Line Broadening

Spectral lines are not arbitrarily sharp. There are a variety of mechanisms that give them finite width, and some of those mechanisms contain significant information. We'll consider a few of these in turn, then have a detailed discussion about how line shapes and profiles have given information about rapidly rotating accretion disks around black holes.

First, **Ask class:** why can't a spectral line be arbitrarily sharp? Ultimately, it comes from the uncertainty principle in the form $\Delta E \Delta t > \hbar/2$. If a line were arbitrarily sharp, this would imply perfect knowledge of E, which can't happen unless the atom spends an infinite amount of time before decaying into a lower state. If instead the decay time is finite, say $\tau_{\rm decay}$, then the approximate width of the line is $\Delta E \sim \hbar/\tau_{\rm decay}$. This is called *natural broadening*, and represents the limit on how sharp a line can be. If one has an atom in state n, and the spontaneous decay rate to a lower energy state n' is $A_{nn'}$, then the spontaneous decay proceeds at a rate

$$\gamma = \sum_{n'} A_{nn'} \ , \tag{1}$$

Ask class: Is this the only contribution to the decay? No, there are also induced decay processes (stimulated emission). These should be added to the spontaneous rates. The energy decays at a rate $\exp(-\gamma t)$. The energy is proportional to the square of the coefficient of the wave function, so that coefficient decays at a rate $\exp(-\gamma t/2)$. The decaying sinusoid that is obtained for the electric field then gives a line profile of the Lorentz form, as we saw in the semiclassical picture:

$$\phi(\nu) = \frac{\gamma/4\pi^2}{(\nu-\nu_0)^2 + (\gamma/4\pi)^2} \ . \tag{2}$$

Ask class: from the above discussion, can they name a state that will have zero breadth because it can persist indefinitely? The ground state is stable, so its energy can be defined with (in principle) arbitrary sharpness. If instead the level n' is itself an excited level, that energy level has breadth as well. Then, approximately, the effective width of the transition is $\gamma = \gamma_u + \gamma_l$, where γ_u and γ_l are respectively the widths of the upper and lower states.

Ask class: what's a way to broaden this line further *for a single atom*? Collisions will do it. Effectively, a collision produces an abrupt change in the phase of the wave function. Suppose that collisions occur at random times with an average frequency $\nu_{\rm col}$. Then the resulting profile still looks like a Lorentzian:

$$\phi(\nu) = \frac{\Gamma/4\pi^2}{(\nu-\nu_0)^2 + (\Gamma/4\pi)^2} \ , \tag{3}$$

where now $\Gamma = \gamma + 2\nu_{\rm col}$ includes contributions from both natural broadening and collisions.

Now suppose we have a collection of many atoms, and we are measuring the line profile from all of them combined. **Ask class:** what is another mechanism that will broaden the observed line? Doppler shifts are one way. Each atom, individually, will emit a line that has the natural width plus a collisional width, but its motion towards us or away from us will produce blueshifts or redshifts, so its line center will be displaced. Many atoms, moving in different directions with different speeds, will produce a line blend with significant width. For example, suppose the atoms are thermalized and thus have a Maxwellian distribution of velocities with some temperature T. Then if the line center frequency is ν_0, the line profile is Gaussian:

$$\phi(\nu) = \frac{1}{\Delta\nu_D\sqrt{\pi}} e^{-(\nu-\nu_0)^2/(\Delta\nu_D)^2} \tag{4}$$

where $\Delta\nu_D$, the Doppler width, is

$$\Delta\nu_D = \frac{\nu_0}{c}\sqrt{\frac{2kT}{m_a}}\ . \tag{5}$$

A similar profile is obtained if one has microturbulence. However, if the Doppler shifts are from ordered motion (e.g., orbits), the profile will be different. Fundamentally, one calculates the Doppler profile by adding up the Doppler shifts from all the atoms individually. One can imagine a situation in which collisions and Doppler shifts are both important. If the Doppler shifts are due to isotropic thermal motion, the resulting line profile is called the Voigt profile, and is a convolution of a Lorentzian and a Gaussian. Note that because a Lorentzian dies off like a power law, whereas a Gaussian dies off exponentially, the line wings sufficiently far from the center will always be dominated by the Lorentzian.

Let's examine a couple of examples in which the line profile gives us physical information.

Suppose you are observing stars moving in the center of a distant galaxy. **Ask class:** If there is a supermassive black hole in the center of the galaxy then what, qualitatively, do you expect to see when you focus on a particular spectral line? It depends on whether the motion near the black hole is ordered or random. If the motion is ordered, then as one scanned across the central regions one would expect the net velocity (as measured by the redshift or blueshift of the line) to increase quickly towards the center, then abruptly change sign when the center was crossed. If the motion is random, then the line would have a width that increased towards the center. Either way, one can define a velocity or velocity dispersion that indicates the mass of the black hole. In more distant galaxies, other methods are used to estimate or constrain the mass of the black hole, because one can't observe the optical lines of stars with enough spatial resolution.

Now consider another example. The inner regions of accretion disks around black holes are hot places, and various processes mean that there are photons of energies reaching up to many keV to tens or even hundreds of keV. When a photon with an energy of 6-7 keV or more hits the accretion disk, it can photoionize the inner K shell electrons of iron,

which is relatively abundant for a metal and has a high cross section for this effect. When an electron drops down into the K shell from the next shell up, it emits a line that, in the rest frame of the atom, is relatively sharp and has an energy of 6.4 keV. Motion of the atoms in an accretion disk can change this sharp rest-frame line into a broader line. Detailed interpretation of this line has given a tremendous amount of information about the properties of accretion disks and strong gravity. Let's try our hand at it. Suppose that the observed line looks like:

Ask class: What effects might account for this profile? We'll need to identify important parts and interpret them separately to put together the picture. We see that the line is (1) broad, (2) asymmetric, (3) sharply peaked. We also note that the line goes a little bit above the rest-frame energy, but a lot below. All this can be understood in terms of a relativistic disk. Emission from deep in a gravitational well can be redshifted significantly, which is why the emission can get to half(!) of the rest-frame energy. If matter is moving in a disk, then when the fast (half the speed of light or more) motion is towards us, then there is significant beaming. Put another way, since the specific intensity scales as $I_\nu \propto \nu^3$, blueshifts can increase the intensity a lot, whereas redshifts decrease it a lot. Thus, the emission is strongly peaked where there is a slight blueshift, and it is asymmetric. The sharp cutoff is also understandable, given some details of a relativistic disk model; deep in the gravitational well, orbital motion produces little if any net blueshift. These results, worked on by Chris Reynolds and Andy Young at Maryland (among others) have provided evidence for rapid rotation of supermassive black holes, and possibly of extraction of the spin energy of black holes in a couple of cases!

Now, since we didn't quite finish the statistics lecture, we're going to shift gears and talk about model comparison. To recall, we talked about parameter estimation last time. But, one cautionary point: since the *value* of the likelihood never enters, one can happily calculate maximum likelihoods and credible regions for models that are awful! It's an automatic procedure. That's why Bayesians draw a distinction between parameter estimation and model comparison, which we will now treat.

Suppose we have a data set, and two models to compare. How do we determine which model is favored by the data? At first glance this may seem easy: just figure out which model matches the data better. But think about models with different numbers of parameters; intuitively, we should give the benefit of the doubt to the model with fewer parameters, based on Ockham's principle. In addition, one could imagine a situation in which the parameters of two models are qualitatively different. For example, some of the parameters could be continuous (e.g., temperature), and some could be discrete (e.g., the quantum spin of a particle). How are these to be taken into account?

This, in my opinion, is where Bayesian statistics shines. It provides a simple procedure that *automatically* takes into account different numbers of parameters in an intuitively satisfying way. As before we'll give the general principles, then try some examples.

Say we have two models, 1 and 2. Model 1 has parameters $a_1,\ a_2, \ldots,\ a_n$, and a prior probability distribution $P_1(a_1,\ a_2, \ldots,\ a_n)$. Model 2 has parameters $b_1,\ b_2, \ldots,\ b_m$ and a prior probability distribution $P_2(b_1,\ b_2, \ldots,\ b_m)$. For a given set of values $a_1,\ a_2, \ldots,\ a_n$, let the likelihood of the data given the model (defined above) for model 1 be $\mathcal{L}_1(a_1,\ a_2, \ldots,\ a_n)$, and similarly for model 2. Then the "odds ratio" of model 1 in favor of model 2 is

$$\mathcal{O}_{12} = \frac{\int \mathcal{L}_1(a_1,\ a_2, \ldots,\ a_n) P_1(a_1,\ a_2, \ldots,\ a_n) da_1\, da_2 \ldots da_n}{\int \mathcal{L}_2(b_1,\ b_2, \ldots,\ b_m) P_2(b_1,\ b_2, \ldots,\ b_m) db_1\, db_2 \ldots db_m} \tag{6}$$

where the integration in each case is over the entire model parameter space. Therefore, it's just a ratio of the integrals of the likelihoods times the priors for each model.

What does this mean? Don't tell a real Bayesian I explained it this way, but consider the following. Suppose you and a friend place a series of bets. In each bet, one has two possible models. You compute the odds ratio as above, and get $\mathcal{O}_{12}$ in each case. Ultimately, it will be determined (by future data, say) which of the two models is correct (we're assuming these are the only two possible models). If your friend puts down \$1 on model 2 in each case, how much money should you place on model 1 in each bet so that you expect to break even after many bets? You put down $\$\mathcal{O}_{12}$. That is, it really does act like an odds ratio. The reason a hard-core Bayesian might get agitated about this analogy is that Bayesian statistics emphasizes considering only the data you have before you, rather than imagining an infinite space of data (as happens in more familiar frequentist statistics). Still, I think this is a good description.

Why does this automatically take simplicity into account? Think of it like this. If your data are informative, then for a given set of data it is likely that only a small portion of the parameter space will give a reasonably large likelihood. For example, if you are modeling the interstellar medium in some region, you might have temperature and density as parameters; with good enough data, only temperatures and densities close to the right ones will give significant $\mathcal{L}$. Now, think about the priors. For a complicated model with many parameters, the probability density is "spread out" over the many dimensions of parameter space. Thus, the probability density is comparatively small in the region where the likelihood is significant. If instead you have few parameters, the prior probability density is less spread out, so it's larger where the likelihood is significant and therefore the integral is larger.

If the parameters have discrete instead of continuous values, you do a sum instead of an integral but otherwise it's the same. Note that we have to use the full Poisson likelihood here. When we did parameter estimation we could cancel out lots of things, but here we have an integral or sum of likelihoods so we can't do the cancellation as easily. The products $\prod \exp(-m_i)$ and $\prod(1/d_i!)$ will be the same for every likelihood, so those can be cancelled, but one still has a sum of likelihoods and so taking the log doesn't help.

Let's try an example. Consider a six-sided die. We want to know the probabilities of each of the six faces. Model 1 is that the probability is the same (1/6) for each face. Model 2 is that the probability is proportional to the number on the face. Normalized, this means a probability of 1/21 for 1; 2/21 for 2; and so on. We roll the die ten times and get 5, 2, 6, 2, 2, 3, 4, 3, 1, 4. What is the odds ratio for the two models?

We're starting with an easy one, in which there are no parameters, so we don't even have to do an integral, just a likelihood ratio. For model 1 the normalized model expectations per bin are $m_1 = 10/6$, $m_2 = 10/6$, and so on. For model 2 we have $n_1 = 10/21$, $n_2 = 20/21$, $n_3 = 30/21$, and so on. Therefore,

$$\mathcal{L}_1 = \left(\frac{10}{6}\right)^1 \cdot \left(\frac{10}{6}\right)^3 \cdot \left(\frac{10}{6}\right)^2 \cdot \left(\frac{10}{6}\right)^2 \cdot \left(\frac{10}{6}\right)^1 \cdot \left(\frac{10}{6}\right)^1 = 165.4 \tag{7}$$

and

$$\mathcal{L}_2 = \left(\frac{10}{21}\right)^1 \cdot \left(\frac{20}{21}\right)^3 \cdot \left(\frac{30}{21}\right)^2 \cdot \left(\frac{40}{21}\right)^2 \cdot \left(\frac{50}{21}\right)^1 \cdot \left(\frac{60}{21}\right)^1 = 20.7\ . \tag{8}$$

Thus, from this data,

$$\mathcal{O}_{12} = \mathcal{L}_1/\mathcal{L}_2 = 7.98\ . \tag{9}$$

Model 1 is strongly favored.

Now try another example, with the same data. Model 1 is the same as before, but now model 2 has a parameter. In model 2, the probability of a 1 is $1 - p$, and the probability of a 2, 3, 4, 5, or 6 is $p/5$. Therefore, model 2 encompasses model 1, so by maximum likelihood

alone it will do better. But will it do enough better to be favored? Let's assume as a prior that p is equally probable from 0 through 1. The numerator is the same as before, but for the denominator we need to do an integral. For probability p and our given data, the Poisson likelihood of the data given the model is

$$\mathcal{L}_2(p) = [10(1-p)] \cdot (2p)^3 \cdot (2p)^2 \ldots = 10(1-p)(2p)^9 \; . \tag{10}$$

Therefore the denominator is

$$\int_0^1 5120(1-p)p^9 \, dp = 46.5 \tag{11}$$

and the odds ratio is

$$\mathcal{O}_{12} = 165.4/46.5 = 3.55 \; , \tag{12}$$

so the first model is still preferred. Note that the maximum likelihood for model 2 occurs for $p = 0.9$ and gives 198.4, so as expected the more complicated model has a higher *maximum* likelihood; it's just not enough to make up for the extra complication.

Model comparison in Bayesian statistics is always between two precisely defined models. There is no analogue to the idea of a null hypothesis. Hard-core Bayesians consider this to be a strength of the approach. For example, suppose that you try to define a null hypothesis and do a standard frequentist analysis, finding that the null hypothesis can be rejected at the 99% confidence level. Should you, in fact, reject the null hypothesis? Not necessarily, according to Bayesians. Unless you know the full space of possible hypotheses, it could be that there are 10,000 competing hypotheses and of those your null hypothesis did the best. For example, suppose I think that gamma-ray bursts should come from isotropically distributed positions in the sky; that's my null hypothesis. A hundred positions are measured, and they are all found to cluster within 1° of each other. Surely I can reject my null hypothesis? Well, if I compare it with another hypothesis that says that all bursts should come from within 1" of each other, my null hypothesis does much better!

I'm not happy with this line of argument. To me, the testing of a null hypothesis as it's done in frequentist statistics is important because it gives you a way to tell if your model is reasonably close or not. That is, a standard chi squared per degree of freedom can give you an idea of whether you need to work a lot harder to get a good model, or if you're nearly there. In my opinion, it's important to have that kind of information, but there is reasoned disagreement on this issue.