Chapter 13
Model Comparison

Abstract The availability of alternative models to fit a dataset requires a quantitative method for comparing the goodness of fit to different models. For Gaussian data, a lower reduced χ^2 of one model with respect to another is already indicative of a better fit, but the outstanding question is whether the value is *significantly* lower, or whether a lower value can be just the result of statistical fluctuations. For this purpose we develop the distribution function of the F statistic, useful to compare the goodness of fit between two models and the need for an additional "nested" model component, and the Kolmogorov–Smirnov statistics, useful in providing a quantitative measure of the goodness of fit, and in comparing two datasets regardless of their fit to a specific model.

13.1 The *F* Test

For Gaussian data, the χ^2 statistic is used for determining if the fit to a given parent function $y(x)$ is acceptable. It is possible that several different parent functions yield a goodness of fit that is acceptable. This may be the case when there are alternative models to explain the experimental data, and the data analyst is faced with the decision to determine what model best fits the experimental data. In this situation, the procedure to follow is to decide first a confidence level that is considered acceptable, say 90 or 99 %, and discard all models that do not satisfy this criterion. The remaining models are all acceptable, although a lower χ^2_{min} certainly indicates a better fit.

The first version of the F test applies to independent measurements of the χ^2 fit statistic, and its application is therefore limited to cases that compare different datasets. A more common application of the F test is to compare the fit of a given dataset between two models that have a *nested* component, i.e., one model is a simplified version of the other. For nested model components one can determine whether the additional component is really needed to fit the data.

M. Bonamente, *Statistics and Analysis of Scientific Data*, Graduate Texts in Physics, DOI 10.1007/978-1-4939-6572-4_13

13.1.1 F-Test for Two Independent χ^2 Measurements

Consider the case of two χ^2_{min} values obtained by fitting data from a given experiment to two different functions, $y_1(x)$ and $y_2(x)$. If both models equally well approximate the parent model, then we would expect that the two values of χ^2 would be similar, after taking into consideration that they may have a different number of degrees of freedom. But if one is a better approximation to the parent model, then the value of χ^2 for such model would be significantly lower than for the other. We therefore want to proceed to determine whether both χ^2_{min} statistics are consistent with the null hypothesis that the data are drawn from the respective model. The statistic to use to compare the two values of χ^2 must certainly also take into account the numbers of degrees of freedom, which is related to the number of model parameters used in each determination of χ^2. In fact, a larger number of model parameters may result in fact result in a lower value of χ^2_{min}, simply because of the larger flexibility that the model has in following the data. For example, a dataset of N points will always be fitted perfectly by a polynomial having N terms, but this does not mean that a *simpler* model may not be just as good a model for the data, and the underlying experiment.

Following the theory described in Sect. 7.4, we define the F statistic as

$$F = \frac{\chi^2_{1,min}/f_1}{\chi^2_{2,min}/f_2} = \frac{\chi^2_{1,min,red}}{\chi^2_{2,min,red}}, \tag{13.1}$$

where f_1 and f_2 are the degrees of freedom of $\chi^2_{1,min}$ and $\chi^2_{2,min}$. Assuming that the two χ^2 statistics are *independent*, then F will be distributed like the F statistic with f_1, f_2 degrees of freedom, having a mean of approximately 1 [see (7.22) and (7.24)].

There is an ambiguity in the definition of which of the two models is labeled as 1 and which as 2, since two numbers can be constructed that are the reciprocal of each other, $F_{12} = 1/F_{21}$. The usual form of the F-test is that in which the value of the statistic is $F > 1$, and therefore we choose the largest of F_{12} and F_{21} to implement a one-tailed test of the null hypothesis with significance p,

$$1 - p = \int_{F_{crit}}^{\infty} f_F(f, x)dx = P(F \geq F_{crit}). \tag{13.2}$$

Critical values F_{crit} are reported in Tables A.8, A.9, A.10, A.11, A.12, A.13, A.14, and A.15 for various confidence levels p.

The null hypothesis is that the two values of χ^2_{min} are distributed following a χ^2 distributions; this, in turn, means that the respective fitting functions used to determine each χ^2_{min} are *both* good approximations of the parent distribution. Therefore the test based on this distribution can reject the hypothesis that both fitting functions are the parent distribution. If the test rejects the hypothesis at the desired confidence level, then only one of the models will still stand after the test—the one at the denominator with the lowest reduced χ^2—even if the value of χ^2_{min} alone was not able to discriminate between the two models.

Example 13.1 Consider the radius vs. ratio data of Table 6.1 (see also Problem 11.1). The linear fit to the entire dataset is not acceptable, and therefore a linear model for all measurements must be discarded. If we consider measurements 1 through 5, 6 through 10, and 11 through 15, a linear fit to these two subsets results in the values of best-fit parameters and χ^2 shown in the table, along with the probability to exceed the value of the fit statistic.

Measurements	a	b	χ^2_{min}	Probability
1–5	0.97 ± 0.09	-0.0002 ± 0.0002	5.05	0.17
6–10	1.27 ± 0.22	-0.0007 ± 0.0011	6.19	0.10
10–15	0.75 ± 0.09	-0.0002 ± 0.0003	18.59	0.0

The third sample provides an unacceptable fit to the linear model, and therefore this subset cannot be further considered. For the first two samples, the fits are acceptable at the 90 % confidence level, and we can construct the F statistic as

$$F = \frac{\chi^2_{min}(6-10)}{\chi^2_{min}(1-5)} = 1.23.$$

Both χ^2 have the same number of degrees of freedom (3), and Table A.13 shows that the value of 1.23 is certainly well within the 90 % confidence limit for the F statistics ($F_{crit} \simeq 5.4$). This test shows that both subsets are equally well described by a linear fit, and therefore the F-test cannot discriminate between them.

To illustrate the power of the F-test, assume that there is another set of five measurements that yield a $\chi^2_{min} = 1.0$ when fit to a linear model. This fit is clearly acceptable in terms of its χ^2 probability. Constructing an F statistic between this new set and set 6–10, we would obtain

$$F = \frac{\chi^2_{min}(6-10)}{\chi^2_{min}(new)} = 6.19.$$

In this case, the value of F is *not* consistent at the 90 % level with the F distribution with $f_1 = f_2 = 3$ degrees of freedom (the measured value exceeds the critical value). The F-test therefore results in the conclusion that, at the 90 % confidence level, the two sets are not equally likely to be drawn from a linear model, with the new set providing a better match. ◇

It is important to note that the hypothesis of independence of the two χ^2 is not justified if the same data are used for both statistics. In practice, this means that the F statistic *cannot* be used to compare the fit of a given dataset to two different models. The test can still be used to test whether two different datasets, derived from the same experiment but with independent measurements, are equally well described by the same parametric model, as shown in the example above. In this case, the null hypothesis is that both datasets are drawn from the same parent model, and a rejection of the hypothesis means that both datasets cannot derive from the same distribution.

13.1.2 F-Test for an Additional Model Component

Consider a model $y(x)$ with m adjustable parameters, and another model $\overline{y}(x)$ obtained by fixing p of the m parameters to a reference (fixed) value. In this case, the $\overline{y}(x)$ model is said to be *nested* into the more general model, and the task is to determine whether the additional p parameters of the general model are required to fit the data.

Example 13.2 An example of nested models are polynomial models. The general model can be taken as a polynomial of second order,

$$y(x) = a + bx + cx^2$$

and the nested model as a linear model,

$$y(x) = a + bx.$$

The nested model is obtained from the general model with $c = 0$ and has one fewer degree of freedom than the general model. ◇

Following the same discussion as in Chap. 10, we can say that

$$\begin{cases} \chi^2_{min} \sim \chi^2(N-m) & \text{(full model)} \\ \overline{\chi}^2_{min} \sim \chi^2(N-m+p) & \text{(“nested” model).} \end{cases} \tag{13.3}$$

Clearly $\chi^2_{min} < \overline{\chi}^2_{min}$ because of the additional free parameters used in the determination of χ^2_{min}. A lower value of χ^2_{min} does not necessarily mean that the additional parameters of the general model are required. The nested model can in fact achieve an equal or even better fit relative to the parent distribution of the fit statistic, i.e., a lower χ^2_{red}, because of the larger number of degrees of freedom. In general, a model with fewer parameters is to be preferred to a model with larger number of parameters because of its more economical description of the data, provided that it gives an acceptable fit.

In Sect. 10.3 we discussed that, when comparing the true value of the fit statistic χ^2_{true} for the parent model to the minimum χ^2_{min} obtained by minimizing a set of p free parameters, $\Delta\chi^2 = \chi^2_{true} - \chi^2_{min}$ and χ^2_{min} are independent of one another, and that $\Delta\chi^2$ is distributed like χ^2 with p degrees of freedom. There are situations in which the same properties apply to the two χ^2 statistics described in (13.3), such that the statistic $\Delta\chi^2$ is distributed like

$$\Delta\chi^2 = \overline{\chi}^2_{min} - \chi^2_{min} \sim \chi^2(p), \tag{13.4}$$

and it is independent of χ^2_{min}. One such case of practical importance is precisely the one under consideration, i.e., when there is a nested model component described by parameters that are independent of the other model parameters. A typical example is an additional polynomial term in the fit function, as illustrated in the example above.

In this case, the null hypothesis we test is that $y(x)$ and $\overline{y}(x)$ are equivalent models, i.e., adding the p parameters does *not* constitute a significant change or improvement to the model. Under this hypothesis we can use the two independent statistics $\Delta\chi^2$ and χ^2_{min}, and construct a bona fide F statistic as

$$F = \frac{\Delta\chi^2/p}{\chi^2_{min}/(N-m)}. \tag{13.5}$$

This statistic tests the null hypothesis using an F distribution with $f_1 = p, f_2 = N-m$ degrees of freedom. A rejection of the hypothesis indicates that the two models $y(x)$ and $\overline{y}(x)$ are not equivalent. In practice, a rejection constitutes a positive result, indicating that the additional model parameters in the nested component *are* actually needed to fit the data. A common situation is when there is a single additional model parameter, $p = 1$, and the corresponding critical values of F are reported in Table A.8. A discussion of certain practical cases in which additional model components may obey (13.4) is provided in a research article by Protassov [36].

Example 13.3 The data of Table 10.1 and Fig. 10.2 are well fit by a linear model, while a constant model appears not to be a good fit to all measurements. Using only the middle three measurements, we want to compare the goodness of fit to a linear model, and that to a constant model, and determine whether the addition of the b parameter provides a significant improvement to the fit.

The best-fit linear model has a $\chi^2_{min} = 0.13$ which, for $f_2 = N-m = 1$ degree of freedom, with a probability to exceed this value of 72 %, i.e., it is an excellent fit. A constant model has a $\overline{\chi}^2_{min} = 7.9$, which, for 2 degrees of freedom, has a probability to exceed this value of $\geq$0.01, i.e., it is acceptable at the 99 % confidence level, but not at the 90 % level. If the analyst requires a level of confidence $\leq$90 %, then the constant model should be discarded, and no further analysis of the experiment is needed. If the analyst can accept a 99 % confidence level, we can determine whether the improvement in χ^2 between the constant and the linear model is significant. We construct the statistic

$$F = \frac{\overline{\chi}^2_{min} - \chi^2_{min}}{\chi^2_{min}} \frac{1}{1} = 59.4$$

which, according to Table A.8 for $f_1 = 1$ and $f_2 = 1$, is significant at the 99 % (and therefore 95 %) confidence level, but not at 90 % or lower. In fact, the critical value of the F distribution with $f_1 = 1, f_2 = 1$ at the 99 % confidence level is $F_{crit} = 4,052$. Therefore a data analyst willing to accept a 99 % confidence level should conclude that the additional model component b is *not* required, since there is ≥1 % (actually, ≥5 %) probability that such an improvement in the χ^2 statistic is due by chance, and not by the fact that the general model is truly a more accurate description of the data. ◇

The example above illustrates the principle of simplicity or parsimony in the analysis of data. When choosing between two models, both with an acceptable fit statistic at the same confidence level (in the previous example at the 99 % level), one should prefer the model with fewer parameters, even if its fit statistic (e.g., the reduced χ^2_{min}) is inferior to that of the more complex model. This general guiding principle is sometimes referred to as *Occam's razor*, after the Middle Ages philosopher and Franciscan friar William of Occam.

13.2 Kolmogorov–Smirnov Tests

Kolmogorov–Smirnov tests are a different method for the comparison of a one-dimensional dataset to a model, or for the comparison of two datasets to one another. The tests make use of the cumulative distribution function, and are applicable to measurements of a single variable X, for example to determine if it is distributed like a Gaussian. For two-variable dataset, the χ^2 and F tests remain the most viable option.

The greatest advantage the Kolmogorov–Smirnov test is that it does not require the data to be binned, and, for the case of the comparison between two dataset, it does not require any parameterization of the data. These advantages come at the expense of a more complicated mathematical treatment to find the distribution function of the test statistic. Fortunately, numerical tables and analytical approximations make these tests manageable.

13.2.1 Comparison of Data to a Model

Consider a random variable X with cumulative distribution function $F(x)$. The data consist of N measurements, and for simplicity we assume that they are in increasing order, $x_1 \leq x_2 \leq \ldots \leq x_N$. This condition can be achieved by re-labelling the measurements, which preserves the statistical properties of the data. The goal is to construct a statistic that describes the difference between the sample distribution of the data and a specified distribution, to test whether the data are compatible with this distribution.

Start with the sample cumulative distribution

$$F_N(x) = \frac{1}{N}[\text{\# of measurements} \leq \text{x}]. \tag{13.6}$$

By definition, $0 \leq F_N(x) \leq 1$. The test statistic we want to use is defined as

$$D_N = \max_x |F_N(x) - F(x)|, \tag{13.7}$$

where $F(x)$ is the parent distribution, and the maximum value of the difference between the parent distribution and the sample distribution is calculated for all values in the support of X.

One of the remarkable properties of the statistic D_N is that it has the same distribution for any underlying distribution of X, provided X is a continuous variable. The proof that D_N has the same distribution regardless of the distribution of X illustrates the properties of the cumulative distribution and of the quantile function presented in Sect. 4.8.

Proof We assume that $F(x)$ is continuous and strictly increasing. This is certainly the case for a Gaussian distribution, or any other distribution that does not have intervals where the distribution functions is $f(x) = 0$. We make the change of variables $y = F(x)$, so that the measurement x_k corresponds to $y_k = F(x_k)$. This change of variables is such that

$$F_N(x) = \frac{(\text{\# of } x_i < x)}{N} = \frac{(\text{\# of } y_k < y)}{N} = U_N(y)$$

where $U_N(y)$ is the sample cumulative distribution of Y and $0 \leq y \leq 1$. The cumulative distribution of Y is

$$U(y) = P(Y < y) = P(X < x) = F(x) = y.$$

The fact that the cumulative distribution is $U(y) = y$ shows that Y is a uniform distribution between 0 and 1. As a result, the statistic D_N is equivalent to

$$D_N = \max_{0 \leq y \leq 1} |U_N(y) - U(y)|$$

where Y is a uniform distribution. Since this is true no matter the original distribution X, D_N has the same distribution for any X. Note that this derivation relies on the continuity of X, and this assumption must be verified to apply the resulting Kolmogorov–Smirnov test. □

The distribution function of the statistic D_N was determined by Kolmogorov in 1933 [25], and it is not easy to evaluate analytically. In the limit of large N, the cumulative distribution of D_N is given by

$$\lim_{N \to \infty} P(D_N < z/\sqrt{N}) = \sum_{r=-\infty}^{\infty} (-1)^r e^{-2r^2 z^2} \equiv \Phi(z). \tag{13.8}$$

Table 13.1 Critical points of the Kolmogorov distribution D_N for large values of N

Confidence level	
p	$\sqrt{N}D_N$
0.50	0.828
0.60	0.895
0.70	0.973
0.80	1.073
0.90	1.224
0.95	1.358
0.99	1.628

The function $\Phi(z)$ can also be used to approximate the probability distribution of D_N for small values of N, using

$$P(D_N < z/(\sqrt{N} + 0.12 + 0.11/\sqrt{N})) \simeq \Phi(z). \tag{13.9}$$

A useful numerical approximation for $P(D_N < z)$ is also provided in [30].

The probability distribution of D_N can be used to test whether a sample distribution is consistent with a model distribution. Critical values of the D_N distribution with probability p,

$$P(D_N \leq T_{crit}) = p \tag{13.10}$$

are shown in Table 13.1 in the limit of large N. For small N, critical values of the D_N statistic are provided in Table A.25. If the measured value for D_N is greater than the critical value, then the null hypothesis must be rejected, and the data are not consistent with the model. The test allows no free parameters, i.e., the distribution that represents the null hypothesis must be fully specified.

Example 13.4 Consider the data from Thomson's experiment to measure the ratio m/e of an electron (page 23). We can use the D_N statistic to test whether either of the two measurement of the variable m/e is consistent with a given hypothesis. It is necessary to realize that the Kolmogorov–Smirnov test applies to a fully specified hypothesis H_0, i.e., the parent distribution $F(x)$ cannot have free parameter that are to be determined by a fit to the data. We use a fiducial hypothesis that the ratio is described by a Gaussian distribution of $\mu = 5.7$ (the true value in units of 10^7 g Coulomb^{-1}, though the units are unnecessary for this test), and a variance of $\sigma^2 = 1$. Both measurements are inconsistent with this model, as can be seen from Fig. 13.1. See Problem 13.1 for a quantitative analysis of the results. ◇

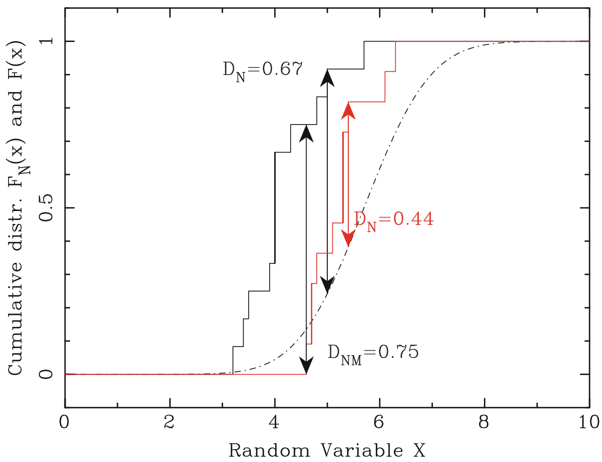

Fig. 13.1 Kolmogorov–Smirnov test applied to the measurements of the ratio m/e from Thomson's experiments described on page 23. The *black line* corresponds to the measurements for Tube 1, and the *red line* to those of Tube 2 (measurements have been multiplied by 10^7). The *dot-dashed line* is the cumulative distribution of a Gaussian with $\mu = 5.7$ (the correct value) and a fiducial variance of $\sigma^2 = 1$

13.2.2 Two-Sample Kolmogorov–Smirnov Test

A similar statistic can be defined to compare two datasets:

$$D_{NM} = \max_x |F_M(x) - G_N(x)| \tag{13.11}$$

where $F_M(x)$ is the sample cumulative distribution of a set of M observations, and $G_N(x)$ that of another independent set of N observations; in this case, there is no parent model used in the testing. The statistic D_{NM} measures the maximum deviation between the two cumulative distributions, and by nature it is a discrete distribution. In this case, we can show that the distribution of the statistic is the same as in (13.9), provided that the change

$$N \rightarrow \frac{MN}{M+N}$$

is made. This number can be considered as the effective number of datapoints of the two distributions. For the two-sample Kolmogorov–Smirnov D_{NM} test we can therefore use the same table as in the Kolmogorov–Smirnov one-sample test, provided N is substituted with $MN/(M+N)$ and that N and M are both large.

As N and M become large, the statistic approaches the following distribution:

$$\lim_{N,M\to\infty} P\left(D_{NM} < z/\sqrt{\frac{MN}{M+N}}\right) = \Phi(z). \tag{13.12}$$

Proof We have already shown that for a sample distribution with M points,

$$F_M(x) - F(x) = U_M(y) - U(y),$$

where U is a uniform distribution in (0,1). Since

$$F_M(x) - G_N(x) = F_M(x) - F - (G_N(x) - G),$$

where $F = G$ is the parent distribution, it follows that $F_M(x) - G_N(x) = U_N - V_N$, where U_M and V_N are the sample distribution of two uniform variables. Therefore the statistic

$$D_{NM} = \max_x |F_M(x) - G_N(x)|$$

is independent of the parent distribution, same as for the statistic D_N.

Next we show how the factor $\sqrt{1/N + 1/M}$ originates. It is clear that the expectation of $F_M(x) - G_N(x)$ is zero, at least in the limit of large N and M; the second moment can be calculated as

$$\begin{aligned} E[(F_M(x) - G_N(x))^2] &= E[(F_M(x) - F(x))^2] \\ +E[(G_N(x) - G(x))^2] &+ 2E[(F_M(x) - F(x))(G_N(x) - G(x))] \\ &= E[(F_M(x) - F(x))^2] + E[(G_N(x) - G(x))^2] \end{aligned}$$

In fact, since $F_M(x) - F(x)$ is independent of $G_N(x) - G(x)$, their covariance is zero. Each of the two remaining terms can be evaluated using the following calculation:

$$E\left[(F_M(x) - F(x))^2\right] = E\left[\frac{1}{M}(\{\#\text{ of } x_i\text{'s} < \text{x}\} - MF(x))^2\right] =$$
$$\frac{1}{M^2}E\left[(\{\#\text{ of } x_i\text{'s} < \text{x}\} - E[\{\#\text{ of } x_i\text{'s} < \text{x}\}])^2\right].$$

For a fixed value of x, the variable $\{\#\text{ of } x_i\text{'s} < \text{x}\}$ is a binomial distribution in which "success" is represented by one measurement being $< x$, and the probability of success is $p = F(x)$. The expectation in the equation above is therefore equivalent to the variance of a binomial distribution with M tries, for which $\sigma^2 = Mp(1-p)$, leading to

$$E\left[(F_M(x) - F(x))^2\right] = \frac{1}{M} F(x)(1 - F(x)).$$

It follows that

$$E[(F_M(x) - G_N(x))^2] = \left(\frac{1}{M} + \frac{1}{N}\right) F(x)(1 - F(x))$$

A simple way to make the mean square of $F_M(x) - G_N(x)$ independent of N and M is to divide it by $\sqrt{1/M + 1/N}$. This requirement is therefore a necessary condition for the variable $\sqrt{NM/(N+M)} D_{NM}$ to be independent of N and M.

Finally, we show that $\sqrt{NM/(N+M)} D_{NM}$ is distributed in the same way as $\sqrt{N} D_N$, at least in the asymptotic limit of large N and M. Using the results from the D_N distribution derived in the previous section, we start with

$$\max_x \left| \sqrt{\frac{MN}{M+N}} (F_M(x) - G_N(x)) \right| = \max_{0 \le y \le 1} \left| \sqrt{\frac{MN}{M+N}} (U_M - V_N)) \right|.$$

The variable can be rewritten as

$$\sqrt{\frac{MN}{M+N}} (U_M - U + (V - V_N)) = \sqrt{\frac{N}{M+N}} (\sqrt{M}(U_M - U)) + \sqrt{\frac{M}{M+N}} (\sqrt{N}(V_N - V)).$$

Using the central limit theorem, it can be shown that the two variables $\alpha = \sqrt{M}(U_M - U)$ and $\beta = \sqrt{N}(V_N - V)$ have the same distribution, which tends to a Gaussian in the limit of large M. We then write

$$\sqrt{\frac{MN}{M+N}} (F_M(x) - G_N(x)) = \sqrt{\frac{N}{M+N}} \alpha + \sqrt{\frac{M}{M+N}} \beta$$

and use the property that, for two independent and identically distributed Gaussian variables α and β the variable $a \cdot \alpha + b \cdot \beta$ is distributed like α, provided that $a^2 + b^2 = 1$. We therefore conclude that, in the asymptotic limit,

$$D_{NM} = \max_x \left| \sqrt{\frac{MN}{M+N}} (F_M(x) - G_N(x)) \right| \sim \max_x \left| \sqrt{N}(V_N - V) \right| = D_N.$$

□

Example 13.5 We can use the two-sample Kolmogorov–Smirnov statistic to compare the data from Tube #1 and Tube #2 of Thomson's experiment to measure the ratio m/e of an electron (page 23). The result, shown in Fig. 13.1, indicates that the two measurements are not in agreement with one another. See Problem 13.2 for a quantitative analysis of this test. ◇

Summary of Key Concepts for this Chapter

- □ *F Test*: A test to compare two independent χ^2 measurements,

$$F = \chi^2_{1,red}/\chi^2_{2,red}.$$

- □ *F Test for additional component*: The significance of an additional model component with p parameters can be tested using

$$F = \frac{\Delta\chi^2/p}{\chi^2_{min}/(N-m)}$$

when the additional component is nested within the general model.
- □ *Kolmogorov–Smirnov test*: A non-parametric test to compare a one-variable dataset to a model or two datasets with one another.

Problems

13.1 Using the data from Thomson's experiment at page 23, determine the values of the Kolmogorov–Smirnov statistic D_N for the measurement of Tube #1 and Tube #2, when compared with a Gaussian model for the measurement with $\mu = 5.7$ and $\sigma^2 = 1$. Determine at what confidence level you can reject the hypothesis that the two measurements are consistent with the model.

13.2 Using the data from Thomson's experiment at page 23, determine the values of the two-sample Kolmogorov–Smirnov statistic D_{NM} for comparison between the two measurements. Determine at what confidence level you can reject the hypothesis that the two measurements are consistent with one another.

13.3 Using the data of Table 10.1, determine whether the hypothesis that the last three measurements are described by a simple constant model can be rejected at the 99 % confidence level.

13.4 A given dataset with $N = 5$ points is fit to a linear model, for a fit statistic of $\overline{\chi}^2_{min}$. When adding an additional nested parameter to the fit, $p = 1$, determine by how much should the χ^2_{min} be reduced for the additional parameter to be significant at the 90 % confidence level.

13.5 A dataset is fit to model 1, with minimum χ^2 fit statistic of $\chi_1^2 = 10$ for 5 degrees of freedom; the same dataset is also fit to another model, with $\chi_2^2 = 5$ for 4 degrees of freedom. Determine which model is acceptable at the 90 % confidence, and whether the *F* test can be used to choose one of the two models.

13.6 A dataset of size *N* is successfully fit with a model, to give a fit statistic $\overline{\chi}^2_{min}$. A model with a nested component with 1 additional independent parameter for a total of *m* parameters is then fit to χ^2_{min}, providing a reduction in the fit statistic of $\Delta\chi^2$. Determine what is the minimum $\Delta\chi^2$ that, in the limit of a large number of degrees of freedom, provides 90 % confidence that the additional parameter is significant.