Bayesian approach to nonlinear regression models

B.Hu C.Hao X.Chen

April 2020

Content

- What is Langmuir Equation?
- Introduction to Data and Two Nonlinear Models.
- MLE and Bayesian Approaches.
- Predictive Distribution Approach and CPO Estimates.
- Model Determination
- Reference

What is Langmuir equation?

- The Langmuir equation (Langmuir, 1918) correlates the amount of adsorbed gases y on plane surfaces of glass, mica, and platinum with the equilibrium aqueous concentration x through this nonlinear function.
$$y = \frac{\alpha\beta x}{1+\alpha x}$$
where $\alpha > 0$ is Langmuir constant and $\beta > 0$ is the maximum adsorption capacity of the solid phase.

Introduction to Data

Adsorption data of aqueous PVA on 9.24 g/L of an Si oxid

OBS	Amount adsorbed (y)	Equilibrium aqueous concentration (x)
1	46.79	3.17
2	46.54	3.48
3	95.82	3.56
4	95.57	3.86
5	201.48	7.14
6	201.28	7.39
7	471.19	101.27
8	469.27	103.65
9	602.63	281.47
10	598.54	286.56
11	696.43	637.41
12	691.17	643.96
13	773.07	1126.94
14	744.45	1162.55
15	835.45	1725.30
16	805.88	1761.88

- Model 1: Langmuir equation linearization:

$$y = \frac{\alpha\beta x}{1+\alpha x} \rightarrow \frac{x}{y} = \frac{1}{\alpha\beta} + \frac{1}{\alpha}x$$

- Model 2: $y = \alpha + \beta \log(x) + \epsilon$.
- We plot the adsorbtion data and least squre fit.
- The adsorbed data could be well fitted by the M1 and M2.

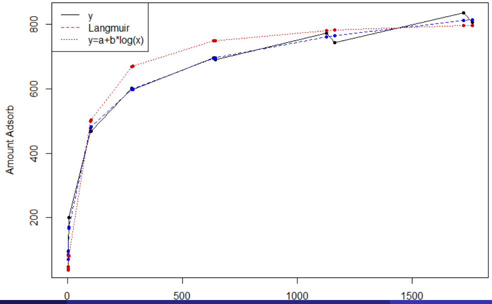


Table 2
Least square fit results for adsorption data

OBS	y	Predicted values under M_1	Residual under M_1	Predicted values under M_2	Residual under M_2	Supporting sign [a]
1	46.79	37.68	9.11	71.12	−24.34	+
2	46.54	41.18	5.36	82.10	−35.56	+
3	95.82	42.07	53.74	84.77	11.04	−
4	95.57	45.43	50.14	94.29	1.28	−
5	201.48	80.28	121.20	166.65	34.83	−
6	201.28	82.81	118.47	170.70	30.58	−
7	471.19	499.27	−28.08	478.64	−7.45	−
8	469.27	503.85	−34.58	481.37	−12.10	−
9	602.63	668.77	−66.14	598.89	3.73	−
10	598.54	671.04	−72.50	601.00	−2.46	−
11	696.43	748.52	−52.09	695.05	1.38	−
12	691.17	749.23	−58.06	696.25	−5.09	−
13	773.07	780.48	−7.41	762.09	10.98	+
14	744.45	781.82	−37.37	765.74	−21.30	−
15	835.45	795.83	39.62	812.19	23.26	−
16	805.88	796.45	9.43	814.66	−8.78	−

'+' means in favor of model M_1 and '−' means in favor of model M_2.

- Transfer modle 1(Langmuir Equation):

$$y = \frac{\alpha\beta x}{1+\alpha x}$$

- Take log

$$\log y = \log\alpha + \log\beta + \log x - \log(1+\alpha x) + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$.

- We reparameterize model 1 by letting $\alpha = e^{\alpha^*}$ and $\beta = e^{\beta^*}$.
- Model 1:

$$\log y = \alpha^* + \beta^* + \log x - \log\left(1 + e^{\alpha^*}x\right) + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$.

- The log-likelihood function for M1

$$\ell_{M1} = -n \log \sigma - \frac{1}{2\sigma^2} \sum_i (\log y_i - \alpha^* - \beta^* - \log x_i + \log(1 + e^{\alpha^*} x))^2$$

- The log-likelihood function for M2

$$\ell_{M2} = -n \log \sigma - \frac{1}{2\sigma^2} \sum_i (y_i - \alpha^* - \beta^* \log x_i)^2$$

- We calculate the 95% confidence interval.

$$\mathrm{SE}\left(\hat{\theta}_{\mathrm{ML}}\right) = \frac{1}{\sqrt{\mathbf{I}\left(\hat{\theta}_{\mathrm{ML}}\right)}} = \frac{1}{\sqrt{-\mathbf{H}\left(\hat{\theta}_{\mathrm{ML}}\right)}}$$

where I is the fisher information and H is the hessian matrix.$\mathbf{H}(\theta) = \frac{\partial^2}{\partial\theta_i \partial\theta_j} \ell(\theta), \quad 1 \leq i, j \leq 3.$

- Transfer modle 1(Langmuir Equation),

$$\log y = \log \alpha + \log \beta + \log x - \log(1 + \alpha x) + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$.

- Using a Bayesian model-fitting technique, we reparameterize model 1 by letting $\alpha = e^{\alpha^*}$ and $\beta = e^{\beta^*}$.
- Model 1:

$$\log y = \alpha^* + \beta^* + \log x - \log\left(1 + e^{\alpha^*} x\right) + \epsilon$$

where $\epsilon \sim N(0, \sigma^2)$.

- We employ noninformative priors on model parameters.

$$\pi\left(\alpha^*, \beta^*, \sigma^2\right) \propto 1/\left(\sigma^2\right)$$

- The conditional posterior distribution of σ^2 given α^*, β^*

$$
\begin{aligned}
&\pi\left(\sigma^2|\alpha^*, \beta^*, D\right) \propto \pi\left(D|\alpha^*, \beta^*, \sigma^2\right) \pi\left(\alpha^*, \beta^*, \sigma^2\right) \\
&\propto (\sigma^2)^{\frac{-n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_i (\log y_i - \alpha^* - \beta^* - \log x_i + \log(1 + e^{\alpha^*} x_i))^2\right\} \cdot \frac{1}{\sigma^2} \\
&= \exp\left\{-\frac{1}{2} \sum_i (\log y_i - \alpha^* - \beta^* - \log x_i + \log(1 + e^{\alpha^*} x_i))^2 \frac{1}{\sigma^2}\right\} \cdot \left(\sigma^2\right)^{-\frac{n}{2}-1}
\end{aligned}
$$

- Therefore, $\sigma^2|\alpha^*, \beta^*, D \sim$ inverse-gamma(a, b), where $a = n/2$, $b = \frac{1}{2}\sum_i \left(\log y_i - \alpha^* - \beta^* - \log x_i + \log\left(1 + e^{\alpha^*} x\right)\right)^2$

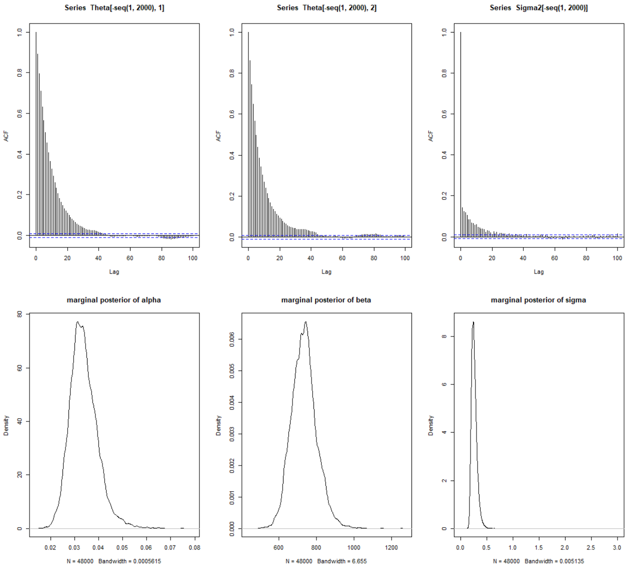
- The conditional posterior distribution of α^*, β^* given σ^2

$$\pi\left(\alpha^*, \beta^* | \sigma^2, D\right) \propto \pi\left(D | \alpha^*, \beta^*, \sigma^2\right) \pi\left(\alpha^*, \beta^*, \sigma^2\right)$$
$$\propto (\sigma^2)^{\frac{-n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_i (\log y_i - \alpha^* - \beta^* - \log x_i + \log(1 + e^{\alpha^*} x_i))^2\right\} \cdot \frac{1}{\sigma^2}$$
$$= \exp\left\{-\frac{1}{2\sigma^2} \sum_i (\log y_i - \alpha^* - \beta^* - \log x_i + \log(1 + e^{\alpha^*} x_i))^2\right\}$$

- Therefore, we use random walk Metropolis Hasting algorithm to sample α^* and β^* from its conditional posterior distribution.
- Proposal distribution is of the form bivariate normal distribution with mean $\begin{pmatrix} \alpha^* \\ \beta^* \end{pmatrix}$ and covariance matrix $V = \begin{pmatrix} 0.1 & 0 \\ 0 & 0.1 \end{pmatrix}$.

MLE and Bayesian Approaches

Bayesian Approach-Random Walk Metropolis Hastings



- Model 2

$$y = \alpha + \beta log x + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$.

- We employ noninformative priors on model parameters.

$$\pi(\alpha, \beta, \sigma^2) \propto \frac{1}{\sigma^2}.$$

- Implementation of the Gibbs sampler under this model is straightforward.

- $\pi(\alpha|\beta, \sigma^2, x, y)$ is a normal distribution.

$$\begin{aligned}\pi(\alpha|\beta,\sigma^2,x,y) &\propto \pi(y|\alpha,\beta,\sigma^2,x) \times \pi(\alpha,\beta,\sigma^2)\\ &\propto exp[-\frac{1}{2\sigma^2}\sum(y-\alpha-\beta logx)^2] \times \frac{1}{\sigma^2}\\ &\propto exp\{-\frac{n}{2\sigma^2}[\alpha^2 - 2\frac{\sum(y-\beta logx)}{n}\alpha]\},\end{aligned}$$

i.e, $\alpha|\beta, \sigma^2, x, y \sim N(\frac{\sum(y-\beta logx)}{n}, \frac{\sigma^2}{n})$.

- $\pi(\beta|\alpha, \sigma^2, x, y)$ is a normal distribution.

$$\begin{aligned}\pi(\beta|\alpha,\sigma^2,x,y) &\propto \pi(y|\alpha,\beta,\sigma^2,x) \times \pi(\alpha,\beta,\sigma^2)\\ &\propto exp[-\frac{1}{2\sigma^2}\sum(y-\alpha-\beta logx)^2] \times \frac{1}{\sigma^2}\\ &\propto exp\{-\frac{\sum(logx)^2}{2\sigma^2}[\beta^2 - 2\frac{\sum(y-\alpha)logx}{\sum(logx)^2}\beta]\},\end{aligned}$$

i.e, $\beta|\alpha, \sigma^2, x, y \sim N(\frac{\sum(y-\alpha)logx}{\sum(logx)^2}, \frac{\sigma^2}{\sum(logx)^2})$.

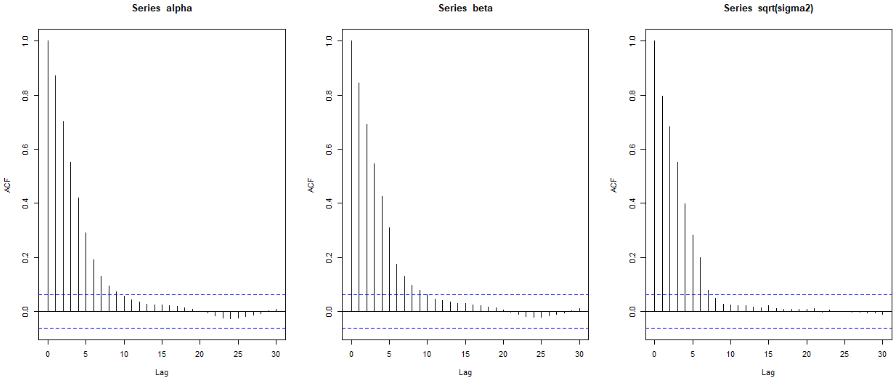
- $\pi(\sigma^2|\alpha,\beta,x,y)$ is an inverse gamma distribution.

$$\begin{aligned}\pi(\sigma^2|\alpha,\beta,x,y) &\propto \pi(y|\alpha,\beta,\sigma^2,x)\times\pi(\alpha,\beta,\sigma^2)\\ &\propto (\frac{1}{\sigma^2})^{\frac{n}{2}} exp[-\frac{1}{2\sigma^2}\sum(y-\alpha-\beta logx)^2]\times\frac{1}{\sigma^2}\\ &\propto (\frac{1}{\sigma^2})^{\frac{n}{2}+1} exp[-\frac{1}{\sigma^2}\frac{\sum(y-\alpha-\beta logx)^2}{2}],\end{aligned}$$

i.e, $\sigma^2|\alpha,\beta,x,y \sim inverse-gamma(\frac{n}{2},\frac{\sum(y-\alpha-\beta logx)^2}{2})$.

MLE and Bayesian Approaches

Bayesian Approach-Gibbs Sampling



MLE and Bayesian Approaches

	M_1		M_2	
Para- meters	MLE ($95\% CI^a$) .	Posterior mean ($95\% CI^b$).	MLE ($95\% CI^a$) .	Posterior mean ($95\% CI^b$).
α	0.033	0.034	-64.591	-64.935
	(0.024,0.044)	(0.0238,0.0472)	(-86.851,-42.323)	(-87.498,-39.867)
β	731.382	734.470	117.640	117.658
	(625.146,854.895)	(612,876)	(113.264,122.007)	(113.104 ,122.005)
σ	1.257	0.259	18.660	22.335
	(1.161,1.359)	(0.179 0.386)	(12.194,25.125)	(14.385, 33.281)

- The predictive density

$$f(\mathbf{Y}) = \int f(\mathbf{Y}|\theta, \mathbf{X})\pi(\theta)d\theta,$$

$\mathbf{Y}$: $n \times 1$ data vector,
$\mathbf{X}$: $n \times k$ matrix of explanatory variables.

- The marginal density of Y_r given the remaining

$$f(Y_r|\mathbf{Y}_{(r)}) = \frac{f(\mathbf{Y})}{f(\mathbf{Y}_{(r)})} = \int f(Y_r|\theta, \mathbf{X}, \mathbf{Y}_{(r)})\pi(\theta|\mathbf{Y}_{(r)})d\theta,$$

Y_r: the rth data vector,
$\mathbf{Y}_{(r)}$: the remaining data vectors $(Y_1, ..., Y_{r-1}, Y_{r+1}, ..., Y_n)^T$,
This is called the **cross-validation approach**, where $f(Y_r|\mathbf{Y}_{(r)})$ is called the **cross-validation predictive density**.

- The cross-validation predictive density is to be checked against y_r. This means if model holds, y_r may be viewed as a random observation from the cross-validation predictive density.
- $g(Y_r; y_r)$: checking function
- d_r: the expectation of $g(Y_r; y_r)$ under $f(Y_r|\mathbf{y}_{(r)})$
- One possible choice of the checking function

$$g_\epsilon(Y_r; y_r) = \frac{1}{2\epsilon} I_{C_r(\epsilon)}(Y_r)$$

$$C_r(\epsilon) = \{Y_r : y_r - \epsilon \le Y_r \le y_r + \epsilon\}$$

$$d_r(\epsilon) = E[g_\epsilon(Y_r; y_r)|\mathbf{y}_{(r)}] = \frac{1}{2\epsilon} P(C_r(\epsilon)|\mathbf{y}_{(r)})$$

- When ϵ is close to 0, Y_r is close to y_r and $lim C_r(\epsilon) = y_r$,

$$g(Y_r; y_r) = lim g_\epsilon(Y_r; y_r) = I_{lim C_r(\epsilon)}(Y_r),$$
$$d_r = E[g(Y_r; y_r)|\mathbf{y}_{(r)}] = f(y_r|\mathbf{y}_{(r)})$$

The quantity of the equation above is called the conditional predictive ordinate(**CPO**).

Predictive Distribution Approach and CPO Estimates

Predictive Distribution

- Three methods to estimate CPO.
- Given the checking function $g(Y_r; y_r)$,

$$d_r = E[g(Y_r; y_r)|\mathbf{y}_{(r)}] = \int\int g(Y_r; y_r) f(Y_r|\theta, \mathbf{X}, \mathbf{y}_{(r)})\pi(\theta|\mathbf{y}_{(r)}) d\theta dY_r$$

  d_r involves a multidimensional integral.
- First method: Monte Carlo integration

$$\hat{d}_r = \frac{1}{B}\sum g(Y_{rs}; y_r)$$

  $(\theta_s, Y_{rs}), s = 1, ..., B$: samples from the joint conditional distribution for θ and Y_r, $f(Y_r|\theta, \mathbf{X}, \mathbf{y}_{(r)})\pi(\theta|\mathbf{y}_{(r)})$.
- Sampling from $\pi(\theta|\mathbf{y}_{(r)})$ is not a easy task.

- Second method: importance sampling

$$\hat{d}_r = \sum g(Y_{rs}; y_r) w_s,$$

$$w_s = \frac{\pi(\theta_s | \mathbf{y}_{(r)}) / h(\theta_s)}{\sum \pi(\theta_s | \mathbf{y}_{(r)}) / h(\theta_s)}, s = 1, ..., B$$

  $h(\theta)$: an importance sampling density for $\pi(\theta | \mathbf{y}_{(r)})$,
  $\theta_s, s = 1, ..., B$: drawn from $h(\theta)$.

- This method will be used in the SIR method, which will be mentioned in detail below.

Predictive Distribution Approach and CPO Estimates

Predictive Distribution

- Third method: we observe that

$$
\begin{aligned}
f(y_r|\mathbf{y}_{(r)}) = \frac{f(\mathbf{y})}{f(\mathbf{y}_{(r)})} &= \frac{\int f(\mathbf{y}|\theta)\pi(\theta)d\theta}{\int f(\mathbf{y}_{(r)}|\theta)\pi(\theta)d\theta} \\
&= \frac{\int \frac{f(\mathbf{y}|\theta)\pi(\theta)}{\pi(\theta|\mathbf{y})f(\mathbf{y})}\pi(\theta|\mathbf{y})d\theta}{\int \frac{f(\mathbf{y}_{(r)}|\theta)\pi(\theta)}{\pi(\theta|\mathbf{y})f(\mathbf{y})}\pi(\theta|\mathbf{y})d\theta} \\
&= \frac{1}{\int \frac{1}{f(y_r|\mathbf{y}_{(r)},\theta)}\pi(\theta|\mathbf{y})d\theta}
\end{aligned}
$$

- The Monte Carlo integration of CPO

$$
\hat{f}(y_r|\mathbf{y}_{(r)}) = \left(\frac{1}{B}\sum \frac{1}{f(y_r|\mathbf{y}_{(r)},\theta_s)}\right)^{-1} = B\left(\sum \frac{1}{f(y_r|\mathbf{y}_{(r)},\theta_s)}\right)^{-1}
$$

If $\{Y_r, r = 1, ..., n\}$ are conditionally independent given θ, $f(y_r|\mathbf{y}_{(r)}, \theta_s) = f(y_r|\theta_s)$.

The values of d_r for adsorption data

OBS	d_r for M_2	d_r for M_2	log_{10} difference	supporting sign
1	0.0007040442	0.006303215	-0.9519622	-
2	0.0007040764	0.001994990	-0.4523209	-
3	0.0006966145	0.013720290	-1.2943708	-
4	0.0006966808	0.015464678	-1.3463071	-
5	0.0006631989	0.003228009	-0.6872910	-
6	0.0006632955	0.004941961	-0.8721922	-
7	0.0004938062	0.015646921	-1.5008723	-
8	0.0004953031	0.014068964	-1.4533911	-
9	0.0003885033	0.015561272	-1.6026504	-
10	0.0003918202	0.015766067	-1.6046365	-
11	0.0003136740	0.014315393	-1.6593247	-
12	0.0003177801	0.014126810	-1.6479174	-
13	0.0002558821	0.010912530	-1.6298855	-
14	0.0002769818	0.008194787	-1.4710865	-

The Predictive Interval

- Sample from predictive distributions, $f(Y_r \mid \mathbf{y_{(r)}})$.
- Count the number of samples that fall within $100 \times (1 - \alpha)\%$ predictive intervals.
- If too many samples are in the predictive interval with a large α (0.5), the predictive distribution might be overdispersed. Conversely, if too few observations are in the interval with a small α (0.05), then the predictive distribuiton might be underdispersed. i.e. The model is inadequate.

Sampling/Importance resampling (SIR) method

- Generate s independent proposed samples $\theta_{\mathbf{1}}, \ldots, \theta_{\mathbf{s}}$ from an importance sampling distribution, $h(\theta) = \pi(\theta \mid y)$.
- Calculate the standardized weights, $w_s = \frac{(f(y_r|\theta_{\mathbf{s}}))^{-1}}{\sum_{j=1}^{S}(f(y_r|\theta_{\mathbf{j}}))^{-1}}$.
- Generate an approximate realization $\theta_{\mathbf{s}}^{\star}$ from $(\theta_{\mathbf{1}}, \ldots, \theta_{\mathbf{s}})$ with probability $w_1, \ldots, w_s$.
- Generate Y_{rs} from $f(y_r \mid \theta_{\mathbf{s}}^{\star})$.

Model Determination

Tool 1: Predictive Interval

Predictive intervals (PI) of Model 1 and Model 2

OBS	y	2.5%	25%	75%	97.5%	Ind1	Ind2	OBS	y	2.5%	25%	75%	97.5%	Ind1	Ind2
1	46.79	39.25	55.88	80.09	113.97	0	1	1	46.79	24.31	52.60	81.96	108.99	0	1
2	46.54	41.46	58.79	87.49	124.60	0	1	2	46.54	31.46	60.88	91.85	121.82	0	1
3	95.82	49.29	68.08	96.64	133.18	1	1	3	95.82	43.89	71.96	100.71	127.96	1	1
4	95.57	52.51	72.15	102.20	143.25	1	1	4	95.57	53.14	80.18	107.78	134.73	1	1
5	201.48	87.29	123.04	178.32	251.98	0	1	5	201.48	125.85	154.75	185.92	214.77	0	1
6	201.28	90.31	127.84	182.54	255.70	0	1	6	201.28	131.10	159.23	188.76	218.29	0	1
7	471.19	339.16	466.71	664.37	921.45	1	1	7	471.19	437.96	463.94	492.49	519.07	1	1
8	469.27	334.07	469.94	670.05	930.69	0	1	8	469.27	440.50	466.45	494.48	520.06	1	1
9	602.63	399.35	553.16	781.38	1091.46	1	1	9	602.63	559.40	585.27	613.38	640.37	1	1
10	598.54	399.37	553.00	783.67	1074.01	1	1	10	598.54	559.48	586.74	613.93	642.78	1	1
11	696.43	428.17	592.47	838.01	1162.60	1	1	11	696.43	654.55	681.41	709.53	735.29	1	1
12	691.17	426.03	597.97	837.43	1169.90	1	1	12	691.17	655.37	681.69	709.63	735.15	1	1
13	773.07	441.98	612.27	860.35	1176.02	1	1	13	773.07	722.77	749.30	777.50	804.07	1	1
14	744.45	438.46	606.98	859.48	1187.04	1	1	14	744.45	721.60	748.77	777.37	803.86	0	1
15	835.45	440.98	614.28	867.07	1207.18	1	1	15	835.45	772.88	800.76	829.18	858.28	0	1
16	805.88	446.52	620.38	871.75	1196.37	1	1	16	805.88	772.04	799.00	827.69	853.04	1	1
						$\Sigma = 11$	$\Sigma = 16$							$\Sigma = 10$	$\Sigma = 16$

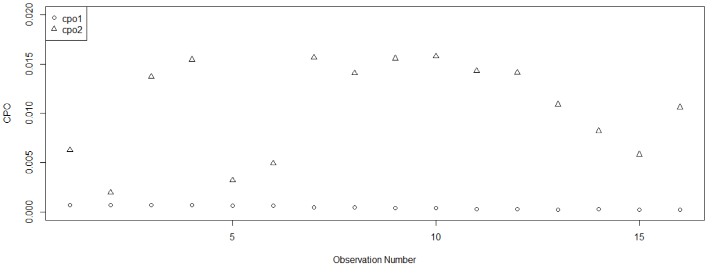
Ind1=1 indicates that the actual observation falls within 50% PI.

Ind2=1 indicates that the actual observation falls within 95% PI.

Model Determination

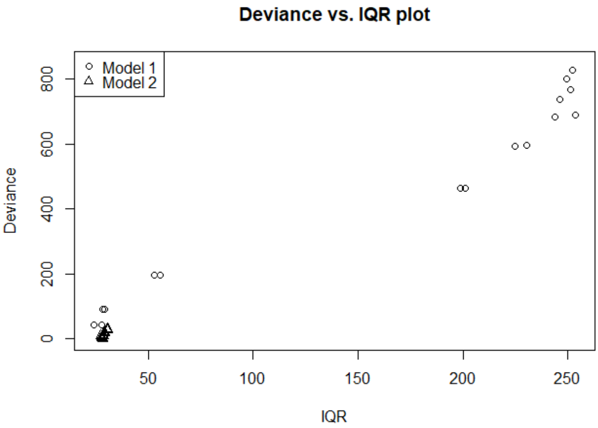
Tool 2: CPO plot

The conditional predictive ordinate (CPO) is a Bayesian diagnostic which detects surprising observations. The better model has the majority of its CPOs (d_rs) above those of the worse one.



Given the samples from the predictive distribution $f(Y_r \mid y_{(r)})$, the deviance measure $|y_r - \mu_r|$ or $|y_r - m_r|$ and spread measure $V_r = var(Y_r \mid y_{(r)})$ or $I_R = IQR(Y_r \mid y_{(r)})$. μ_r and m_r represent the mean and median of predictive values, Y_{rs}, we generated for the rth observation. V_r and I_r represent the variance and interquartile range of the predictive distribution $f(Y_r \mid y_{(r)})$. The deviance plot could be either a plot of $|y_r - \mu_r|$ vs. V_r or a plot of $|y_r - m_r|$ vs. I_r to compare several models.

Tool 3: Deviance plot



Model Determination

Tool 4: l_r plot

We define the quantity $l_r = \log_{10} PBF_{12} - \log_{10} PBF_{12}^r$, where PBF_{12}^r represents pseudo-Bayes factor excluding the rth observation, which could be used to measure the effect of observation r on the pseudo-Bayes factor. A negative l_r indicates less support for the model from observation r and a positive l_r suggests the observation r favors the model. Thus, the model with more positive l_rs is better than that with less positive l_rs.

By the definition of pseudo-Bayes factor (PBF), we have

$$PBF_{12} = \frac{\prod_{r=1}^{n} \pi(y_r \mid y_{(r)} \mid M_1)}{\prod_{r=1}^{n} \pi(y_r \mid y_{(r)} \mid M_2)}$$

where the cross validated predictive density is

$$\begin{aligned}\pi(y_r \mid y_{(r)}) &= \int \pi(y_r \mid \theta, y_{(r)})\pi(\theta \mid y_{(r)})d\theta \\ &= \frac{1}{\int \frac{1}{\pi(y_r \mid \theta, y_{(r)})}\pi(\theta \mid y_{(r)})d\theta}\end{aligned}$$

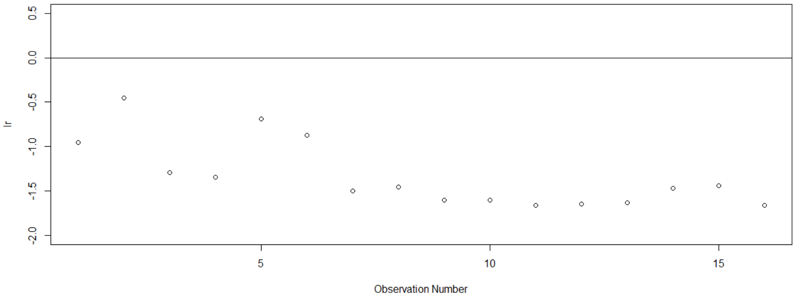
Since d_r is estimated by $\hat{\pi}(y_r \mid y_{(r)}) = B\left(\sum_{s=1}^{B} \frac{1}{\pi(y_r \mid \theta_s)}\right)^{-1}$, PBF could also be written as

$$PBF_{12} = \frac{\prod_{r=1}^{n} d_r^1}{\prod_{r=1}^{n} d_r^2}$$

and $\frac{PBF_{12}}{PBF_{12}^r} = \frac{d_r^1}{d_r^2}$. i.e. $l_r = \log_{10} PBF_{12} - \log_{10} PBF_{12}^r = \log_{10} d_r^1 - \log_{10} d_r^2$

Model Determination

Tool 4: l_r plot



Reference

Dipak K. Dey. (Dec. 1997). Bayesian Approach for Nonlinear Random Effects Models. *Biometrics.* Vol. 53, No.4.

Aitkin, M. (1991). Posterior Bayes factors (with discussion). *Journal of the Royal Statistical Society, Series B 53, 111-142.*

Devroye, L. (1986). *Non-uniform Random Variate Generation.* New York: Springer-Verlag.