

Rayleigh's Test

Let's do this example in two dimensions - three adds nothing conceptually and is a little more complicated in practice.

The null hypothesis in this case is that we have a set of angles Θ_i (perhaps phases) that are independent and uniformly distributed between zero and 2π . The geometrical picture is that each angle defines a unit vector $(\cos \Theta_i, \sin \Theta_i)$; adding up all these unit vectors is a random walk which results in a vector of length \mathcal{R} .

We have

$$\mathcal{R}^2 = \left(\sum_{i=1}^N \cos \Theta_i \right)^2 + \left(\sum_{i=1}^N \sin \Theta_i \right)^2$$

Write s_i for $\sin \Theta_i$; then for large N we expect the summations $\sum s_i$ to tend to a Gaussian, by the central limit theorem. We also have

$$\left(\sum_i s_i \right)^2 = \sum_i s_i^2 + \sum_{i \neq j} s_i s_j.$$

The second term averages to zero, and in the first we note that the average of \sin^2 is $1/2$. So $2/N(\sum_i s_i)^2$ must be a chi-square variable (with one degree of freedom), since it is a Gaussian variable squared. This is for $N \rightarrow \infty$.

The same argument applies to the cosine terms, so we appear to have the sum of two chi-square variables, each with one degree of freedom; this gives a chi-square with two degrees of freedom *if* the sine and cosine summations produce statistically independent answers. It's not very obvious that this is so. In general, to show that two variables a and b are independent, we have to show that $\text{prob}(a, b) = \text{prob}(a)\text{prob}(b)$, and the probability densities in this case are horrible (see Exercise 3.7).

However, since the distributions are asymptotically Gaussian we can get away with looking at the *correlation coefficient* ρ between $\sum \cos \Theta_i$ and $\sum \sin \Theta_i$. Without cranking through too many details, we can see that this will average to zero because terms like $\cos \Theta_i \sin \Theta_i$ will average to zero. But if a bivariate Gaussian has $\rho = 0$ it follows that it factorizes, and so the variables are independent. This is the result we need to finally establish the asymptotic form of Rayleigh's statistic.

As always with an asymptotic result, it's interesting to ask: how big is infinity? In this case, the answer is about 4. The graph shows the empirical and asymptotic distributions for \mathcal{R} , for just 4 angles. The agreement is remarkable.

Of course, for a test we are usually interested in the wings of the distribution. The second Figure shows the empirical and asymptotic cumulative distributions; the differences are quite appreciable. Evidently we will need larger and larger N if we wish to use the asymptotic form to test at high levels of confidence.

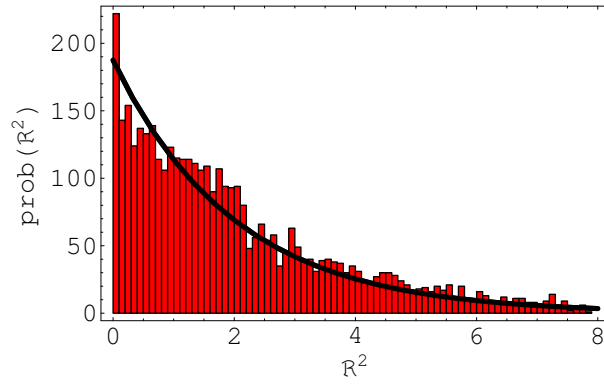


Figure 1: Empirical and asymptotic distributions for \mathcal{R} for $N = 4$ and 4000 repetitions of the experiment.

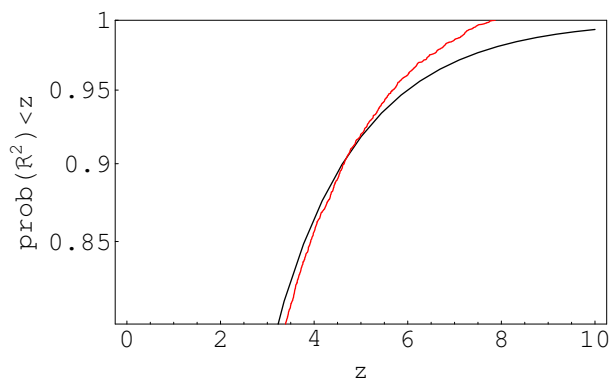


Figure 2: Empirical and asymptotic cumulative distributions for \mathcal{R} for $N = 4$ and 4000 repetitions of the experiment.