1}^{\infty} (-1)^{r-1} e^{-2r^2z^2}$$

which is valid for $n \geq 80$.

For $n \geq 100$, the following table can be used to define rejection region:

$P(D_n \leq d_\alpha) = 1 - \alpha$:

α	0.20	0.10	0.05	0.01
d_α	$1.07/\sqrt{n}$	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.63/\sqrt{n}$

Exercise 13

Apply (a) Pearson's χ^2 and (b) Kolmogorov - Smirnov tests to the experimental samples produced in exercise 11.

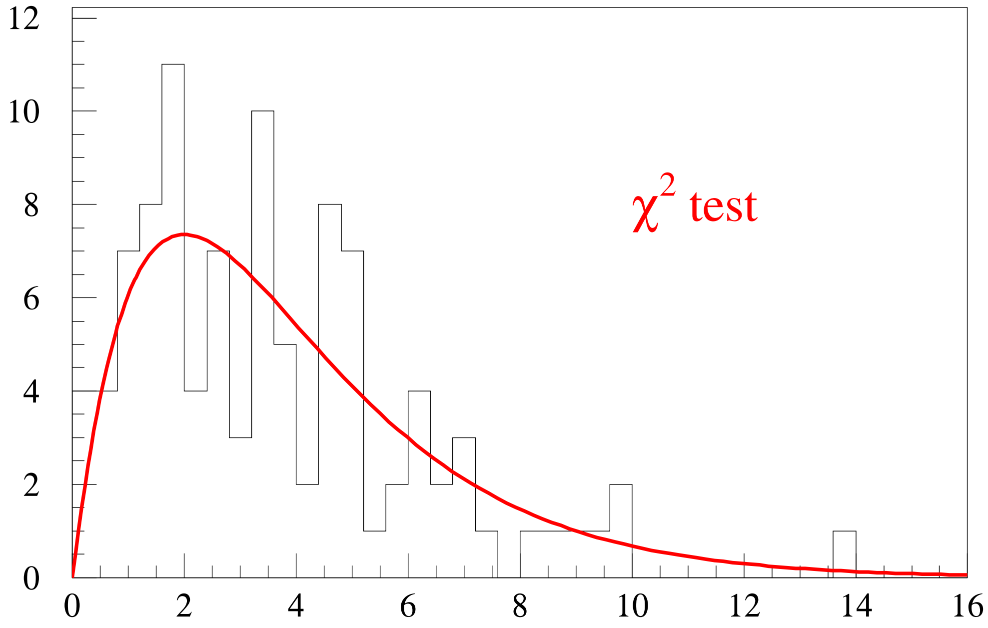
(a) Compare the observed time distribution to the model with $\tau = 1$ s, $\sigma_t = 0.5$ s. To do this, make 5 one-second bins starting at t=-1 s. Show the X^2 distribution for the 100 repititions and compare to the appropriate χ^2 distribution. What is the rejection region for a test of size (α) of 10% ? How many of your 100 experiments fail this?

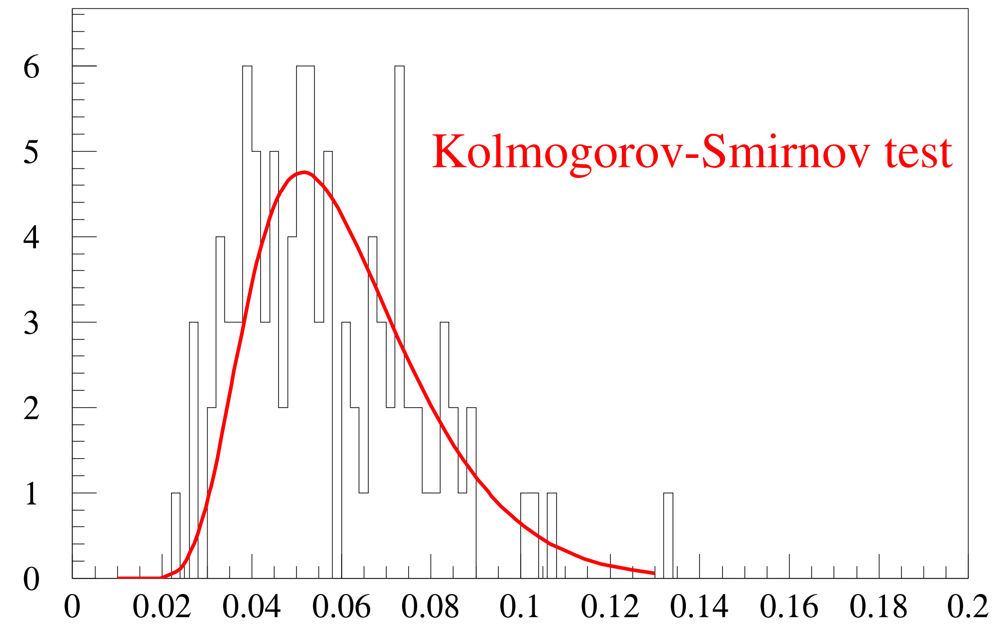
(b) Compare the cumulative distribution with the model for $\tau = 1$ s, $\sigma_t = 0.5$ s. Show the D_n distribution for your data. How many experiments fail a test of size α=10%?
Hint: The cumulative distribution is given by:

$$F_0(t) = \int_{-\infty}^{t} f(t|\tau,\sigma_t)$$

$$= \frac{1}{2}\left(\mathrm{erfc}\left(\frac{-t}{\sqrt{2}\,\sigma_t}\right) - \exp\left(\frac{\sigma_t^2}{2\,\tau^2} - \frac{t}{\tau}\right) \mathrm{erfc}\left(\frac{\sigma_t}{\sqrt{2}\,\tau} - \frac{t}{\sqrt{2}\,\sigma_t}\right)\right)$$

Solution to Exercise 13





Test of Independence

Are two different properties measured in an experiment correlated? The hypothesis under test is then:

$$H_0 \;:\; f(x,y) = f_1(x)\, f_2(y)$$

χ^2 can again be used. This time bin in 2 D:
$n_{ij} \equiv$ the # in bin i of x, bin j of y , $n_{i\bullet} \equiv \sum_j n_{ij}$,
$n_{\bullet j} \equiv \sum_i n_{ij}$

If true probability is given by p_{ij}, then

$$H_0 \;:\; p_{ij} = p_{i\bullet}\, p_{\bullet j} \quad \text{for all } i,j$$

So form the variable,

$$X^2 = \sum_i \sum_j (n_{ij} - n_{i\bullet} n_{\bullet j}/n)^2/(n_{i\bullet} n_{\bullet j}/n)$$

$$= n\{\sum_i \sum_j n_{ij}^2/(n_{i\bullet} n_{\bullet j}) - 1\}$$

Then X^2 follows χ^2 distribution with # of d.o.f: (I-1)(J-1)

Run test for comparing 2 samples

Given two sets of measurements, $(x_1, ..., x_n)$ and $(y_1, ..., y_m)$, where $n \leq m$, we wish to test the hypothesis,

$$H_0 \quad : \quad f_x(x) = f_y(y)$$

To do this, make an ordered list of the combined sample, for example

$$x_1 \; x_2 \; y_1 \; x_3 \; y_2 \; y_3 \; x_4 \; x_5$$

and count the number of *runs* (groups of elements from the same set of measurements). If H_0 is true, there should be a large number of runs. To find the probability to find r runs for two random samples from the same distribution is a problem in combinatorics,

$$\begin{aligned} p(r = 2k) &= 2\frac{\binom{n-1}{k-1}\binom{m-1}{k-1}}{\binom{n+m}{n}}, \quad r \text{ even} \\ p(r = 2k-1) &= \frac{\binom{n-1}{k-2}\binom{m-1}{k-1} + \binom{n-1}{k-1}\binom{m-1}{k-2}}{\binom{n+m}{n}}, \quad r \text{ odd} \end{aligned}$$

The distribution has the mean and variance,

$$\mu_r = \frac{2nm}{n+m} + 1 \qquad V(r) = \frac{2nm(2nm - n - m)}{(n+m)^2(n+m-1)}$$

and for large n, m $(n, m > 10)$, $d = (r - \mu_r)/\sqrt{V(r)}$ approximately follows the standard Gaussian, $G_{(0,1)}$.

Run test to supplement Pearson's χ^2 test

Recall that X^2 is insensitive to the sign of $(n_i - np_i^0)$. An additional test comes from considering the sign of this quantity in subsequent bins and counting the number of runs. In the case where no parameters of the model have been determined from the data, the run test and the χ^2 tests are independent, and so they can be combined into a simple test, and the quantity $u = -2(\ln P_{\chi^2} + \ln p(r))$ will follow χ^2 with 4 d.o.f.

Proof: Suppose x is uniformly distributed in $[0, 1]$, consider $u = -2\ln x$. To work out its distribution function:

$$\begin{aligned} g(u)\, du &= f(x)\, dx \qquad (f(x) = 1) \\ g(u) &= \left|\frac{dx}{du}\right| = \frac{1}{2}e^{-\frac{1}{2}u} \end{aligned}$$

This is the χ^2 distribution function for 2 degrees of freedom.

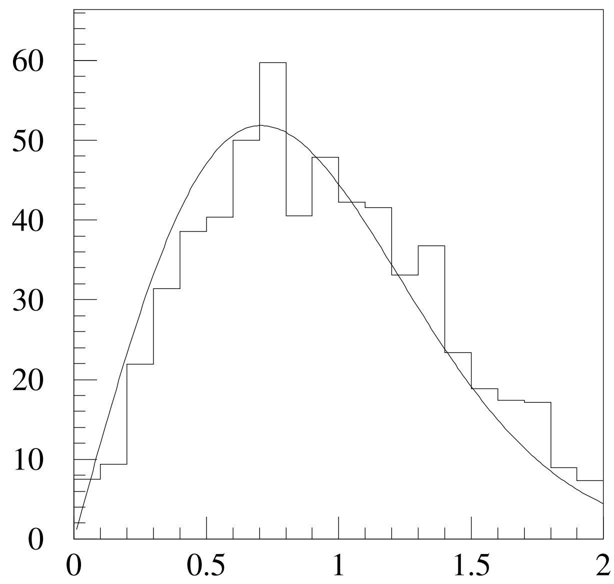
Given x_1 and x_2 which are both uniformly distributed in $[0, 1]$, the variable u,

$$u = -2\,(\ln x_1 + \ln x_2)$$

will follow a χ^2 distribution with 4 degrees of freedom.

Example of Run test with χ^2 test

A simulated data sample is shown below along with a distribution function that was not used to generate the data sample. There are 20 bins shown, and the distribution function was normalized to match the number of events in the data sample.



The value of X^2 for this distribution is 25.2, for 19 d.o.f., resulting in $P_{\chi^2} = 0.16$, which alone gives little reason to suspect the model. There are 7 bins with negative $(n_i - np_i^0)$ and 13 positive bins, with only 5 runs in the signs, so that $p(r = 5) = 0.0074$ is reason to reject the hypothesis.

The combination $u = -2(\ln P_{\chi^2} + \ln p(r)) = 13.5$ corresponds to probability ≈ 0.009.

References for Statistics

1. *Probability and Statistics in Particle Physics*, A. G. Frodesen, O. Skjeggestad, H. Tøfte, Columbia University Press, 1978.
2. *Statistical Methods in Experimental Physics*, W. T. Eadie, D. Drijard, F. E. James, M. Roos, B. Sadoulet, North Holland, 1971.
3. *Statistics for Nuclear and Particle Physicists*, L. Lyons, Cambridge University Press, 1986. (Very elementary.)
4. *Probability, Statistics, and Monte Carlo*, in Review of Particle Properties, *Phys. Rev.* D50 Part I (1994) 1271–1284.