WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS

Editors: *David J. Balding, Peter Bloomfield, Noel A. C. Cressie, Nicholas I. Fisher, Iain M. Johnstone, J. B. Kadane, Geert Molenberghs, Louise M. Ryan, David W. Scott, Adrian F. M. Smith, Jozef L. Teugels;*
Editors Emeriti: *Vic Barnett, J. Stuart Hunter, David G. Kendall*

A complete list of the titles in this series appears at the end of this volume.

# Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives

## An essential journey with Donald Rubin's statistical family

Edited by

**Andrew Gelman**
*Department of Statistics, Columbia University, USA*

**Xiao-Li Meng**
*Department of Statistics, Harvard University, USA*

John Wiley & Sons, Ltd

# 26

# Efficient EM-type algorithms for fitting spectral lines in high-energy astrophysics

**David A. van Dyk and Taeyoung Park**[1]

## 26.1 Application-specific statistical methods

In recent years, a progressive new trend has been growing in applied statistics: It is becoming ever more popular to build application-specific models that are designed to account for the hierarchical and latent structures inherent in any particular data generation mechanism. Such multilevel models have long been advocated on theoretical grounds, but the development of methodological and computational tools for statistical analysis has now begun to bring such model fitting into routine practice. In this chapter, we discuss one such application, the use of highly structured models to analyze spectral and spatial data obtained with modern high-resolution telescopes that are designed to study the high-energy end of the electromagnetic spectrum (e.g., X-rays and Gamma-rays). In particular, we consider the high-resolution data that is available from the space-based *Chandra X-ray Observatory*.

[1]Department of Statistics, University of California, Irvine and Department of Statistics, Harvard University, Cambridge, Mass. The authors gratefully acknowledge funding for this project partially provided by NSF grant DMS-01-04129 and by NASA Contract NAS8-39073 (Chandra X-ray Center). This chapter summarizes one thread of the work of the California–Harvard Astrostatistics Collaboration (`www.ics.uci.edu/~dvd/astrostat.html`). In addition to the authors, active participants include A. Connors, D. Esch, P. Freeman, H. Kang, V. L. Kashyap, X. L. Meng, A. Siemiginowska, E. Sourlas, Y. Yu, and A. Zezas.

*Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives.*
Edited by A. Gelman and X-L. Meng  ISBN: 0-470-09043-X

Launched in 1999 by the space shuttle *Columbia*, *Chandra* provides a new class of high precision instrumentation that allows for much more precise imaging of distant X-ray sources. The first author has been working on developing methods for handling this data since before *Chandra* was launched; there are a number of citations listed below that fill in many details of what is presented here and discuss related topics in astrostatistics.

X-ray telescopes such as *Chandra* can map nearby stars with active magnetic fields, the remnants of exploding stars, areas of star formation, regions near the event horizon of a black hole, very distant but very turbulent galaxies, or even the glowing gas embedding a cosmic cluster of galaxies. The production of X-ray emission requires temperatures of millions of degrees and indicates the release of stored energy such as that in very strong magnetic fields, extreme gravity, explosive nuclear forces, or shock waves in hot plasma. Thus, X-ray observations give astrophysicists a window into the physical processes involved in these turbulent regions of the universe, which is unavailable from observations of visible light. Because of the complexity of the sources themselves as well as the data collection process, however, unlocking this window requires sophisticated statistical modeling and analysis. For example, the recorded X-rays are a mixture of X-rays from a number of physical processes within the source. The X-rays are also subject to the so-called effective area, a nonignorable stochastic censoring process: The probability that an X-ray is observed depends on its energy, one of the variables of primary interest. The energy and the originating sky coordinates of each X-ray are observed with error and X-ray observations are subject to background contamination. (More background on the relevant astrophysics and instrumentation appears in van Dyk et al. (2004)). To handle these various factors we generally adopt a Bayesian perspective and construct highly structured multileveled models. Sophisticated computational tools such as EM-type algorithms and MCMC samplers are required for model fitting.

In this chapter, we describe a particular applied question that has come up in our work in astrophysics. Namely, we describe computational methods for fitting narrow emission lines in high-energy spectral analysis. Spectral analysis aims to describe the distribution of the energy of photons emitted from a particular source; here we focus on a high-energy interval of energies, the X-ray band. An emission line is a narrow range of energy with excess electromagnetic emission, relative to nearby intervals of energy. Such emission lines appear as sharp jumps in the distribution over a narrow range of energies. Emission lines are formed when electrons of energized ions fall down to lower energy shells and the excess energy is emitted in the form of a photon. Because of the distinct quantum differences between the energies of the electron shells of a particular ion, photons are emitted with one of a number of particular energies. Thus, we observe excess electromagnetic emission at these energies. The emission lines can be used to identify the ions and thus the composition of the source. The redshift of the emission lines can be used to compute the relative velocity of and the distance to the source. For these reasons, the precise fitting of emission lines is of key interest to astrophysicists.



This chapter is organized into four sections. In Section 26.2, we describe the statistical issues involved with the *Chandra's* data-generating mechanism and introduce a highly structured model that accounts for this mechanism. Our model is formulated in terms of several levels of missing data, which are critical in our formulation of the necessary computational techniques. The specific problem that we address in this chapter is introduced in Section 26.3, where we discuss the computational challenges that are involved with fitting narrow emission lines in high-energy spectral analysis. In Section 26.4, we discuss model checking techniques based on the posterior-predictive distribution.

## 26.2 The Chandra X-ray observatory

Data are collected on each X-ray photon that arrives at one of the detectors on board *Chandra*; the time of arrival, the two dimensional sky coordinates, and the energy are all recorded. Because of instrumental constraints, each of these four variables is discrete. Thus, the data can be complied into a four-way table of photon counts with margins corresponding to time, energy, and the two sky coordinates. Spectral analysis focuses on the one-way energy margin and image analysis models the two-way marginal table of sky coordinates. In this chapter, we focus on spectral analysis; see van Dyk and Hans (2002), Esch (2003), Esch, Connors, Karovska, and van Dyk (2004), and van Dyk et al. (2004) for discussion of image analysis of *Chandra* data.

A spectral model aims to describe the distribution of the energy of photons emitted from an astronomical source. This distribution can be formulated as a finite mixture model, in which the photon count in each energy bin is modeled as the sum of several independent Poisson random variables. A simplified form of this model might consist of a continuum term and an emission line. These terms represent two physical processes in the source; the continuum emission is a smooth function across a wide range of energies, while the emission line is highly focused at a particular energy. Thus, we might model the expected Poisson count in energy bin $j$ as

$$\Lambda_j(\theta) = \Delta_j f\big(\theta^C, E_j\big) + \lambda p_j(\mu, \sigma), \quad \text{for } j = 1, \ldots, J \tag{26.1}$$

where $\Delta_j$ is the width of bin $j$, $f(\theta^C, E_j)$ is the expected counts per unit energy due to the continuum term at energy $E_j$, $\theta^C$ is the set of free parameters in the continuum model, $\lambda$ is the expected counts due to the emission line, and $p_j(\mu, \sigma)$ is the proportion of an emission line centered at $\mu$ and with width $\sigma$ that falls into bin $j$. A Gaussian or Lorentzian density function is often used to model the emission line, in which case $\sigma$ might represent the standard deviation or some other measure of variability. There are a number of standard forms for the continuum term; here we use a power law, $f(\theta^C, E_j) = \alpha E_j^{-\beta}$, with $\theta^C = (\alpha, \beta)$.

While the model in (26.1) is of primary scientific interest, a more complex model is needed to address the data distortion introduced by instrumental effects

and other aspects of the data collection procedure. For example, photons have a certain probability of being *absorbed* by interstellar or intergalactic media. Since this probability depends on the energy of the photon, the missing-data mechanism is nonignorable (Rubin, 1976a). A similar effect occurs inside the detector; rather than being reflected onto the detector, some photons are reflected away from or pass right through the mirror. The likelihood of this occurring again depends on the energy of the photon; this effect is known as the *effective area* of the detector. Even for photons that are recorded, their energy may be recorded with error; given the energy of the photon, there is a multinomial distribution that characterizes the likely energy that is recorded by the instrument. (In practice, the number of cells in these multinomial distributions is different from $J$; we index the cells that correspond to the observed data with $l = 1, \ldots, L$.) To account for these processes along with *background contamination*, (26.1) is modified via

$$\Xi_l(\theta) = \sum_{j=1}^{J} M_{lj} \Lambda_j(\theta) d_j u(\theta^A, E_j) + \lambda_l^B \tag{26.2}$$

where $M_{lj}$ is the probability that a photon with true energy in bin $j$ is recorded in the multinomial cell $l$, $d_j$ is the effective area of bin $j$, $u(\theta^A, E_j)$ is the probability that a photon with energy $E_j$ is *not* absorbed, and $\lambda_l^B$ is a Poisson intensity of the background counts in channel $l$. The multinomial distributions and effective area are presumed known from calibration. The absorption probability is parameterized using a smooth function, see van Dyk and Hans (2002) for details. Background contamination is quantified using a second observation from an area of black space near the source of interest, where all counts are assumed to be due to background contamination. More details, more general forms, and applications of this model can be found in van Dyk, Connors, Kashyap, and Siemiginowska (2001) and van Dyk and Kang (2003).

This data generation process can be described in terms of a number of steps and intermediate unobservable quantities. Each step starts with the output from the previous step and updates it in some possibly stochastic fashion. We begin with the energies of the continuum photons and the energies of the emission line photons. In the first step, these energies are mixed together. Next, a Bernoulli random variable is generated for each photon, with the probability of success depending on the energy of the photon. If this random variable comes up positive, the photon is observed; otherwise the photon is lost to absorption or the effective area of the instrument. In another step, error is added to the remaining photon energies via the conditional multinomial distributions. Finally, the data is contaminated with Poisson background counts.

This formulation of the data generation process leads naturally to a multilevel model that formalizes each of these intermediate quantities as missing data. Given the layers of missing data, the model falls into a sequence of simple standard models. For example, we might use a loglinear model for the Poisson counts from the continuum, or a binomial regression to account for absorption (van Dyk,



Connors, Kashyap, and Siemiginowska, 2001; van Dyk and Hans, 2002). Likewise, given the parameters for each of the stochastic steps, it is a simple application of the Bayes theorem to compute the conditional distribution of each of the layers of missing data. Thus, from a computational point of view, such tools as the EM or expectation/maximization algorithm (Dempster, Laird, and Rubin, 1977), the Data Augmentation algorithm (Tanner and Wong, 1987), the Gibbs sampler (e.g., Gelfand and Smith, 1990; Smith and Roberts, 1993), and other Markov chain Monte Carlo (MCMC) methods are ideally suited to highly structured models of this sort; see van Dyk (2003). The modular structure of these algorithms fits hand in glove with the hierarchical structure of our models. This allows us to divide a complex model-fitting task into a sequence of much easier tasks. The modular structure also allows us to take advantage of well-known algorithms that exist for fitting certain components of our model. For example, using the EM algorithm to handle a blurring matrix and background contamination in Poisson image analysis is a well-known (and often rediscovered) technique (Fessler and Hero, 1994; Lange and Carson, 1984; Lucy, 1974; Meng and van Dyk, 1997; Richardson, 1972; Shepp and Vardi, 1982). Even though this standby image reconstruction algorithm is unable to handle the richness of our highly structured model, we utilize it and its stochastic generalization as a step in our mode-finding and posterior-sampling algorithms.

In this short chapter, we only present one model that we hope illustrates the complexity of the data generation process, the models of this process, the algorithms required to fit the models, and the required inference and model-checking techniques. We emphasize, however, that the multilevel structure in the data generation process is inherent to the complex scientific processes studied and the instruments used in high-energy astrophysics. Thus, the missing-data framework, the related computational techniques, and methods for Bayesian inference and model checking have many waiting applications in high-energy astrophysics.

## 26.3 Fitting narrow emission lines

In this section, we outline some of the difficulties involved in fitting the location of a narrow emission line in (26.1). Our proposed solutions including EM-type algorithms, MCMC samplers, and data-analysis techniques along with detailed examples can be found in Park and van Dyk (2004) and Park, Siemiginowska, and van Dyk (2004).

When a Gaussian density function is used to model the emission line in the simplified spectral model given in (26.1), the standard EM algorithm renders straightforward calculation of the maximum likelihood estimate (MLE) or the posterior mode; to streamline our discussion, we focus on maximum likelihood estimation in this section. We construct a multilevel missing-data structure that accounts for background contamination, the effective area of the instrument, photon absorption, the blurring of photon energies, and the mixture of continuum and emission line photons, see van Dyk, Connors, Kashyap, and Siemiginowska (2001). For clarity, we consider an *ideal instrument* that produces counts that are

a mixture of continuum and emission line photons, but these counts are not subject to the data distortion processes described in Section 26.2. Accounting for the various forms of data distortion causes no conceptual difficulty, but obscures the ideas involved with fitting an emission line. The counts from an ideal instrument are one of the levels of missing data in our formulation of the model that does account for data distortion; we call these counts the *ideal counts*. Notationally, we write $Y_j^{\text{ideal}} = Y_j^C + Y_j^L$ for each $j$, where $Y_j^{\text{ideal}}$, $Y_j^C$, and $Y_j^L$ are the total ideal counts, the counts due to the continuum, and the counts due to the emission line in bin $j$ respectively. Given the ideal counts, it is easy to construct an EM algorithm to fit (26.1); the missing data are the ideal counts split into continuum counts and emission line counts, that is, $\{(Y_j^C, Y_j^L), j = 1, \ldots, J\}$. Since the augmented-data log-likelihood is linear in these counts, the E-step simply computes the conditional expectation of the missing data. Because given the ideal counts, the photon counts due to the emission line in each bin follow binomial distributions, the conditional expectation of each is simply the total (ideal) photon counts times the relative magnitude of the emission line intensity and the combined continuum and emission line intensities at that bin. Specifically, given the current iterate of the model parameters, $\theta^{(t)} = (\theta^{C(t)}, \lambda^{(t)}, \mu^{(t)}, \sigma^{(t)})$, the E-step is given by

**E-step:** Compute $\mathrm{E}\big[Y_j^L \big| \theta^{(t)}, Y_j^{\text{ideal}}\big]$ for each bin $j = 1, \ldots, J$, that is,

$$\widehat{Y}_j^L \equiv \mathrm{E}\big[Y_j^L \big| \theta^{(t)}, Y_j^{\text{ideal}}\big] = Y_j^{\text{ideal}} \frac{\lambda^{(t)} p_j(\mu^{(t)}, \sigma^{(t)})}{\Delta_j f(\theta^{C(t)}, E_j) + \lambda^{(t)} p_j(\mu^{(t)}, \sigma^{(t)})}. \quad (26.3)$$

Next, the M-step of EM completes the update of the emission line location by computing the weighted average of the bin energies using the photon counts due to the emission line at every bin as the weights. In particular, the M-step updates the emission line location using

**M-step:** Compute $\mu^{(t+1)} = \sum_{j=1}^J E_j \widehat{Y}_j^L / \sum_{j=1}^J \widehat{Y}_j^L$,

where $E_j$ is the mean energy in bin $j$. Generally the model includes other unknown parameters such as the continuum parameters and the emission line intensity, which are also updated in the M-step. Iteration between the E-step and the M-step forms what we call the "standard EM algorithm" for maximum likelihood estimation.

Given physical constraints on emission lines, it is often appropriate to replace the Gaussian line profile with a delta function. In this case, however, the standard EM algorithm breaks down. A delta function is a limiting case of a Gaussian density that results when the Gaussian variance goes to zero. Since the data are binned, the success probability of the binomial random variable in the E-step is zero for all of the bins except the one containing the previous iterate of the line location. That is, all of the photon counts attributed to the emission line are in one bin. Thus, the M-step necessarily returns the next iterate of the line location that is the same as the previous iterate. Since the EM algorithm begins with an arbitrarily specified value of a parameter, the algorithm will converge to the mean energy of



the bin closest to the starting value in one iteration; the standard EM algorithm does not return the maximum likelihood estimate in this case.

To avoid this difficulty, we can update the line location at each iteration by maximizing the observed-data log-likelihood conditional on the other parameters in the model. To accomplish this, we simply compute the conditional observed-data log-likelihood at each possible value of the line location; because of the binning of the data, possible line locations within each bin are indistinguishable, and thus we are left with a finite number of possible distinguishable line locations. That is, this strategy updates the other model parameters by maximizing the augmented-data log-likelihood conditional on the line location using an EM iteration, and then updates the line location given the other model parameters without a missing-data formulation; this is an example of the ECME or expectation/conditional maximization either algorithm (Liu and Rubin, 1994). This algorithm allows groups of parameters to be updated by maximizing *either* the augmented-data log-likelihood or the observed-data log-likelihood while conditioning on the other parameters. The ECME algorithm is especially easy to formulate in this case because the conditional independence between the line location and the other parameters given the augmented data means that the E-step and conditional M-steps (CM-STEPS) for the other parameters are the same as in the standard EM algorithm. A difficulty with the ECME algorithm when used with real data that are subject to the data distortion processes described in Section 26.2, however, is that each iteration of the algorithm is computationally expensive, requiring the computation of the observed-data log-likelihood at each possible line location. Each evaluation involves computing, (26.2) which is time consuming because of the large dimension of the blurring matrix, $M$; this difficulty persists even when sparse matrix techniques are implemented.

As an alternative to ECME, we consider an AECM or alternating expectation/conditional maximization algorithm (Meng and van Dyk, 1997) that is computationally less expensive per iteration in this case. The AECM algorithm is so named because it allows the missing-data formulation to alternate for different groups of parameters. In terms of its use of missing data, the AECM algorithm finds middle ground between the EM and ECME algorithms. The AECM algorithm offers a more general formation than the ECME algorithm in that the CM-steps of AECM may conditionally maximize not only the observed-data log-likelihood or the conditional expectation of the augmented-data log-likelihood but also the conditional expectation of a *partially* augmented-data log-likelihood. That is, a portion of the missing data may be used to formulate some of the CM-steps in AECM. Thus, in our example, the ECME algorithm uses no missing data to formulate the CM-step for the emission line location, the EM algorithm uses all of the missing data, and the AECM algorithm allows us to formulate the CM-step using part of the missing data. In particular, we construct an augmented-data log-likelihood using the ideal counts as missing data, but do not separate the ideal counts into continuum and emission line counts. To update the line location in the AECM algorithm, we maximize this augmented-data log-likelihood conditional on

the other model parameters. As with the ECME algorithm, the CM-step evaluates the augmented-data log-likelihood at each possible value of the line location, while fixing the other parameters at their current iteration. Because this evaluation does not involve the high-dimensional blurring matrix, each iteration is much quicker than those of the ECME algorithm.

The computational advantage of AECM, however, comes at a price: For some starting values, AECM exhibits the same pathological convergence as the EM algorithm, that is, the AECM algorithm can also get stuck at a point near its starting value and thus never reach a mode of the likelihood. In order to combine the stability of ECME and the speed of AECM, we propose a *Rotation* algorithm. In the Rotation algorithm, we run one ECME iteration followed by a number of AECM iterations, and repeat this procedure, rotating between ECME and AECM until convergence. For clarity, we refer to a rotation algorithm that runs $m$ AECM iterations per ECME iteration as a Rotation($m$) algorithm. In our experience, the rotation algorithms not only find the same mode as ECME for any starting value, but also outperform ECME in terms of required computation time.

To illustrate the application of our spectral model and the various EM-type algorithms, we use a *Chandra* observation of the high redshift quasar PG1634+706. Quasars are the most distant distinct detectable objects in the universe and their study has important consequences for cosmological theory. In particular, by measuring the location of the emission line of a quasar and accounting for the expansion of the universe, we can estimate the distance of the quasar from Earth. Thus, accurate fitting of emission line locations is central to the substantive scientific questions. We modeled this data using a power law continuum with the absorption model of Morrison and McCammmon (1983) to account for absorption due to the interstellar and intergalactic media, and a power law continuum for background contamination. The model was fitted via maximum likelihood using ECME, AECM, and the Rotation(1) and Rotation(9) algorithms. Figure 26.1(a) shows the fixed values of the AECM runs for each of the 51 equally spaced starting values between 1.0 keV and 6.0 keV. If a point in this plot does not lie on a horizontal line near 2.885 keV, it indicates that the AECM iteration is fixed at a point other than the mode of the likelihood, that is, AECM did not attain the maximum likelihood as shown in Figure 26.1(b). The middle two panels of Figure 26.1 plot the log-likelihood against computation time and against iteration number using the ECME, Rotation(1), and Rotation(9) algorithms; all three algorithms were started at 4.9 keV. These plots illustrate that the use of more AECM iterations in the Rotation algorithm can make the algorithm converge to a mode significantly more quickly and that the increase of the log-likelihood per iteration is about the same in these algorithms. The bottom two panels of Figure 26.1 compare the ECME, Rotation(1), and Rotation(9) algorithms in terms of total required computation time and the number of iterations required for convergence for 51 different starting values equally spaced between 1.0 keV and 6.0 keV. Figure 26.1(e) illustrates that the Rotation(9) algorithm converged most quickly among these algorithms with every starting value. In particular, the Rotation(1) and the Rotation(9) algorithms



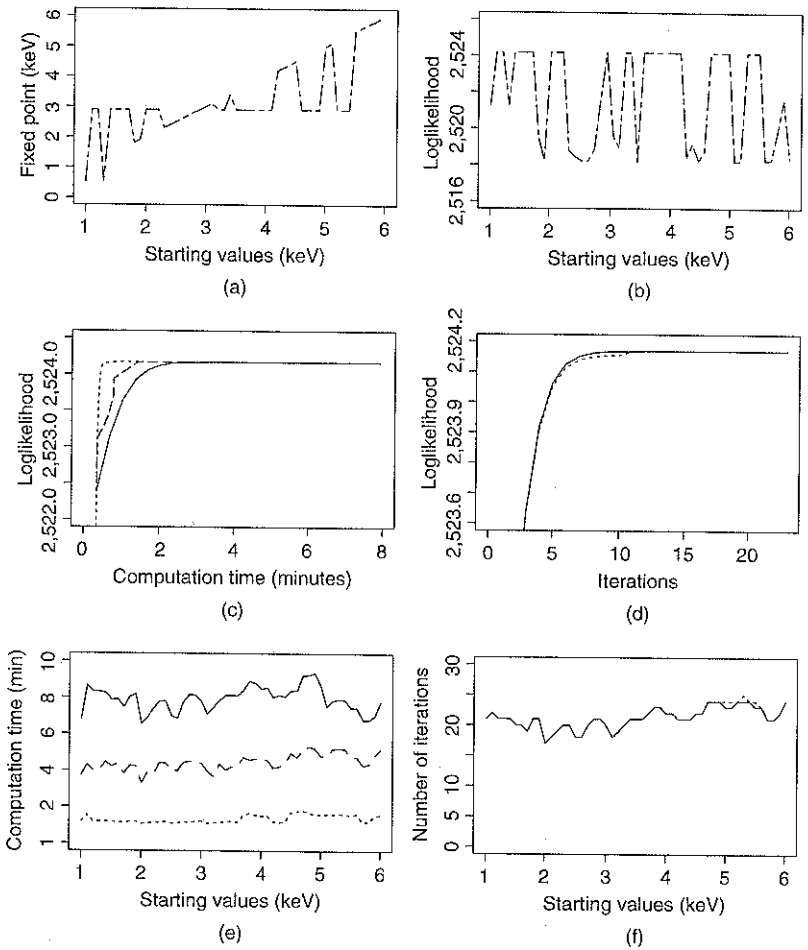


Figure 26.1 Comparison of the ECME, AECM, Rotation(1), and Rotation(9) algorithms. (a) and (b) illustrate that AECM can get stuck at a point other than a mode, thereby never reaching a mode of the likelihood; the AECM algorithm is represented by a dashed-dotted line. (c) and (d) illustrate the behavior of the log-likelihood evaluated at the iterates as a function of computation time and the iteration number using the ECME, Rotation(1), and Rotation(9) algorithms. The ECME, Rotation(1), and Rotation(9) algorithms are represented by solid, dashed, and dotted lines respectively. (e) and (f) compare the ECME and Rotation(1) and (9) algorithms in terms of computation time and the number of iterations required for convergence using 51 equally spaced starting values. The Rotation(9) algorithm is the quickest to converge among all of the algorithms considered for every starting value.

required only 1/2 and 1/10 of the computation time required by ECME respectively. While the algorithms all attain the same mode, 2.885 keV, for every starting value, Figure 26.1(f) indicates that the number of iterations required for convergence is almost identical for all of the algorithms with this data set; for a few starting values, the Rotation(9) algorithm takes more iterations than the other algorithms. Thus, in this example, the computation time for each AECM iteration is trivial relative to that of ECME, and when AECM is combined with ECME the average gain per iteration is equal to that of ECME alone.

## 26.4 Model checking and model selection

Residual plots and posterior-predictive methods (Gelman and Meng, 1996; Gelman, Meng, and Stern, 1996; Meng, 1994b; Rubin, 1981a, 1984) can be employed to check our spectral model specification. Both methods aim to check the self-consistency of the model, that is, the ability of the fitted model to predict the data to which the model was fit. The methods illustrated in this section were suggested for the spectral model by van Dyk and Kang (2003).

We consider the same model for Quasar PG1634+706 as discussed in Section 26.3 except that we compare three models for the emission line:

**Model 0:** There is no emission line.

**Model 1:** There is an emission line with fixed location in the spectrum.

**Model 2:** There is an emission line with unknown location.

The top two panels of Figure 26.2 compare the observed data with the fitted models under Models 0 and 1 in the first and second column respectively. The expected count per channel, $\Xi_I(\hat{\theta})$, is represented by a solid line and the predictive errors by dotted lines; $\hat{\theta}$ is the maximum likelihood estimate. The errors are computed using two standard deviations under the sampling model conditioning on $\hat{\theta}$; thus, these errors are based on a Gaussian approximation and do not account for the posterior variability of $\theta$. The middle two panels of Figure 26.2 are mean subtracted versions of the first two panels, that is, these panels are residual plots. To better account for the Poisson nature of the data and the posterior variability in $\theta$, we can compute residual errors using the posterior-predictive distribution. These plots appear as the final two panels in Figure 26.2; the jagged nature of the posterior-predictive residual errors is due to our Monte Carlo evaluation of this distribution. The advantage of the posterior-predictive errors is evident for the low counts in the high-energy tail of the spectra as shown in the residual plots of Figure 26.2. Comparing the two columns in Figure 26.2 near 2.885 keV also provides evidence for the inclusion of the emission line.

Posterior-predictive p-values can be used to compare the three models and, thus, to quantify the evidence in the data for the emission line. We base our comparisons



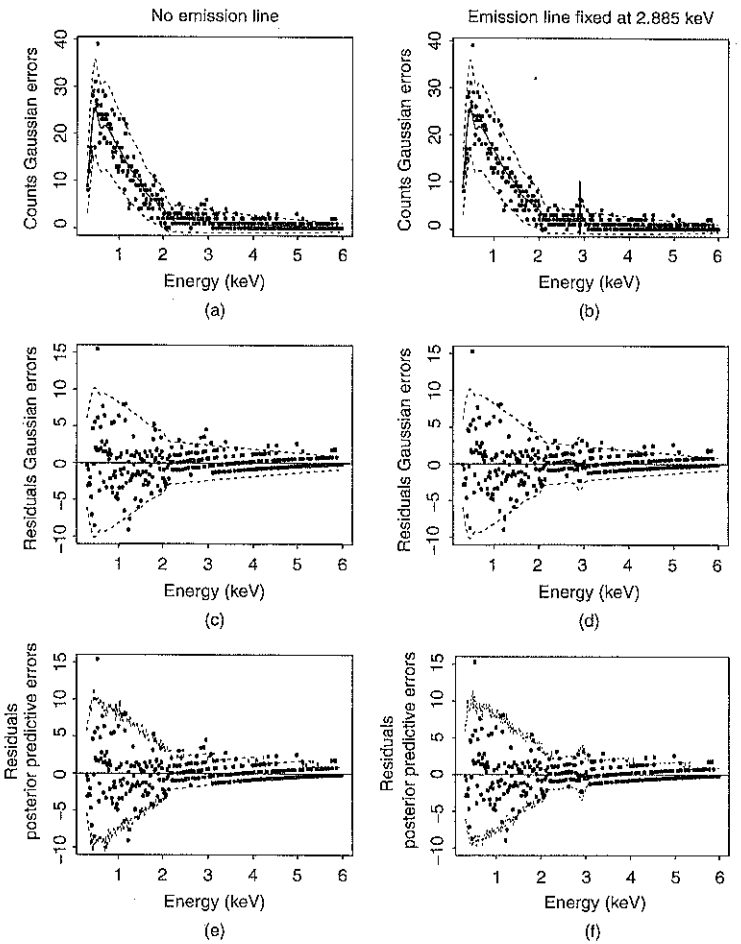


Figure 26.2 Model diagnostic plots. (a) and (b) show the data with predictive errors based on a Gaussian approximation; (c) and (d) show the residuals with errors based on a Gaussian approximation; and (e) and (f) show the residuals with errors based on the posterior predictive distribution. The two columns of the figure correspond to Models 0 and 1 respectively. The excess counts near 2.885 keV are apparent in the top two panels, thereby indicating evidence for the inclusion of the emission line in the model; the location of the emission line is represented by a vertical line in (b).

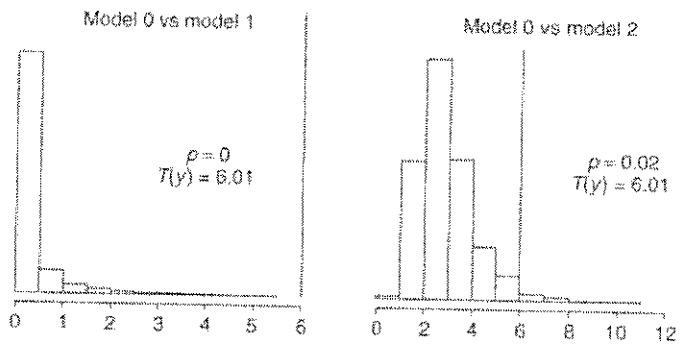


Figure 26.3 The posterior-predictive check. In each of the two histograms, the observed likelihood ratio test statistic (the vertical line) is compared with the posterior-predictive distribution of the test statistic under Model 0.

on the likelihood ratio test statistic,

$$T_i(y_{\text{rep}}) = \log\left\{\frac{\sup_{\theta\in\Theta_i} L(\theta|y_{\text{rep}})}{\sup_{\theta\in\Theta_0} L(\theta|y_{\text{rep}})}\right\}, \quad i = 1, 2,$$

where $\Theta_0$, $\Theta_1$, and $\Theta_2$ represent the parameter spaces under Models 0, 1, and 2 respectively, and $y_{\text{rep}}$ is a replicate data set. We can generate a sample from the posterior-predictive distribution of $T_i(y_{\text{rep}})$ under Model 0; we use the EM-type algorithms described above to compute $T_i(y_{\text{rep}})$. Histograms of $T_1(y_{\text{rep}})$ and $T_2(y_{\text{rep}})$ appear in Figure 26.3. Comparing these distributions with the observed values of the test statistics yields the posterior-predictive p-values in Figure 26.3. There is strong evidence for the presence of the emission line in the spectrum. Thus, Models 1 and 2 are preferable to Model 0.