

例 4.2 例 4.1 の続き 図 4.2 には, CPO_i と KL_i を計算した結果が示されている. 図において, 左側は各値のプロット (CPO については対数値), 右側は箱ひげ図である. 比較のため, 最小自乗推定値から計算された残差プロットもあわせて示されている. この図より, いくつかの観測値についてはモデルの当てはまりがよくないことが分かる. 特に $i = 372$ 番目のデータについては, もっとも当てはまりがよくない結果となっている.

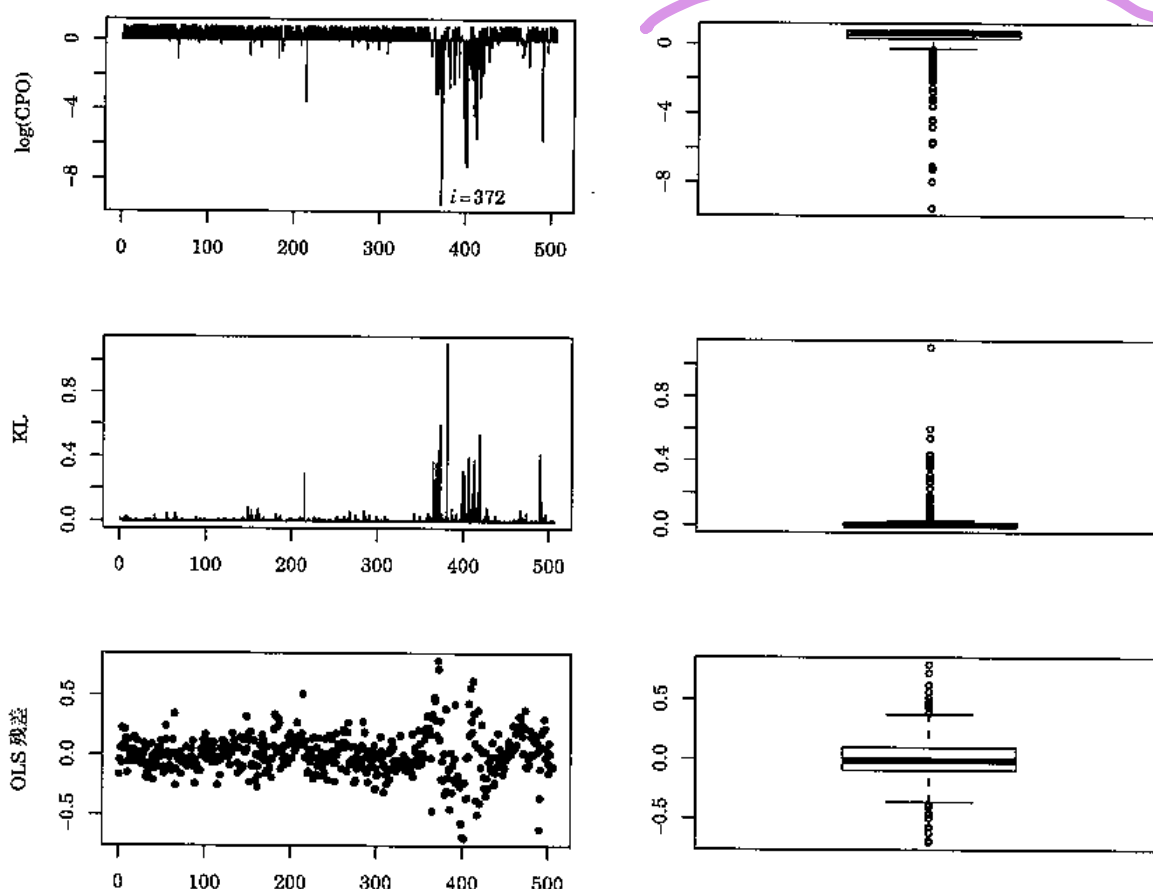


図 4.2 ボストン住宅データ: CPO (上段), KL (中段), 残差 (下段) のプロット (左) と箱ひげ図 (右)

4.2.2 モデルの選択

第 1 章で説明したように, ベイズ分析におけるモデル選択は周辺尤度に基づいて行われる. しかしながら, (4.3) 式の事前分布のもとでは周辺尤度を解析的に求めることができないため, 数値的に計算しなければならない.

(4.10) 式から, 周辺尤度は事前分布に関する期待値となっていることが分か

$$\beta \sim N(\beta_0, B_0), \quad \sigma^2 \sim IG\left(\frac{n_0}{2}, \frac{s_0}{2}\right) \quad (4.3)$$

\downarrow 事前分布
 \downarrow 事後分布

る。そこで、モンテカルロ積分によって周辺尤度を計算することが考えられる。すなわち、事前分布 $\pi(\boldsymbol{\theta})$ から $\boldsymbol{\theta}^{(t)}$ ($t=1, \dots, T$) をサンプリングし、

$$\hat{m}_{\text{MC}}(\mathbf{y}) = \frac{1}{T} \sum_{t=1}^T f(\mathbf{y}|\boldsymbol{\theta}^{(t)})$$

によって周辺尤度を推定する。この方法は、簡単に実行することができるが、事前分布と尤度関数とが大きく異なるときには非効率であり、また精度もよくない。

Gelfand and Dey (1994) が示したように、任意の正則な確率分布 $q(\boldsymbol{\theta})$ に対して、

$$\begin{aligned} \frac{1}{m(\mathbf{y})} &= \frac{1}{m(\mathbf{y})} \int_{\Theta} q(\boldsymbol{\theta}) d\boldsymbol{\theta} & m(\mathbf{y}) &= \int_{\Theta} f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}, \quad m(\mathbf{y}_{-i}) = \int_{\Theta} f(\mathbf{y}_{-i}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \quad (4.10) \\ &= \int_{\Theta} \frac{q(\boldsymbol{\theta})}{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})} \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{m(\mathbf{y})} d\boldsymbol{\theta} \\ &= \int_{\Theta} \frac{q(\boldsymbol{\theta})}{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})} \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \end{aligned}$$

ベイズの定理

$$\pi(\boldsymbol{\theta}|\mathbf{x}) = \frac{f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int_{\Theta} f(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

が成立する。そこで、 $\boldsymbol{\theta}^{(t)}$ を事後分布からのサンプルとすれば、

$$\hat{m}(\mathbf{y}) = \left[\frac{1}{T} \sum_{t=1}^T \frac{q(\boldsymbol{\theta}^{(t)})}{f(\mathbf{y}|\boldsymbol{\theta}^{(t)})\pi(\boldsymbol{\theta}^{(t)})} \right]^{-1}$$

によって周辺尤度を推定することができる。 $q(\boldsymbol{\theta})$ として事前分布を選べば、

$$\hat{m}_{\text{HM}}(\mathbf{y}) = \left[\frac{1}{T} \sum_{t=1}^T \frac{1}{f(\mathbf{y}|\boldsymbol{\theta}^{(t)})} \right]^{-1}$$

が得られる。この $\hat{m}_{\text{HM}}(\mathbf{y})$ は、周辺尤度の調和平均推定量 (harmonic mean estimator) と呼ばれている (Newton and Raftery 1994; Raftery *et al.*, 2007). 調和平均推定量は周辺尤度の一致推定量であるが、尤度関数の逆数を含んでいることから、その分散が無限大になってしまう場合がある。→ 尤度関数の分散が無限大になる。

調和平均推定量のこうした問題点を解決するため、Geweke (1999) は、事後分布からのサンプル $\boldsymbol{\theta}^{(t)}$ を使って、

$$\hat{\boldsymbol{\theta}} = \frac{1}{T} \sum_{t=1}^T \boldsymbol{\theta}^{(t)}, \quad \hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T (\boldsymbol{\theta}^{(t)} - \hat{\boldsymbol{\theta}})(\boldsymbol{\theta}^{(t)} - \hat{\boldsymbol{\theta}})'$$

を計算し、事前分布のかわりに切断正規分布

$$q(\boldsymbol{\theta}) = N(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}})I(\boldsymbol{\theta} \in \Theta^*)$$

を用いることを提案している。ここで、 Θ^* は

$$\Theta^* = \{\boldsymbol{\theta} | (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq \chi_p^2(\alpha)\}$$

によって定義される領域である。また、 p は $\boldsymbol{\theta}$ の次元 (線形回帰モデルでは $p = k + 1$) であり、 $\chi_{df}^2(\alpha)$ は自由度 df の χ^2 分布の $100\alpha\%$ 点を表す (α としては、0.5, 0.75, 0.9 がよく用いられる)。実際に周辺尤度を計算するときには、切断正規分布の正規化定数が必要となる。これについては、正規分布 $N(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\Sigma}})$ から多数のサンプルを発生させ、それらが Θ^* に入る割合を計算することによって求めることができる。

周辺尤度については、

$$m(\mathbf{y}) = \frac{f(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{\theta}|\mathbf{y})} \quad (4.11)$$

← ベイズの定理を切。

で与えられる恒等式が成立する。この式は任意の $\boldsymbol{\theta}$ に対して成り立つので、 $\boldsymbol{\theta}$ のある値 $\boldsymbol{\theta}^* = (\boldsymbol{\beta}^*, \sigma^{*2})$ で評価し、両辺の対数をとれば、

$$\log m(\mathbf{y}) = \log f(\mathbf{y}|\boldsymbol{\theta}^*) + \log \pi(\boldsymbol{\theta}^*) - \log \pi(\boldsymbol{\theta}^*|\mathbf{y}) \quad (4.12)$$

を得る ($\boldsymbol{\theta}^*$ の選択については後述)。Chib (1995) は、(4.12) 式に基づいて周辺尤度を推定する方法を提案しており、以下、彼の方法について説明する。

線形回帰モデルでは、尤度関数と事前分布は簡単に計算することができるので、 $\pi(\boldsymbol{\theta}^*|\mathbf{y})$ を求めることができれば周辺尤度が推定できることになる。そこで、 $\pi(\boldsymbol{\theta}^*|\mathbf{y})$ を

$$\pi(\boldsymbol{\theta}^*|\mathbf{y}) = \pi(\boldsymbol{\beta}^*|\mathbf{y})\pi(\sigma^{*2}|\boldsymbol{\beta}^*, \mathbf{y}) \quad (4.13)$$

と分解して表すことにする。(4.7) 式より、 $\pi(\sigma^{*2}|\boldsymbol{\beta}^*, \mathbf{y})$ は逆ガンマ分布の密度関数であるから、これを評価することは簡単である。また、 $\pi(\boldsymbol{\beta}^*|\mathbf{y})$ については、

$$\pi(\boldsymbol{\beta}^*|\mathbf{y}) = \int_0^\infty \pi(\boldsymbol{\beta}^*|\sigma^2, \mathbf{y})\pi(\sigma^2|\mathbf{y})d\sigma^2$$

と表すことができ、 $\pi(\boldsymbol{\beta}^*|\sigma^2, \mathbf{y})$ は正規分布の密度関数であるから、ギブス・サンプリングによってサンプリングした $\sigma^{2(t)}$ ($t = 1, \dots, T$) を用いれば、

$$\hat{\pi}(\beta^*|\mathbf{y}) = \frac{1}{T} \sum_{t=1}^T \pi(\beta^*|\sigma^{2(t)}, \mathbf{y})$$

によって推定することができる。したがって、周辺尤度の推定量として

$$\log \hat{m}(\mathbf{y}) = \log f(\mathbf{y}|\theta^*) + \log \pi(\theta^*) - \log \hat{\pi}(\theta^*|\mathbf{y})$$

が得られることになる。ここで、

$$\log \hat{\pi}(\theta^*|\mathbf{y}) = \log \hat{\pi}(\beta^*|\mathbf{y}) + \log \pi(\sigma^{*2}|\beta^*, \mathbf{y})$$

である。(4.11) 式は任意の θ について成立するので、実際の計算ではどのような θ^* 値を選択しても構わない。しかし、推定の効率性や安定性などを考慮して、事後平均や事後モードなどがよく用いられている。線形回帰モデルでは、パラメータが2つのブロックに分割されていたが、ブロック数が3つ以上の場合についても Chib の方法を用いることができる(章末の補論を参照)。また、 $\log \hat{m}(\mathbf{y})$ の標準誤差の求め方については、Chib (1995) を参照してほしい。

線形回帰モデルにおける周辺尤度の推定について説明してきたが、これらの方法は、尤度関数を評価することができ、事前分布が正則であれば、線形回帰モデル以外に対しても適用することができる。

例 4.3 例 4.1 の続き (4.8) 式で与えられる回帰モデルを M_1 と表すことにする。また、 M_1 の説明変数から ZN, INDUS, AGE を除いた回帰モデルを M_2 とする。このとき、両モデルの周辺尤度を Chib (1995) と Geweke (1999) の方法によって計算した結果が表 4.2 に示されている。この表から、モデル M_1 よりもモデル M_2 の方が対数周辺尤度の値が大きくなっており、説明変数を除いた方がよいと判断できる。また、周辺尤度の値は2つの方法でほぼ同じとなっていることが確認できる。

表 4.2 ポストン住宅データ：対数周辺尤度の計算結果

| 方法 | M_1 | M_2 |
|----------------------------|--------|--------|
| Chib | 35.705 | 63.847 |
| Geweke ($\alpha = 0.5$) | 35.707 | 63.845 |
| Geweke ($\alpha = 0.75$) | 35.692 | 63.847 |
| Geweke ($\alpha = 0.9$) | 35.690 | 63.844 |

周辺尤度やモデル選択を行うとき、パラメータの事後分布について注意しなければならないことがある。例えば、パラメータ $\theta_i \in (-\infty, \infty)$ の事後分布として $\pi(\theta_i|M_i) \propto 1$ を推定したとする。この事後分布は非正則であり、任意の

20 1. ベイズ統計学の基礎
定数 $c_1 (> 0)$ を用いて $\pi(\theta_i|M_i) = c_1$ と表すことができる。このとき、周辺尤度は

$$m(\mathbf{y}|M_i) = \int_{\theta_i} f(\mathbf{y}|\theta_i, M_i) \pi(\theta_i|M_i) d\theta_i \\ = c_1 \int_{\theta_i} f(\mathbf{y}|\theta_i, M_i) d\theta_i$$

と表される。 c_1 は任意の定数であったから、周辺尤度の値も任意となり、模型選択やモデル選択を周辺尤度に基づいて行うことはできないことになる。この例が示すように、周辺尤度やベイズ・ファクターを用いて模型選択やモデル選択を行う場合には、パラメータの事後分布は正則でなければならない。もし非正則な事後分布を用いたときには、標準化定数や他の統計量を用いる必要がある(第4章を参照)。

線形回帰モデルのベイズ分析では、しばしば無情報事前分布が用いられる。第1章で述べたように、無情報事前分布を用いたときには周辺尤度に基づいてモデル選択を行うことができない。このようなとき、Akaike (1973, 1974) によって提案された赤池情報量規準 (Akaike's information criterion : 以下 AIC)

$$\text{AIC} = -2 \log f(\mathbf{y}|\hat{\boldsymbol{\theta}}_{\text{ML}}) + 2p$$

あるいは、Schwarz (1978) によるベイズ情報量規準 (Bayesian information criterion : 以下 BIC)

$$\text{BIC} = -2 \log f(\mathbf{y}|\hat{\boldsymbol{\theta}}_{\text{ML}}) + p \log n$$

などの情報量規準が周辺尤度のかわりに用いられることがある (情報量規準については小西・北川, 2004 に詳しい)。ここで、 $\hat{\boldsymbol{\theta}}_{\text{ML}}$ は $\boldsymbol{\theta}$ の最尤推定量、 p は $\boldsymbol{\theta}$ の次元を表す。

BIC については、周辺尤度との関係が知られている。このことを説明するために、

$$h(\boldsymbol{\theta}) = -\frac{1}{n} \log f(\mathbf{y}|\boldsymbol{\theta}) - \frac{1}{n} \log \pi(\boldsymbol{\theta})$$

とおき、周辺尤度を

$$m(\mathbf{y}) = \int_{\boldsymbol{\theta}} f(\mathbf{y}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} = \int_{\boldsymbol{\theta}} e^{-nh(\boldsymbol{\theta})} d\boldsymbol{\theta}$$

と書き直すことにする。この式にラプラス法を適用すれば、

$$m(\mathbf{y}) \approx \left(\frac{2\pi}{n}\right)^{p/2} |\Sigma_h(\boldsymbol{\theta}^*)|^{1/2} e^{-nh(\boldsymbol{\theta}^*)}$$

なる近似が得られる ((1.10) 式を参照)。ここで、 $\boldsymbol{\theta}^*$ は事後モードを表し、 $\Sigma_h^{-1}(\boldsymbol{\theta}^*) = \partial^2 h(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \partial \boldsymbol{\theta}' |_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ である。事前分布が $\pi(\boldsymbol{\theta}) = 1$ であるとすれば、事後モードと最尤推定量は同じものとなるので、 $\boldsymbol{\theta}^*$ を $\hat{\boldsymbol{\theta}}_{\text{ML}}$ に置き換えることにする。さらに、 n が十分大きいときに定数となる項を無視すれば、

$$\log m(\mathbf{y}) \approx -\frac{p}{2} \log n + \log f(\mathbf{y}|\hat{\boldsymbol{\theta}}_{\text{ML}}) = -\frac{1}{2} \text{BIC}$$

が得られる。この式より、BIC は対数周辺尤度の近似であると理解することができる。

Handwritten derivation of the BIC approximation:

$$\begin{aligned} \log m(\mathbf{y}) &= \int \log f(\mathbf{y}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ &\approx \int \left[\log f(\mathbf{y}|\hat{\boldsymbol{\theta}}_{\text{ML}}) + \frac{1}{n} \frac{\partial \log f(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{ML}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{ML}}) \right. \\ &\quad \left. + \frac{1}{2n} \frac{\partial^2 \log f(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{ML}}} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{ML}})^2 \right] \pi(\boldsymbol{\theta}) d\boldsymbol{\theta} \end{aligned}$$

Since $\frac{\partial \log f(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{ML}}} = 0$, the first-order term vanishes. The second-order term is a constant, so it can be pulled out of the integral:

$$\log m(\mathbf{y}) \approx \log f(\mathbf{y}|\hat{\boldsymbol{\theta}}_{\text{ML}}) + \frac{1}{2n} \frac{\partial^2 \log f(\mathbf{y}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{\text{ML}}} \int (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{ML}})^2 \pi(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

The integral term is a constant, so it can be absorbed into the log-prior term, leading to the final result:

$$\log m(\mathbf{y}) \approx -\frac{p}{2} \log n + \log f(\mathbf{y}|\hat{\boldsymbol{\theta}}_{\text{ML}}) = -\frac{1}{2} \text{BIC}$$

最近では, Spiegelhalter *et al.* (2002) によって提案された偏差情報量規準 (deviation information criterion: 以下 DIC) によるモデル選択もよく目にするようになった. DIC も, AIC や BIC 同様に, モデルの当てはまりのよさとモデルの複雑さを考慮した規準となっており,

$$D(\boldsymbol{\theta}) = -2 \log f(\mathbf{y}|\boldsymbol{\theta})$$

と定義すれば, DIC ではモデルの当てはまりのよさを

$$\bar{D}(\boldsymbol{\theta}) = \int_{\Theta} D(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} = \int_{\Theta} -2 \log f(\mathbf{y}|\boldsymbol{\theta}) \cdot \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}$$

によって評価する. 一方, モデルの複雑さについては, 有効パラメータ数 (effective number of parameters) と呼ばれる

$$p_D = \bar{D}(\boldsymbol{\theta}) - D(\hat{\boldsymbol{\theta}})$$

によって評価する. ここで, $\hat{\boldsymbol{\theta}}$ は $\boldsymbol{\theta}$ の事後平均を表す. これら 2 つを組み合わせれば, DIC は

$$\text{DIC} = \bar{D}(\boldsymbol{\theta}) + p_D = D(\hat{\boldsymbol{\theta}}) + 2p_D$$

によって定義され, DIC の値が小さいほどよいモデルということになる. 多くの場合, $\bar{D}(\boldsymbol{\theta})$ を解析的に求めることができないが, 事後分布からのサンプル $\boldsymbol{\theta}^{(t)}$ から

$$\hat{\bar{D}}(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T D(\boldsymbol{\theta}^{(t)})$$

によって推定することができる.

ここで説明した以外にも, さまざまな情報量規準が提案されている (Ando, 2010). 中でも, Watanabe (2010, 2013) によって提案された WAIC (widely applicable information criterion) と WBIC (widely applicable Bayesian information criterion) は, 新しい情報量規準として注目を浴びている.

例 4.4 スプライン非線形回帰モデル

$$y_i = \sum_{j=1}^k \beta_j B_j(x_i) + u_i, \quad u_i \sim N(0, \sigma^2) \quad (i = 1, \dots, n) \quad (4.14)$$

を考えることにする. ここで, x_i は 1 次元の説明変数, β_j ($j = 1, \dots, k$) は未知

パラメータである。また、 $B_j(x)$ は節点 (knot) $\kappa_1, \dots, \kappa_m$ に対して定義される d 次の **B-スプライン基底関数** (B-spline basis function) を表し、 $k = m + d + 1$ である (B-スプラインの性質や計算方法については、de Boor, 2001 に詳しい)。パラメータの事前分布については、 σ^2 に対して

$$\sigma^2 \sim IG\left(\frac{n_0}{2}, \frac{s_0}{2}\right)$$

を仮定する。また、 $\beta = (\beta_1, \dots, \beta_k)'$ については、過剰適合の問題を避けるため、Lang and Brezger (2004) にならい、

$$\beta_j - \beta_{j-1} \sim N(0, \phi^2) \quad (j = 2, \dots, k) \quad (4.15)$$

であるとし、 β_1 については無情報事前分布 $\pi(\beta_1) \propto 1$ を仮定する (したがって、 β の事前分布は非正則である)。さらに、 ϕ^2 に対しては逆ガンマ事前分布

$$\phi^2 \sim IG\left(\frac{m_0}{2}, \frac{r_0}{2}\right)$$

を採用し、階層的な事前分布を考えることにする。

(4.14) 式は、説明変数ベクトルが $\mathbf{x}_i = (B_1(x_i), \dots, B_k(x_i))'$ である線形回帰モデルとなっていることから、 σ^2, β, ϕ^2 の完全条件付き分布は以下のように導出される。

- σ^2 の条件付き分布：

$$\pi(\sigma^2 | \beta, \phi^2, \mathbf{y}) = IG\left(\frac{n + n_0}{2}, \frac{\sum_{i=1}^n (y_i - \mathbf{x}_i' \beta)^2 + s_0}{2}\right)$$

- β の条件付き分布：

$$\pi(\beta | \phi^2, \sigma^2, \mathbf{y}) = N(\hat{\beta}, \hat{B})$$

ここで、

$$\hat{B}^{-1} = \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i'}{\sigma^2} + \frac{1}{\phi^2} \mathbf{K}_0, \quad \hat{\beta} = \hat{B} \left(\sum_{i=1}^n \frac{\mathbf{x}_i y_i}{\sigma^2} \right)$$

である。また、 \mathbf{K}_0 は (4.15) 式より導出される $k \times k$ 行列で、

$$\mathbf{K}_0 = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ & & & -1 & 2 & -1 \\ & & & & -1 & 1 \end{pmatrix}$$

によって定義される (空白の要素は 0 である)。

• ϕ^2 の条件付き分布：

$$\pi(\phi^2 | \beta, \sigma^2, \beta, \mathbf{y}) = IG \left(\frac{k-1+m_0}{2}, \frac{\beta' \mathbf{K}_0 \beta + r_0}{2} \right)$$

$k-1$ は行列 \mathbf{K}_0 のランクに対応している。

ここでは、オートバイ衝突実験データ (Silverman, 1985; Härdle, 1990; Eilers and Marx, 1996) を使い、先に説明した AIC, BIC, DIC を用いて節点の数 m を選択することにする。そのために、B-スプライン基底関数の次数は $d=3$ とし、節点については Ruppert *et al.* (2003) を参考にして、

$$\kappa_l = \text{重複を除いた説明変数の } \frac{l+1}{m+2} \text{ の分位点}$$

表 4.3 オートバイ衝突実験データ：情報量規準の計算結果

| m | AIC | BIC | DIC |
|-----|----------------|----------------|----------------|
| 5 | 1246.66 | 1275.56 | 1250.67 |
| 6 | 1223.20 | 1255.00 | 1221.70 |
| 7 | 1219.90 | 1254.59 | 1217.23 |
| 8 | 1219.75 | 1257.32 | 1216.84 |
| 9 | 1222.70 | 1263.16 | 1218.99 |
| 10 | 1224.46 | 1267.82 | 1219.44 |
| 11 | 1225.41 | 1271.66 | 1218.57 |
| 12 | 1224.41 | 1273.54 | 1217.37 |
| 13 | 1227.35 | 1279.38 | 1219.30 |
| 14 | 1228.23 | 1283.14 | 1218.86 |
| 15 | 1229.22 | 1287.03 | 1218.70 |

(太字は最小値)

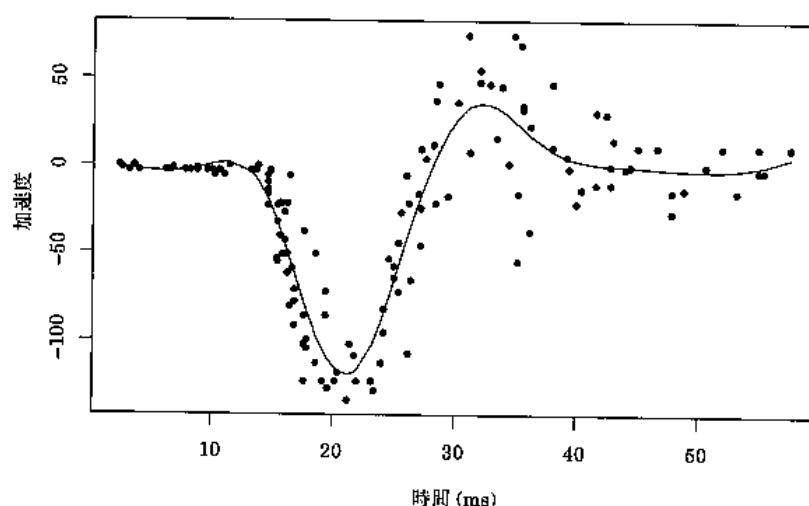


図 4.3 オートバイ衝突実験データ： $m=8$ のスプライン曲線 (実線) とデータ (点)

とした。また、事前分布のハイパーパラメータについては、 $m_0 = n_0 = 5$, $r_0 = s_0 = 0.01$ を用いた。このような設定のもと、20000 回のギブス・サンプリングを走らせることによって計算を行い、その結果が表 4.3 に示されている。なお推定にあたっては、最初の 5000 個のサンプルは収束までの期間として棄て、残りの 15000 個のサンプルを用いている。表より、AIC と DIC は $m = 8$, BIC では $m = 7$ において最小となっていることが分かる。また、図 4.3 には、 $m = 8$ としたときのスプライン曲線の推定値が示されている。

4.3 変数選択

回帰分析では、どの変数を回帰モデルに含めるかという変数選択 (variable selection) も重要な問題の 1 つである。もし、説明変数の候補が k 変数あれば、 2^k とおりの可能な組み合わせ (モデル) があり、説明変数の数が少なければ、前節で説明した周辺尤度や情報量規準を計算することによって、必要な変数を選ぶことができるであろう。しかし、説明変数の候補が多いときには、周辺尤度や情報量規準による方法はかなりの計算量を必要とし、実行が難しくなる (例えば $k = 15$ であれば、周辺尤度や情報量規準を $2^{15} = 32768$ 回計算しなければならない)。ベイズ的推測では、回帰係数の事前分布を工夫することによって、計算量をさほど増やすことなく変数選択を行うことができる (O'Hara and Sillanpää, 2009)。

4.3.1 確率的探索変数選択

説明変数の候補をすべて含む線形回帰モデルを

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i, \quad u_i \sim N(0, \sigma^2) \quad (i = 1, \dots, n) \quad (4.16)$$

と表すことにする。ただし、 $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})'$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$ である。ここでは説明を簡略化するため、すべての変数を変数選択の対象とする (定数項など必ずモデルに含めたい変数がある場合でも、若干の修正を行うことで以下で説明する方法を適用することができる)。

George and McCulloch (1993) によって提案された確率的探索変数選択