**Easy computation of the Bayes Factor to fully quantify Occam’s razor**

D.J. Dunstan,\* J. Crowne, and A.J. Drew

School of Physics and Astronomy, Queen Mary University of London, London E1 4NS, UK.

**Supplementary Information**

**§S1. Least-squares and Likelihood:** Let the probability (pdf) that a residual is of magnitude *r* be Gaussian,

 (S1)

so that when *Pr*(*r*) is Gaussian, as least-squares fitting assumes, minimising the sum of the squares of the residuals is the same thing as maximising *L* or ln*L*. Any other pdf *Pr*(*r*) can be used.

**§S2. Parameter Covariance Matrix:** Some least-squares fitting routines return the parameter covariance matrix **Cov**. When that is not so, it can be readily calculated. The residuals are expressed as functions of the parameters *pi*, and then the likelihood *L* can be calculated (§S1). Then the fitted values *pi*0 for all parameters except one, *pj*, or two, *pj* and *pk*, can be substituted in to give the *m* functions *L*(*pj*) and the ½(*m*2 – *m*) functions *L*(*pj*, *pk*). From these, the second derivatives can be calculated quite quickly. The Hessian matrix is

 (S2)

and then **Cov** = –**H–1.**

**§3. Ranges of parameters:** Ranges should be set *a priori*, using all knowledge that may be available, but before any inspection of the data. No great precision is needed, since it is the logarithms of the ranges that matter for lnBF. Choosing the ranges has been criticized as introducing subjectivity; however, the key point is that the ranges are given quantitatively and should be justified.2,15 The values used can be defended or opposed on objective grounds. For physically-meaningless parameters such as polynomial coefficients, it is helpful to consider the data rescaled to the ranges 0 to 1 in *x* and *y* (that can be done automatically without inspecting the data). For example, if some data are to be fitted with *a* + *bx* + *cx*2 and there is no prior information about the parameters, when the data are rescaled to the ranges 0 – 1 on both axes, all parabolae that could fit in this would be encompassed by the ranges 0 ≤ *a* ≤ 1, –4 ≤ *b* ≤ 4 and –4 ≤ *c* ≤ 4. The rescaled data can be fitted and these ranges used, or these ranges can be rescaled for the original data. For the real examples of Section 4, see §§6-8 below.

**§4. Theory of Eq.1:** Figure S1 shows why the square-root of the determinant of the parameter covariance matrix in Eq.1 is used in place of the product of the parameter uncertainties. Data from Fig.2 is shown in Fig. S1a with the LS fit using Δ*Eg* = *a*0 + *bP* + *cP*2 (*a* is fixed at its fitted value *a*0 to make a two-parameter problem). The Eqn.S5 fit is also shown as the difference between the two fits is clearer here than in Fig.2. The MLI comes from the integral of *L* over the parameter space. From the LS fit, we have the fitted values *b*0 and *c*0at which *L* = *L*max. We write *L* as a function of the residuals with *b* and *c* as variables, and it is easy to calculate a few values of *L* with either *b* or *c* varying and the other at its fitted value. These values are plotted in Fig.S1b, normalised to unity peak height, for parameter *p* = *b* and for *p* = *c*, and with the *x*-axis normalised by the uncertainties σ*b*and σ*c* which are returned by the LS fit. These functions are accurate Gaussians, justifying the use of the Laplace approximation.

 

 

 

**Fig.S1.**  In (a), data from Fig.2 (*Eg*(*P*) of GaAs from Goñi *et al.*16) are shown after subtraction of the straight line *a*0 + 9.5*P* to make the curvature more visible. Two least-squares fits are shown, quadratic (dashed red line) and Eq.S5 (solid blue line). In (b), the likelihood of the quadratic fit is plotted as a function of the polynomial parameters *pi* (*i* = *b*, *c*) scaled by their uncertainties σ*i*, and fitted with Gaussians of width δ*b* and δ*c*. For clarity the data and fit for *c* are shifted up by 0.2. In (c) and (d), *L*(*b*, *c*) is plotted with normalised axes, and the heavy blue ellipse is the contour at the height *e*–½ . On the contour plot in (d), the area of the large circle of radius 1 corresponds to the product of the uncertainties σ*b*σ*c*. Rotating the parameter basis set to *v*1 and *v*2 gives the corresponding plots of (e) and (f) from which the normalised products and are equal and both correctly give the area of the contour line.

Notice that the widths of the two Gaussians of Fig.S1b, δ*b* and δ*c*, are much less than the uncertainties σ*b*and σ*c* by which the *x*-axis is normalised. Using the product σ*b*σ*c*will give a gross overestimate of the two-dimensional integral of *L*. Equally, using the product δ*b*δ*c* will give a gross underestimate. Plotting the full function *L*(*b*, *c*) (Fig.S1c) shows why this happens. The function is Gaussian about its peak in all directions, but it is skew to the axes. The skewness expresses the parameter covariance. If the random errors in the data were to put, e.g., the initial gradient *b* higher, *c* will be fitted with a higher value as well. This corresponds to positive off-diagonal elements in the correlation and covariance matrices. To obtain the volume of *L*(*b*, *c*), we need the area of the elliptical contour shown in Fig.S1c. The contour map of *L*(*b*, *c*) (Fig.S1d) shows that σ*b*σ*c* gives the area of the circle radius 1, and that using δ*b*δ*c* instead gives the area of the small circle of radius 0.21.

The correct estimate can be found by taking linear combinations of the parameters *pi* corresponding to a rotation in parameter space, such that the new parameters *vi* will have zero correlation or off-diagonal covariance – that is, to change the parameter basis set. These new parameters are found by diagonalising the parameter covariance matrix. (If the fitting routine used does not return the parameter covariance matrix, it is not difficult to calculate; see SI §S2). The resulting eigenvectors specify the *vi*. Repeating the LS fit with these parameters, we get the identical fit and we can recover the original *b*0 and *c*0 by a back transform from the fitted values of *vi*. Plotting *L*(*v*1, *v*2) in Fig.S1e (contour plot in Fig.S1f) we see that now the product of the uncertainties is the same as the product δ*v*1δ*v*2 of the widths of the Gaussians *L*(*v*1*f*, *v*2) and *L*(*v*1, *v*2 *f*) and either gives us exactly the integral of *L* over the parameter space that we require.

To complete the calculation of the MLI with *v*1 and *v*2, we would need their parameter pdfs in parameter space. It is usual to estimate these as top-hat or rectangle functions, of width Δ*pi* and area unity. This is justified as a consequence of Laplace’s Principle of Indifference.15 Then the MLI is the integral of *L* that we have just found, divided by the product of the parameter ranges, Δ*v*1Δ*v2*. However, we do not need to estimate the ranges Δ*vi* if we have already estimated the ranges Δ*b* and Δ*c*, because the product of the ranges is an area in either parameter basis, and areas are invariant under rotation. Better still, we do not need to make this transformation of the basis set nor to carry out the fit in this basis. The diagonal elements of the parameter covariance matrix in this basis, Cov*v*, are the squares of the uncertainties on the parameters, so what we require for the calculation of the MLI in this basis is the product of these elements. Since the off-diagonal elements are zero, this product is also the determinant of Cov*v*. But the determinant of a matrix is invariant to rotations, and so it is sufficient to take the determinant of the parameter covariance matrix in the *p* basis, Cov*p*. That leads to Eq.1 for the MLI.

**§6. Example 1: Muon data fits.**

Lorentzian peaks were used, with peak positions *Pi*, widths *Wi* and areas *Ai*:

 (S7)

Light-induced effects were added by adding to each parameter *p* a term *z*δL*p*, where δL*p* is the change in *p* under illumination, and writing the data in 3D as *yi*(*xi*,*zi*) with *zi* = 0 for data in the dark and *zi* = 1 for the photoexcited data. This enabled least-squares fitting to all data simultaneously, using the Mathematica function NonlinearModelFit[data, func, {x,z}]. Linear baselines were added, defined by the four corner points, *y*00 at (*xlo*,0), *yx*0 at (*xhi*,0), *y*0*z*at (*xlo*,1) and *yxz*at (*xhi*,1).

 To calculate the MLI, ranges were set as follows. For the first peak, Δ*P*1= 400 as it could be anywhere in the data. The second peak could not be far away as there was doubt whether one peak or two would be the best fit, so Δ*P*2= 100. For widths and areas, Δ*W* = 50 and Δ*A* = 40 considering the data range. Light-induced effects were expected to be small, so ΔδL*P* = 100, and for the widths and areas a 20% change was considered the most that could be expected, so ΔδL*W* = 10 and ΔδL*A* = 8. For the four background values, given that the positron background had already been subtracted, 20% of the data range was considered the largest plausible error in that subtractions, so Δ*yij* = 0.03. When we add the third peak, the same values were used except for Δ*P*3= 60 (reduced because there isn’t much room), Δ*A*3 = 4 because it has to be a weak peak, and accordingly ΔδL*A*3 = 1.

**§7. Example 2: Pressure dependence of the GaAs bandgap.**

The Murnaghan equation of state derives from standard interatomic potentials and finite elasticity theory, using the approximation that the pressure derivative of the bulk modulus *B*0­ is constant. Then the equation for the volume as a function of pressure is

 (S4)

Assuming that the bandgap of GaAs, , is linear with the density,28

 (S5)

where Ξ/*B*0 is equivalent to *b* in the polynomial fit, . For both fits we add a constant *a* to allow for uncertainty in the ambient-pressure value , so the quadratic fit has three terms and is denoted QP3.

 The physical meanings of *a* and of *b* ≡ Ξ/*B*0 are the same for all the fits. We choose Δ*a* = 100 meV and Δ*b* = ΔΞ/*B*0 = 5 meV kbar–1 to be larger than the variations in the published fits. These values are not important as they mostly cancel in comparisons of different models. Δ*c* = 100 μeV kbar–2 and = 1 are chosen as described in Section 4.2.

 ML fitting is used because the Perlin data shows outliers among the residuals after LS fitting (Fig.S2). For this, the pdf of the residuals for these data is written as the sum of two Gaussian distributions, and the likelihood is written as

 (S6)

where *data*1 is the Perlin27 data (datapoints *i* = 1 to 88) and *data*2 is the Goñi26 data (datapoints *j* = 89 to 146). The three Gaussians have σG for the Goñi data, σP for the good Perlin datapoints, and σO for the outlier Perlin datapoints. The fraction *f* may be interpreted as the fraction of the Perlin data that are outliers.



**Fig.S2.** A histogram of the residuals returned by the least-square fit to the Perlin *et al.*27 data. The double Gaussian pdf used in the maximum-likelihood fitting of these data is plotted, multiplied by the number of data points (88).

 **Table S1.** Fits as described in the text to the data of Goñi *et al.*26 (Go) and Perlin *et al.*27 (Pe). The methods are least-squares (LS) and maximum likelihood (ML). In the 4th column, the values are *b* for the QP3 fit and Ξ/*B*0 for Eq.S5, in units of meV kbar–1; in the 5th column *c* for the QP3 fit has units μeV2 kbar–2. is dimensionless; for it, 4.49\* was fixed not fitted.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | Method | Data | *b*, Ξ/*B*0 | *c*,  | ln*L* | SBIC | lnMLI | lnBF | Evidence |
| QP3 | LS | Go | 10.9 ± 0.2 | –14 ± 1  | –247.18 | –255.34 | –258.34 |  0 |  |
| Eq.S5 | LS | Go | 11.6 ± 0.3 | 4.5 ± 0.3 | –247.17  | –255.33 | –254.55 | 3.8 | Strong |
| Eq.S5 | LS | Go | 11.6 ± 0.1 | 4.49\* | –247.17 | –253.28 | –254.40 | 3.9 | Strong |
| Eq.S5 | ML | Pe + Go | 11.42 ± 0.04 | 4.49\* | –417.6 | –428.1 | –429.9 | 0 |  |
| Eq.S5 | ML | Pe & Go | 11.28 ± 0.0611.60± 0.04 | 4.49\* | –408.8 | –421.6 | –425.3 | 4.6 | Strong |

**§8. Example 3. Nanotube Raman fit.**

For the Raman data, Fig3, the ranges for the 23 parameters common to the three models were Δ*Pi* = 120 cm–1 for the Raman peak positions as they must fall in the range of the data, 238 to 364 cm–1. For their widths, Δ*Wi* = 10 cm–1, on physical grounds (we know they are nanotube RBM modes). For the areas, Δ*Ai* = 500 cm–1, about twice the largest peaks present. The zero offset (dark count rate) is expected to be of the order of 1. The fraction of Gaussian in the Voigt lineshape, *G*, is necessarily between 0 and 1, so Δ*G* = 1. In the Fourier and Poly models, the extra parameters were given ranges Δ*ci* = Δ*si* = Δ*a*i =200, except for *c*0 and *a*0 which represent the zero offset (dark count rate), expected to be of the order of 1, and so Δ*c*0 = Δ*a*0 = 1. For the extra peaks in the Peaks model, Δ*Pi* = 120 cm–1 again, and here and in the Tails model, Δ*Wi* = 100 cm–1 as these can be much broader than the sharp peaks. The intensities can be correspondingly large, for Peaks we used Δ*Ai* = 1000 cm–1. For the Tails model, we gave the tail intensity for peak *i* as *f* × *Ai* and the fitting parameter *f* was given the range Δ*f* = 5.

While we tried adding broad Lorentzian tails to each peak, obtaining at best ln*L* = –80, the function

 (S7)

gave much better fits with the best ln*L* = –50. This function has symmetrical exponential tails with a little rounding at *x* = 0 to make it continuously differentiable.

 

**Fig.S3.** The normalised likelihood function *L*(*G*2) (solid line) and the pdf of this parameter (dashed line) are plotted.

Finally, Fig.S3 shows a case earlier in the fitting, where the calculated Bayes Factor is in error but the error is readily detected and corrected. We gave each of the seven peaks its own fraction *Gi* of Gaussian content in the pseudo-Voigt peak shape. Fitted values were sometimes outside the meaningful range of 0 ≤ *Gi* ≤ 1, so this condition was added as a constraint. Fig.3c shows that in this case, with a fitted value of *G*2 = 0, half of the Gaussian function *L*(*G*2) is outside the range of non-zero probability for this parameter. Its contribution to the MLI integral is thus twice what it should be. This can be corrected by dividing the MLI by 2, or subtracting ln2 from lnMLI and lnBF.