9.8 Baselines

Working through this example is intended to stimulate thought about three key issues in Bayesian model-fitting. What are the priors on the model parameters? Is our model sufficiently comprehensive, or should we be testing it against alternatives? How do we do the marginalization integrations to get the error bounds on the parameters of interest?
Starting with a spectral line superimposed on some sort of continuum, it is easy enough to model the baseline and take it away; once we are left with the line, we can fit a Gaussian (or something else) and get error bounds on its parameters by a version of the procedure outlined in Example 6.6. Instinctively we feel that this is missing a key source of error; have we chosen the right form for the baseline? If we expand the baseline model, inevitably we get more parameters. Should we include these? What are the odds on a model with one more parameter? The Bayes factor helps here, but ... we are going to integrate over priors to do this. How do we choose them for some more or less complex model of a baseline? The original analysis of scale and location parameters helps here. If a term in a polynomial baseline model is

$$\left(\frac{x - x_0}{\alpha}\right)^2,$$

where x is position in the spectrum, then α is a scale parameter (depends on the units we use, but the prior should be invariant under change of units). By contrast, x_0 is a location parameter. So we could try priors of the form $\mathrm{prob}(\alpha) \propto 1/|\alpha|$ and $\mathrm{prob}(x_0) = \text{constant}$.
In the case of removing a baseline by Fourier techniques, the prior on the amplitudes is not so apparent, because they can be positive or negative; experimental design should be a guide to the range. We might argue that the baseline is "noise", and since the transform of noise of standard deviation σ is itself of amplitude proportional to σ, we might assign a Jeffreys-like prior $1/|\beta|$ to the amplitudes β.
Hidden away in the Fourier baseline technique is the choice of the number of Fourier components to fit. This is a special case of the problem of sufficiently comprehensive models, but illustrates the Bayesian techniques nicely. Find the odds on a model with N components compared to one with $N + 1$ – until the odds are even, I suggest you should carry on adding components. In a layer below this reasoning, I am assuming that we create our baseline from successively higher frequencies in the Fourier transform. Why? Maybe a bandpass filter would be more appropriate – but over what range of frequencies? What about a polynomial? At some level, information about what *causes* the baseline must inform our choices.
Both marginalization and the calculation of odds will require multi-dimensional integrations. There are only two games in town – use the Gaussian approximation (see Example 6.6) to the likelihood. Then only simple – wrong? – priors are allowed, to keep things Gaussian. Otherwise, if the asymptotic nature of the Gaussian worries you, then the powerful numerical techniques of Markov-chain-Monte-Carlo are what you need. At this stage, you also need the bibliography!

Now for the example. The data are a spectrum 256 points long; the noise level is 8 units (standard deviation) and the noise is Gaussian.
The assumed baseline is

$$a + b \sin 2\pi x/256 + b \sin 4\pi x/256 + c \cos 2\pi x/256 + d \cos 4\pi x/256$$

and the line is just a Gaussian with free parameters height, width and position. The first step is to fit the baseline; the region $100 - 160$ was "patched" out of the fit. The resulting fit is seen in the top left panel of the figure.
Now, fit the Gaussian to the line. To make the analysis easier, use the asymptotic Gaussian form of the likelihood, so that the problem is characterized by the covariance matrix. Often a least-squares minimization routine will give you the Hessian matrix at the minimum, which closely related to the covariance matrix (check your implementation details for definitions of stray factors of 2). For this fit, I find the covariance matrix to be (columns ordered as height, position and width):

$$\begin{bmatrix} 5.20 & 0 & -1.48 \\ 0 & 1.26 & 0 \\ -1.48 & 0 & 1.26 \end{bmatrix}$$

The fit to the baseline-subtracted data is in the top right panel of the Figure.
Let's focus on the height; the distribution of height, location and width is asymptotically Gaussian, defined by the covariance matrix. If we assume diffuse priors on these, we can marginalize out location and width and get a distribution for just the height. See Exercise 6.9 for the reference for this. The result for the height is 24.6 ± 2.3.
If we now fit the baseline and the Gaussian *simultaneously* (no patching) the covariance matrix is much bigger, but the fit looks very similar – bottom left panel in the figure. Now, however, the correlations between the baseline parameters and the Gaussian's parameters are kept in the analysis. The marginalization, although apparently lengthier, is still easy in the Gaussian approximation, and the result for the height is 23.7 ± 3.1 – somewhat bigger than for the "two part" analysis. However it is clear that the crude patching technique is pretty good.
Finally, out of curiosity let us add two more terms to the baseline, namely

$$e \sin 6\pi x/256 + f \cos 6\pi x/256.$$

Cranking through the same analysis as before (see the bottom right panel of the figure), we find a rather different value for the height: 26.4 ± 5.6. The error is somewhat bigger but should we take the extra terms seriously? Simply integrating the posterior distributions (the likelihood functions in our case) gives us the Bayes factor or evidence; the ratio of the Bayes factors is about 7 to 1 in favor of adding the two extra terms (if the prior odds are even). These are fairly weak odds,but do suggest we should add these extra terms. The covariance matrix is actually quite similar to the one obtained in the original 'patching' approach, although scaled up with bigger errors in the parameters.

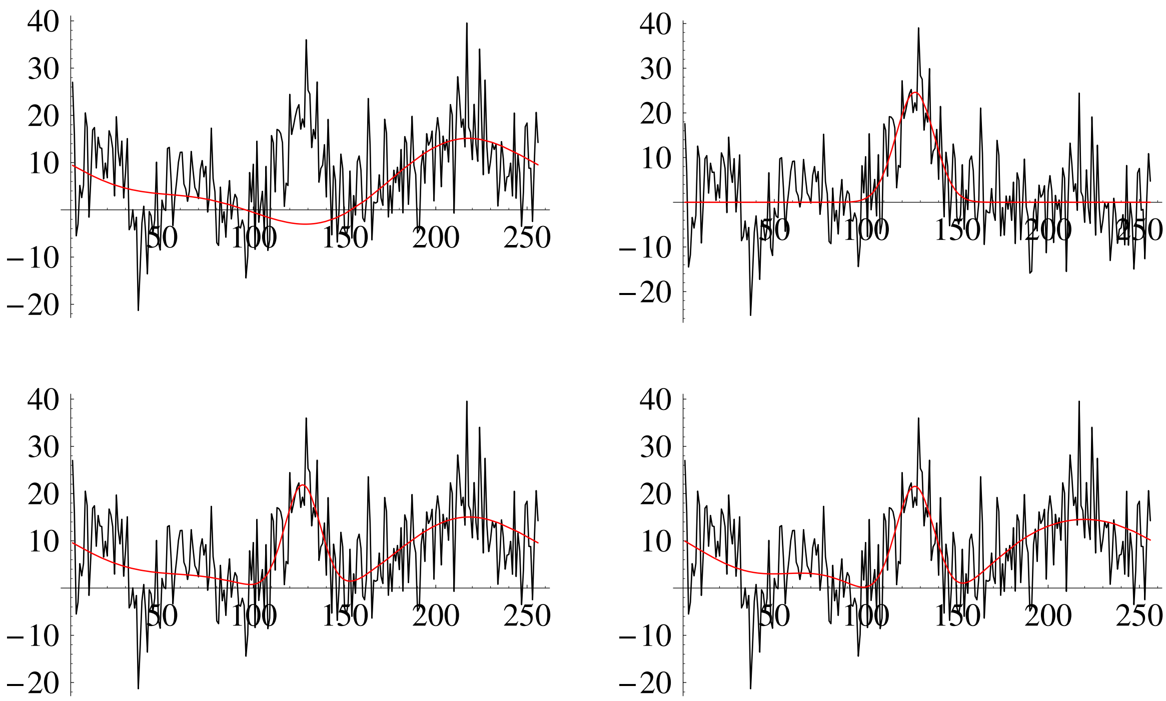

Figure 1: Various fits to the spectrum, as explained in the text.