BEHR-Lite

Michael Nowak, MIT-CXC, Nov. 21, (Revision 3)

The goal of this document is to come up with a simpler, more easily described and programmed version of the "Bayesian Estimation of Hardness Ratios" (BEHR) code. Since our primary concern is the Level 3 catalog, we will treat the problem somewhat less generally than in the paper by Park et al. (2007, ApJ, 652). Specifically, we will not look at all definitions of the color (hardness ratio), nor will we consider mutliple types of prior probability distributions. For our purposes the requirements are that: a) the algorithm is valid for color definitions of the form $(A - B)/(A + B)$, b) is valid for 0 observed counts in one of the bands, c) allows effective area variations, etc., and backgrounds to be incorporated, and d) the algorithm is comprehensible to a typical astrophysicist, but reasonably accurately reproduces the results of Park et al. (2007).

1 No Background

The algorithm described in this section meets all of the above goals, with the exception of easily incorporating a significant background. In the next section, I describe the version with background incorporated.

Bayes Theorem states that $P(\theta|D, M) \propto P(D|\theta, M)P(M, \theta)$, i.e., the probability of the model parameters (θ) given the data (D) and the model (M), is proportional to the probability of the data given the model and the parameters times a prior ($P(M, \theta)$). If we consider a Poisson process, and we take a uniform prior probability for the value of the expected counts (over a given time interval), then Bayes theorem yields for the intrinsic (i.e., expected) counts in a channel, $A_{\rm int}$:

$$P(A_{\rm int}|A_{\rm obs})\, dA_{\rm int} \;=\; \frac{(\epsilon_A A_{\rm int})^{A_{\rm obs}} \exp(-\epsilon_A A_{\rm int})}{A_{\rm obs}!}\, d(\epsilon_A A_{\rm int}) \;\; . \qquad (1)$$

Although $A_{\rm int}$ is the expected, real-valued counts, $A_{\rm obs}$ is the observed integer-valued counts in the channel. ϵ_A is a scaling factor that incorporates any desire to normalize the intrinsic rate to another reference (via different effective areas, different integration times, etc.). One possible use for ϵ_A is to reference backside/frontside effective areas in bands to one another (eventhough realistically this factor should also include spectral shape). Note also that the above probability is properly normalized, and is also valid for 0 observed counts.

We'd now like to use this probability distribution to help us arrive at a probability distribution for the hardness ratio defined by:

$$\mathcal{H} \;\equiv\; \frac{A_{\rm int} - B_{\rm int}}{A_{\rm int} + B_{\rm int}} \;\; , \qquad (2)$$

given counts observed in two channels, $A_{\rm obs}$ and $B_{\rm obs}$. We can calculate this probability distribution by effecting a simple coordinate transformation. If we look at the $A_{\rm int}$-$B_{\rm int}$ plane, then lines of constant $\mathcal{H}$ are radial spokes ($\mathcal{H} = 1$ along the $A_{\rm int}$-axis, $\mathcal{H} = 0$ at 45°, and $\mathcal{H} = -1$ along the $B_{\rm int}$-axis), and curves of constant $\sqrt{A_{\rm int}^2 + B_{\rm int}^2}$, $\mathcal{R}$, are quarter-circles. Thus we have $A_{\rm int} = \mathcal{R}\cos\theta$, $B_{\rm int} = \mathcal{R}\sin\theta$, and $dA_{\rm int}dB_{\rm int} = \mathcal{R}d\mathcal{R}d\theta$. The hardness ratio is then simply defined by:

$$\mathcal{H} \;=\; \frac{\cos\theta - \sin\theta}{\cos\theta + \sin\theta} \quad , \quad d\mathcal{H} \;=\; \frac{-2}{(\cos\theta + \sin\theta)^2}\, d\theta \;\; . \qquad (3)$$

As the observed counts in each channel are independent Poisson variables, we treat the probability distributions for the intrinsic counts independently. Thus, we can write the joint probability density for $A_{\rm int}$ and $B_{\rm int}$ as:

$$P(A_{\rm int}, B_{\rm int}|A_{\rm obs}, B_{\rm obs})\, dA_{\rm int}dB_{\rm int} \;=\; F(A_{\rm int})F(B_{\rm int})\, \mathcal{R}d\mathcal{R}d\theta \;\; . \qquad (4)$$

Using eq. (1) above, we can write this as:

$$P(A_{\rm int}, B_{\rm int})dA_{\rm int}dB_{\rm int} = \frac{\epsilon_A^{A_{\rm obs}+1}\epsilon_B^{B_{\rm obs}+1}\,(\cos\theta)^{A_{\rm obs}}\,(\sin\theta)^{B_{\rm obs}}}{A_{\rm obs}!B_{\rm obs}!}\mathcal{R}^{A_{\rm obs}+B_{\rm obs}+1}\, e^{-\mathcal{R}(\epsilon_A\cos\theta + \epsilon_B\sin\theta)}d\mathcal{R}d\theta \qquad (5)$$

Marginalizing over $\mathcal{R}$, and using eq. (3) to replace $d\theta$, we obtain:

$$P(\mathcal{H})d\mathcal{H} = 2\left[\frac{(A_{\rm obs}+B_{\rm obs}+1)!}{A_{\rm obs}!B_{\rm obs}!}\right]\left(\epsilon_A^{A_{\rm obs}+1}\epsilon_B^{B_{\rm obs}+1}\right)\left[\frac{(\cos\theta)^{A_{\rm obs}}(\sin\theta)^{B_{\rm obs}}(\cos\theta+\sin\theta)^2}{(\epsilon_A\cos\theta+\epsilon_B\sin\theta)^{A_{\rm obs}+B_{\rm obs}+2}}\right]d\mathcal{H} \quad (6)$$

John Davis pointed out to me that this equation simplifies even further by explicitly substituting $\mathcal{H}$ back into the equation, to find:

$$P(\mathcal{H})\,d\mathcal{H} \;=\; 2\left[\frac{(A_{\rm obs}+B_{\rm obs}+1)!}{A_{\rm obs}!B_{\rm obs}!}\right]\left(\epsilon_A^{A_{\rm obs}+1}\epsilon_B^{B_{\rm obs}+1}\right)\frac{(1+\mathcal{H})^{A_{\rm obs}}(1-\mathcal{H})^{B_{\rm obs}}}{\left[\epsilon_A+\epsilon_B+(\epsilon_A-\epsilon_B)\,\mathcal{H}\right]^{A_{\rm obs}+B_{\rm obs}+2}}\,d\mathcal{H}\ . \quad (7)$$

This is very straightforward to program, and is even analytically integrable (although numerical integration might be more convenient). For the special case of $\epsilon_A = \epsilon_B = 1$, it's easy to show that this peaks at $\mathcal{H} = (A_{\rm obs} - B_{\rm obs})/(A_{\rm obs} + B_{\rm obs})$.

The only shortcomings in the above is that it does not yet include a background term. As is, it could be used for the (many) cases where the background is negligible (predominantly on-axis sources). It needs to be generalized for cases where the background becomes significant (off-axis cases, most likely).

2 Background

When considering the background, we want to begin by altering eq. (1) by setting:

$$\epsilon_A A_{\rm int} \;\rightarrow\; \epsilon_A A_{\rm int} + \beta_A\ , \quad (8)$$

where β_A is the intrinsic background rate for channel A, and $A_{\rm int}$ remains the intrinsic, background-free source counts for channel A. Performing a transformation of variables as above, we then have:

$$\begin{aligned}P(A,B)\,dAdB \;\propto\; & \frac{\epsilon_A\epsilon_B}{A_{\rm obs}!B_{\rm obs}!}\left(\epsilon_A\mathcal{R}\cos\theta+\beta_A\right)^{A_{\rm obs}}\left(\epsilon_B\mathcal{R}\sin\theta+\beta_B\right)^{B_{\rm obs}} \\ & e^{-[\mathcal{R}(\epsilon_A\cos\theta+\epsilon_B\sin\theta)+(\beta_A+\beta_B)]}\;P(\beta_A)P(\beta_B)\;\mathcal{R}d\mathcal{R}d\theta\ , \end{aligned} \quad (9)$$

or, upon expanding using binomial coefficients,

$$\begin{aligned}P(A,B)\,dAdB \;\propto\; & \left[\sum_{j=0}^{A_{\rm obs}}\frac{A_{\rm obs}!}{j!(A_{\rm obs}-j)!}(\epsilon_A\mathcal{R}\cos\theta)^{A_{\rm obs}-j}\beta_A^j\exp(-\beta_A)P(\beta_A)\right] \\ & \left[\sum_{k=0}^{B_{\rm obs}}\frac{B_{\rm obs}!}{k!(B_{\rm obs}-k)!}(\epsilon_B\mathcal{R}\sin\theta)^{B_{\rm obs}-k}\beta_B^k\exp(-\beta_B)P(\beta_B)\right] \\ & \frac{\epsilon_A\epsilon_B}{A_{\rm obs}!B_{\rm obs}!}\,e^{-\mathcal{R}(\epsilon_A\cos\theta+\epsilon_B\sin\theta)}\;\mathcal{R}d\mathcal{R}d\theta\ . \end{aligned} \quad (10)$$

Note that the above is not actually properly normalized. Also note that we have introduced probability functions, $P(\beta_A)$ and $P(\beta_B)$, for the background rates. Since these, too, will come from measurements of a Poisson variable, e.g., counts in some large, neighboring region, we choose:

$$P(\beta_A) \;=\; \gamma_A\frac{(\gamma_A\beta_A)^{\delta_A}\exp(-\gamma_A\beta_A)}{\delta_A!}\ , \quad (11)$$

where δ_A are the actual, measured (integer) counts in the background region and γ_A^{-1} is the scaling factor that takes us from the measured background counts to the expected background counts in the source region. In practice, we expect $\gamma_A > 1$, and usually hope that $\gamma_A \gg 1$. (For the latter case, we recover the formulae for the case of no background.)

We can now integrate over the various nuisance parameters, β_A, β_B, and as before, $\mathcal{R}$. And also as before, we can substitute $\mathcal{H}$ back into the probability equation. We then obtain:

$$P(\mathcal{H})\, d\mathcal{H} \;\propto\; 2\left(\epsilon_A^{A_{\rm obs}+1}\epsilon_B^{B_{\rm obs}+1}\right)\left(\sum_{j=0}^{A_{\rm obs}}\sum_{k=0}^{B_{\rm obs}}\frac{(\delta_A+j)!}{\delta_A!}\frac{(\delta_B+k)!}{\delta_B!}(\gamma_A+1)^{-j}(\gamma_B+1)^{-k}\right.$$
$$\left.\frac{(A_{\rm obs}+B_{\rm obs}+1-j-k)!}{(A_{\rm obs}-j)!(B_{\rm obs}-k)!j!k!}\frac{(1+\mathcal{H})^{A_{\rm obs}-j}(1-\mathcal{H})^{B_{\rm obs}-k}}{[\epsilon_A+\epsilon_B+(\epsilon_A-\epsilon_B)\mathcal{H}]^{A_{\rm obs}+B_{\rm obs}+2-j-k}}\right) d\mathcal{H} \qquad (12)$$

The above is not normalized; however, in the limit of γ_A, $\gamma_B \rightarrow \infty$, it does reduce to the properly normalized form of $P(\mathcal{H})$ found for the case of zero background.

There are a few possible strategies for calculating this probability function. For small numbers of counts in *both* channels A and B, one could try calculating each term in the above and summing over the $(A_{\rm obs}+1)\times(B_{\rm obs}+1)$ elements. The second strategy is to only retain small powers of $(\gamma_A+1)^{-j}(\gamma_B+1)^{-k}$. (I have not yet tested the latter scheme to see how well it works.)

3 Multiple Observations

After the probability function has been calculated, whether with or without background, for an individual observation, indexed by i, one can then calculate the joint probability function for a series of observations as:

$$P(\mathcal{H}_{\rm joint}) \;\propto\; \prod_i P(\mathcal{H}_i | A^i_{\rm obs}, B^i_{\rm obs}, \epsilon^i_A, \epsilon^i_B, \gamma^i_A, \gamma^i_B, \delta^i_A, \delta^i_B) \;\; . \qquad (13)$$

This form is manifestly symmetric with respect to the order of the individual observations, and it is not any more difficult to calculate than the individual probability functions.

4 Code & Figures

Attached below is really simple (but fairly dumbass) `S-lang` code that seems to work well for the above cases by producing normalized probability distributions for 10^3 values of $\mathcal{H}$. (This should be sufficient resolution on $\mathcal{H}$ to obtain 95% confidence intervals.) The case with no background is almost instantaneous. For the case of background included, with a 100×100 array, the code runs in about 5 seconds on my 3 year old laptop. I have faith that Davis can come up with a better version that will run 10 times faster. As proof of concept, I think this demonstrates that we can come up with a replacement for the BEHR code that will be far easier to maintain, and yields comparably good results. I have attached several figures of sample probability distributions.

```
require("gsl");

public variable h_lo, h_hi;
(h_lo,h_hi) = linear_grid(-1,1,1000);
public variable h = (h_lo+h_hi)/2.;
public variable dh = (h_hi-h_lo);

% No background case

public define ph_nb(a,b,ea,eb)
{
   variable ph = lngamma(a+b+2) + log(2.) - lngamma(a+1) - lngamma(b+1)
              + (a+1)*log(ea) + (b+1)*log(eb) + a*log(1+h)
              + b*log(1-h) - (a+b+2)*log(ea+eb+(ea-eb)*h);
```

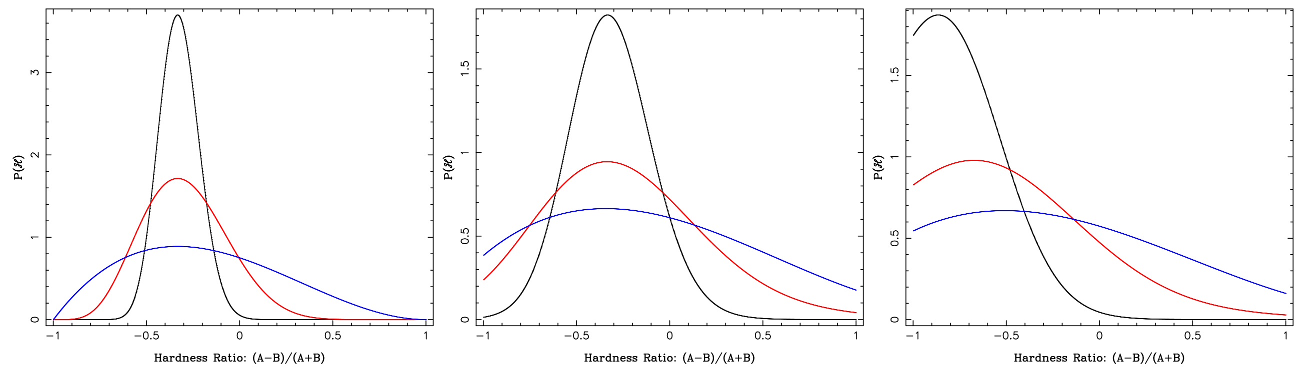

Figure 1: *Left:* Probability distributions without background, for $A = 1$, $B = 2$ (blue), $A = 5$, $B = 10$ (red), and $A = 24$, $B = 50$ (black). Compare this to Fig. 8 of Park et al. (2007). *Middle:* Probability distributions *with* background. The values of A and B are the same as on the left, however, for each $\gamma_A = \gamma_B = 10$, and $\delta_A = 5$, $\delta_B = 10$ (blue), $\delta_A = 25$, $\delta_B = 50$ (red), and $\delta_A = 125$, $\delta_B = 250$ (black). I.e., for each channel, background makes up about half of the observed counts. The peak of the color is therefore the same as on the left, but the distributions are wider. *Right:* Probability distributions *with* background. The values of A, B, δ_A, and δ_B are the same as for the middle, but $\gamma_A = 5$, and $\gamma_B = 10$ for each. I.e., background makes up nearly 100% of the observed counts for channel A, and about half of the observed counts for channel B. Hence the distributions are very broad and skewed towards $\mathcal{H} = -1$.
(For all figures, $\epsilon_A = \epsilon_B = 1$.)

```
   return exp(__tmp(ph));
}

% Case with background

public define ph_bkg(a,b,ea,eb,da,db,ga,gb)
{
   variable j,k,pjk;
   variable psum = 0;
   _for j (0,int(a),1)
   {
      _for k (0,int(b),1)
      {
         pjk = lngamma(da+j+1)+lngamma(db+k+1)-lngamma(da+1)-lngamma(db+1)
               -j*log(ga+1.)-k*log(gb+1)+lngamma(a+b+2-j-k)
               -lngamma(a-j+1)-lngamma(b-k+1)-lngamma(j+1)-lngamma(k+1)
               +(a-j)*log(1+h)+(b-k)*log(1-h)
               -(a+b+2-j-k)*log(ea+eb+(ea-eb)*h);
         psum += exp(__tmp(pjk));
      }
   }
   return psum/sum(psum*dh);
}
```