Accounting for Absorption Lines in High Energy Spectra

Christopher Hans and David A. van Dyk
Department of Statistics, Harvard University
hans@stat.harvard.edu, vandyk@stat.harvard.edu

Abstract

The increasing popularity of Markov chain Monte Carlo methods and the limitations of “classical” astrophysical data analysis in the face of a new class of instruments (e.g. the Chandra X-Ray Observatory) make Bayesian analysis of high-resolution low-count energy spectra both feasible and attractive. van Dyk et al. (ApJ, 548:224-243, 2001) describes a Bayesian hierarchical model which directly models counts as a Poisson process, avoiding problems resulting from the Gaussian assumptions of standard chi square fitting (see also the poster by Sourlas et al.). We extend this model to account for absorption lines. Estimating absorption line parameters (location, width and intensity) is challenging, as it is often difficult to distinguish between random fluctuations in the continuum and weak absorption lines. Additionally, lines cause the already low counts to become even smaller, making Gaussian assumptions even less tenable. We assume the same basic model as van Dyk et al. by modeling the counts as Poisson, and use a generalized linear model with a log-log link function to incorporate absorption lines. This requires an additional Metropolis-Hastings step in the Gibbs sampler. Incorporating our model into the larger model is trivial given its hierarchical structure and the construction of the Gibbs sampler. We finish by investigating the properties of our model using simulated data.
Absorption Lines

Absorption lines are downward spikes in the continuous spectrum, i.e. the continuum, of energy that is emitted by a source, and represent wavelenghts where photons from the continuum have been absorbed by elements in the source. We are interested in estimating three parameters of the absorption process: $\mu$ (the location of the line), $\sigma^2$ (the width parameter) and $\tilde{\lambda}$ (the intensity parameter). A simulated spectrum with a strong absorption line is shown below.

![Observed Counts](image1.png)  ![Expected Counts](image2.png)

**Figure 1:** The plot on the left shows a simulated spectrum. The continuum model is a power law, and the absorption line parameters are $\mu = 1.25$, $\sigma^2 = 0.0005$ and $\tilde{\lambda} = 15$. The plot on the right shows the expected number of counts as a function of energy.
A Model for Absorption Lines

Statistical Model: In the absence of an absorption line, we model the true number of counts at energy $E_j$ as independent observations from a Poisson distribution,

$$Y_j \sim \text{Poisson}(f(E_j, \theta)),$$

where $f(E_j, \theta)$ is the expected number of counts at energy $E_j$ from a continuum model with parameters $\theta$. To account for an absorption line we use a multiplicative model to describe the observed counts,

$$Y_j^{\text{obs}} \sim \text{Poisson}(f(E_j, \theta)\pi(E_j, \phi)),$$ \hspace{1cm} (1)

where $\pi(E_j, \phi)$ is the probability that a photon of energy $E_j$ is not absorbed by a line with parameters $\phi = (\bar{\lambda}, \mu, \sigma^2)^T$. We allow $f(E_j, \theta)$ to represent any continuum model but restrict $\pi(E_j, \phi)$ to be the double exponential absorption line model used by Freeman et al. (ApJ, 524:753-771, 1999)

$$\pi(E_j, \phi) = \exp \left\{ -\bar{\lambda} \exp \left\{ \frac{-(E_j - \mu)^2}{2\sigma^2} \right\} \right\}.$$

(a) \hspace{1cm} (b)

![Graphs showing the flexibility of $\pi(E_j, \phi)$]

Figure 2: These plots show the flexibility of $\pi(E_j, \phi)$. Plot (a) has $\bar{\lambda} = 1.5$ and plot (b) has $\bar{\lambda} = 85$. Both have $\mu = 1.25$ and $\sigma^2 = 0.0005$. 


Bayesian Prior and Posterior Distributions

Prior Specifications: In the current analysis we use flat, non-informative, prior distributions on the absorption line parameters. However, it is straightforward to include prior scientific information for these parameters. (In particular we use an inverse-$\chi^2$ prior distribution for $\sigma^2$, a Gaussian prior distribution for $\mu$ and a gamma prior distribution for $\lambda$, and assume the parameters are a priori independent.)

Posterior Analysis: We report parameter estimates and error bars that are summaries of the posterior distribution

$$p(\phi|Y^{obs}, \theta) \propto p(Y^{obs}|\phi, \theta)p(\phi),$$

which is obtained through Bayes’ Theorem. $p(Y^{obs}|\phi, \theta)$ is the Poisson likelihood of the observed data. The posterior distribution is not a standard distribution, so we use Markov chain Monte Carlo (MCMC) methods such as the Gibbs sampler and the Metropolis-Hastings algorithm to obtain samples from this distribution. Note that in this analysis the continuum parameters, $\theta$, are fixed. It is relatively easy, however, to incorporate this model into a model such as the hierarchical model of van Dyk et al., which could then estimate all model parameters simultaneously.
**Analysis of Simulated Data**

**Model:** We use the simulated data set pictured in Figure 1. The data were simulated according to Equation 1 with a power law continuum and absorption line parameters $\mu = 1.25$, $\sigma^2 = 0.0005$ and $\tilde{\lambda} = 15$.

**Posterior Modes:** Because random fluctuations in the continuum can often appear to be weak absorption lines, it is likely that the posterior will be multimodal. For example, using the simulated data, fixing $\sigma^2 = 0.0005$ and using a flat prior on $\mu$ and $\tilde{\lambda}$, ten different posterior modes were found using a handful of starting values. The mode with the highest likelihood value corresponds to the line in the simulated data; the nine other modes all have $\tilde{\lambda} < 0.25$, corresponding to possible weak lines. Because an exhaustive search of the parameter space was not performed, it is likely that there are many more of these small modes.

**Figure 3:** The plot on the left shows convergence to the largest mode from several starting values. Starting from values of $\mu$ outside of $[1.16, 1.45]$ generally resulted in convergence to a smaller mode. The second plot shows some of the smaller modes, labeled in order of decreasing likelihood value. $\sigma^2$ was fixed at 0.0005.
**MCMC Simulation:** We use the Gibbs sampler to draw from the posterior distribution of the parameters, $p(\sigma^2, \mu, \lambda | Y, \theta)$. Three chains of 1500 draws each were constructed from different starting values, and the latter 1000 draws from each chain were combined to form a sample from the posterior; flat prior distributions were used on all three parameters. Posterior summaries are shown in Table 1, plots and histograms of the draws are shown in Figure 4, and a scatterplot matrix of the three parameters appears in Figure 5.

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>median</th>
<th>mode</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$</td>
<td>5.266e-04</td>
<td>5.23e-04</td>
<td>5.36e-04</td>
<td>[4.33e-04, 6.34e-04]</td>
</tr>
<tr>
<td>$\mu$</td>
<td>1.2498</td>
<td>1.2498</td>
<td>1.2498</td>
<td>[1.2481, 1.2515]</td>
</tr>
</tbody>
</table>

Table 1: Posterior summaries for each of the parameters. All of the 95% credible intervals contain the model’s parameter values.

**Conclusions and Further Work**

As shown above, even without any prior information this model is able to detect weak absorption lines in the continuum. However, distinguishing weak lines from random fluctuations in the continuum is difficult and will require the use of statistics such as posterior predictive p-values (ppp-values). In the case of a strong line, the case where there is strong prior information (from past observations or known locations), or in the case when a parameter can be fixed at a specified value, posterior modes can be easily found and the Gibbs sampler converges relatively quickly to the posterior distribution of interest. The next step with this model is to integrate it into the larger hierarchical model of van Dyk et. al. and begin analyzing real data.
3000 Draws of log(intensity)

3000 Draws of mu

3000 Draws of log(sig2)

Figure 4: Posterior draws for the three parameters, with $\sigma^2$ and $\tilde{\lambda}$ on the log scale as this produces a roughly symmetric distribution.
Figure 5: Scatterplot matrix for the three parameters, with $\sigma^2$ and $\bar{\lambda}$ on the log scale. The width and intensity parameters are highly correlated, with narrower lines corresponding to more intense lines.
Computation Methods

Data Augmentation: For computational reasons we formalize the absorption line model in terms of observed and “missing” data. In particular, we let $Y^{\text{aug}} = \{Y^{\text{obs}}, Y^{\text{mis}}\}$ be the augmented data set, where $Y^{\text{obs}}$ are the observed data and $Y^{\text{mis}}$ represent the photons absorbed at the line. We also define $Y^{\text{com}}_j = Y^{\text{obs}}_j + Y^{\text{mis}}_j$ to be the “complete” data at energy $E_j$, the count that would have been observed had there been no absorption line. Computational methods use this formulation via two-step algorithms:

**Step 1:** Update the missing data given the parameters and observed data,

**Step 2:** Update the model parameters given the missing and observed data.

Fitting the Absorption Line Parameters: Given $Y^{\text{com}}$, maximum likelihood estimation of $\phi$ can be simplified by noting that under a log-log link function, $\pi(E_j, \phi)$ is linear in $E_j$ and $E_j^2$:

$$- \log \left( - \log (\pi(E_j, \phi)) \right) = \left[ - \log \hat{\lambda} + \frac{\mu}{2\sigma^2} \right] + \left[ - \frac{\mu}{\sigma^2} \right] E_j + \left[ \frac{1}{2\sigma^2} \right] E_j^2.$$

We can identify the GLM coefficient $\beta = (\beta_0, \beta_1, \beta_2)^T$ with the expressions in the square brackets above and then find maximum likelihood estimates of

$$\hat{\sigma}^2 = \frac{1}{2\hat{\beta}_2}, \quad \hat{\mu} = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}, \quad \hat{\lambda} = \exp \left\{ \frac{\hat{\beta}_1^2}{4\hat{\beta}_2 - \log \hat{\beta}_0} \right\},$$

using the computational method of Iterative Reweighted Least Squares (IRLS). We can also choose to fix either $\sigma^2$ or both $\sigma^2$ and $\mu$, which changes the model to a GLM with an offset.
MCMC and EM Algorithms

**MCMC Methods:** We use a Gibbs sampler with a Metropolis-Hastings step to obtain draws from the posterior distribution of the absorption line parameters. In Step 1 of the sampler we draw the missing data given the current value of the parameters, $\phi^{[t]}$. Given the probability of absorption and observed counts the distribution of the absorbed counts is readily available, Poisson with expected value $f(E_j, \theta)(1 - \pi(E_j, \phi))$. In Step 2 we draw $\phi^{[t+1]}$ conditioned on the current draw of the missing data. This conditional distribution is not easily sampled from, so it is approximated using the maximum likelihood method described above. This approximation is then corrected using the Metropolis-Hastings algorithm.

**MCMC Sampler:**

**Step 1:** Draw $(Y_{\text{mis}}^{\text{mis}})^{[t]}|Y_{\text{obs}}, \phi^{[t]}, \theta \sim \text{Poisson}(f(E_j, \theta)(1 - \pi(E_j, \phi)))$,

**Step 2a:** Compute $\hat{\beta}|Y_{\text{aug}}$ and $\hat{\Sigma}_{\hat{\beta}}|Y_{\text{aug}}$ via the GLM,

**Step 2b:** Draw $\beta^{[t+1]}$ using the Metropolis-Hastings algorithm with $t_4(\hat{\beta}, \hat{\Sigma})$ jumping distribution, and transform $\beta^{[t+1]}$ to $\phi^{[t+1]}$,

where $\hat{\Sigma}_{\hat{\beta}}$ is the estimated covariance matrix of the GLM parameters, and $t_4$ is the $t$-distribution with 4 degrees of freedom.

**EM Algorithm:** The EM algorithm is a computational method for computing maximum likelihood estimates and is also based on data augmentation. When using MCMC methods to sample from a posterior distribution, it is often useful to first use a mode finding algorithm to gain some knowledge of the distribution, and we use the EM algorithm for this purpose. The details of the algorithm are similar to those of the MCMC sampler.