Sankhyā: The Indian Journal of Statistics
Special Issue on Bayesian Analysis
1998, Volume 60, Series B, Pt. 1, pp. 145-160

BAYESIAN MODELLING OF A PHARMACEUTICAL EXPERIMENT WITH HETEROGENEOUS RESPONSES

By CHRISTIAN P. ROBERT
Laboratoire de Statistique, CREST, Paris
Université de Rouen, France
and
ANNIE REBER
Université de Rouen, France

SUMMARY. We show in this paper how a hierarchical Bayes modelling can lead to better insights in a pharmaceutical experiment assessing the performances of a drug against vertigo, by providing direct access to different extensions of the model. The availability of such extensions relies on the use of Gibbs sampling techniques which allow for simulation from arbitrary distributions. The Bayesian processing of the experiment is noninformative, which allows for an objective perspective, and we consider in detail the extension when the population is heterogeneous at some stage of the experiment, and the corresponding modelling by a normal mixture distribution.

1. Introduction

Consider an experiment where animals (rats, say) are intoxicated by a substance, then treated by either a placebo or a drug with several possible concentrations. This setup is typical of many pharmaceutical experiments and has been thoroughly treated in the theoretical literature as well as in biometric, animal breeding and psychological applications. We still undertake the study of this approach in order to show how easily a Bayesian hierarchical analysis can be implemented in this setup and also how readily it allows for extensions otherwise unattainable (e.g., by a frequentist analysis). The extensions we will consider in this paper deal with the possible heterogeneity of the animals involved in the study and the representation of this heterogeneity through a mixture structure, and with the processing of missing data. Other possible extensions include the introduction of restrictions on the different parameters of the model and the

AMS (1990) *subject classification.* 62F15, 62J10, 62J10, 62M99, 62P10.
Key words and phrases. Analysis of variance; Gibbs sampling; missing data; mixtures; normal linear model; placebo effect; probit model.

involvement of covariates either in the weights of the mixture through a probit link function or in the mean effects at each stage of the experiment.

The statistical model associated with the above experiment is chosen to be the "classical" linear additive effect; given x_{ij}, y_{ij} and z_{ij}, j-th responses of the i-th rat at the control, intoxication and treatment stages, respectively, we assume that $(1 \leq i \leq I)$

$$\begin{array}{lll} x_{ij} & \sim \mathcal{N}(\theta_i, \sigma_c^2), & 1 \leq j \leq J_i^c \\ y_{ij} & \sim \mathcal{N}(\theta_i + \delta_i, \sigma_a^2), & 1 \leq j \leq J_i^a \\ z_{ij} & \sim \mathcal{N}(\theta_i + \delta_i + \xi_i, \sigma_t^2), & 1 \leq j \leq J_i^t \end{array} \qquad \ldots(1.1)$$

where θ_i is the average control measurement, δ_i the average intoxication effect and ξ_i the average treatment effect for the i-th rat, the variances of these measurements being constant for the control, the intoxication and the treatment effects. (This assumption could be easily relaxed.)

So far, this model is quite similar to a regular analysis of variance model, although the parameterization may appear slightly excessive if the number of observations per rat is reduced to one, i.e. if $J_i^c = J_i^a = J_i^t = 1$. The purpose of the experiment, however, is not in the analysis of the individual rat destinies but rather in the assessment of the overall effect of the tested drug, with respect to the intoxication level ("*Does the drug improve the condition of the rat?*"), to the control level ("*Does the drug allow the rat to come back to the original level?*") and to the placebo effect ("*Does the drug effect differ from the placebo effect at all?*"). We thus relate the different individual averages through a common (conjugate) prior distribution, namely by assuming that $(1 \leq i \leq I)$

$$\theta_i \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2), \qquad \delta_i \sim \mathcal{N}(\mu_\delta, \sigma_\delta^2), \qquad \ldots(1.2)$$

and

$$\xi_i \sim \mathcal{N}(\mu_P, \sigma_P^2) \qquad \text{or} \qquad \xi_i \sim \mathcal{N}(\mu_D, \sigma_D^2), \qquad \ldots(1.3)$$

depending on whether the i-th rat receives a placebo treatment or a drug treatment, assuming for simplicity's sake that there is only one concentration of the drug. We denote by ℓ_i the indicator of the treatment allocated to the i-th rat, namely $\ell_i = 0$ if the treatment is the placebo and $\ell_i = 1$ if the treatment is the drug to be tested, and by I_P the number of rats which receive the placebo and by $I_D = I - I_P$ the number of rats which receive the drug. The parameters of interest are therefore the unknown mean effects μ_θ, μ_δ, μ_P and μ_D, and the difference $\mu_D - \mu_P$. We propose in Section 2 a Bayesian noninformative approach for this hierarchical model.

Formally, the introduction of two levels in the modelling of the individual responses is equivalent to assuming directly that

$$\begin{array}{ll} x_{ij} & \sim \mathcal{N}(\mu_\theta, \sigma_c^2 + \sigma_\mu^2), \\ y_{ij} & \sim \mathcal{N}(\mu_\theta + \mu_\delta, \sigma_a^2 + \sigma_\mu^2 + \sigma_\delta^2), \\ z_{ij} & \sim \mathcal{N}(\mu_\theta + \mu_\delta + \mu_P, \sigma_t^2 + \sigma_\mu^2 + \sigma_\delta^2 + \sigma_P^2) \end{array}$$

if the i-th rat receives the placebo and

$$z_{ij} \sim \mathcal{N}(\mu_\theta + \mu_\delta + \mu_D, \sigma_t^2 + \sigma_\mu^2 + \sigma_\delta^2 + \sigma_D^2)$$

otherwise. But to consider this two-level decomposition does not hinder the inferential process while allowing for easier missing data processing and mixture extensions. Moreover, this modelling seems closer to the actual phenomenon, in the sense that the responses vary from rat to rat and that observation errors have to be taken into account for a given rat. Distinguishing between observation and population errors is therefore sensible. (Note that the marginal representation above ensures that all the hyperparameters are identifiable.)

We consider in Section 3 the case when

$$\delta_i \sim p\mathcal{N}(\mu_{\delta 1}, \sigma_{\delta 1}^2) + (1-p)\mathcal{N}(\mu_{\delta 2}, \sigma_{\delta 2}^2), \qquad \ldots(1.4)$$

allowing for two levels of intoxication, i.e. for two different reactions in the rat population. While this mixture structure also transfers to the marginal distribution of the y_{ij}'s, it differs from a regular mixture model since it requires that the y_{ij}'s belong to the same component of the mixture for $1 \leq j \leq J_i^a$. The extension mentioned above would be to consider the case when p in (1.4) depends on a covariate ϱ_i through a *link function*, for instance a probit link function

$$p_i = \Phi(\alpha + \beta^t \varrho_i).$$

The actual experiment was conducted on 31 rats, assessing the effect of a drug used to remedy to vertigo—one *ml* of a solution containing $1 \cdot 10^{-13}$M picrotoxin in a saline solution—on the restoration of gaze stability; this stability was impaired by the injection of a toxic substance—$6.6 \cdot 10^{-5}$M picrotoxin in a saline solution. A placebo was injected to 16 rats instead of the drug, which was injected to 15 rats, in order to compare the effect of the drug with a natural return to stability and also to test for a placebo effect. In normal life, gaze is stabilized by a very precise adjustment of the eye velocity on the velocity of the moving surroundings. Instability usually causes blurring of vision, vertigo or symptoms of motion sickness (see Reber *et al*, 1989, for more details on the experimental conditions and the biological motivations of the experiment). The quantitative response associated with this experiment is the gain, namely the ratio of eye velocity over visual stimulus velocity, measured by derivation of the eye position recorded by an electro-physiological method based on amplitude detection (see Sarrau *et al*, 1989). As this ratio is always between 0 and 1, we transformed responses by a logit function, $\log(t/(1-t))$, in order to get real support variables x_{ij}, y_{ij} and z_{ij} for which the normal distribution is a convenient approximation. Moreover, as happens frequently for *in vivo* experiments, some responses are incomplete, i.e. failed to be reported at some stage of the study. In addition, the population appears to be heterogeneous, as shown by different levels of instability of the gaze in the intoxicated state. Therefore, the modelling

proposed in this paper includes processing of the missing data and enhances an extension which takes into account the possible heterogeneity of the population via the mixture (1.4).

This pharmaceutical study induces a pathological state for the rats in order to investigate the efficacy of a drug. However, the modelling presented here exhibits many advantages which could be replicated for most clinical studies. First, responses are frequently missing. Secondly, populations are often heterogeneous (young and old, males and females,...) with a lack of covariates to explain this heterogeneity. The heterogeneity may even not be suspected at the time of the experiment. A mixture model is then appropriate to exhibit such latent structures, at very little cost in terms of computing time.

2. Bayesian Implementation for the Classical Model

The hierarchical model corresponding to the equations (1.1)–(1.3) can be associated with a noninformative prior distribution on the different parameters of the model, i.e. with a σ-finite measure $\pi(\mu_\theta, \mu_\delta, \mu_P, \mu_D, \sigma_c, \sigma_a, \sigma_t, \sigma_\theta, \sigma_\delta, \sigma_P, \sigma_D)$ leading to a Bayesian processing of the model. The measure usually considered in location-scale setups is Jeffreys' prior,

$$\pi(\mu_\theta, \mu_\delta, \mu_P, \mu_D, \sigma_c, \sigma_a, \sigma_t, \sigma_\theta, \sigma_\delta, \sigma_P, \sigma_D) \propto (\sigma_c\sigma_a\sigma_t\sigma_\theta\sigma_\delta\sigma_P\sigma_D)^{-1}, \quad \ldots(2.1)$$

although some alternatives for analysis of variance models have been proposed by Ye and Berger (1991). This prior leads to a well-defined posterior distribution if there are at least two observations for each stage of the experiment. (This condition is weaker than the usual identifiability condition because the θ_i's and the corresponding individual mean effects for intoxication and treatment can be eliminated by integration.) The role of (2.1) is to allow for a formal Bayesian treatment of the problem without requiring the derivation of a subjective prior distribution, thus keeping both an objective perspective and the advantages of a Bayesian approach. In fact, while (2.1) cannot be interpreted as a prior distribution on the parameters, since it is not a probability measure, it still allows for a true posterior distribution on $(\mu_\theta, \mu_\delta, \mu_P, \mu_D, \sigma_c, \sigma_a, \sigma_t, \sigma_\theta, \sigma_\delta, \sigma_P, \sigma_D)$ by the formal use of Bayes' theorem. (See Berger (1985), Bernardo and Smith (1994) or Robert (1994) for more details on improper and noninformative prior distributions.) We do not introduce restrictions on the means μ_δ, μ_P and μ_D, although the intoxication effect is often considered negative and the treatment effects positive, in order to keep the analysis as objective as possible. (Again, such restrictions could be introduced in the analysis at little cost in processing and computing times.)

The posterior distribution of the complete parameter vector is given by

$$\pi((\theta_i,\delta_i,\xi_i),\mu_\theta,\mu_\delta,\mu_P,\mu_D,\sigma_c,\sigma_a,\sigma_t,\sigma_\theta,\sigma_\delta,\sigma_P,\sigma_D|Data) \propto$$

$$\prod_{i=1}^{I}\left\{\prod_{j=1}^{J_i^c}\exp-\{(x_{ij}-\theta_i)^2/2\sigma_c^2\}\prod_{j=1}^{J_i^a}\exp-\{(y_{ij}-\theta_i-\delta_i)^2/2\sigma_a^2\}\right.$$

$$\left.\prod_{j=1}^{J_i^t}\exp-\{(z_{ij}-\theta_i-\delta_i-\xi_i)^2/2\sigma_t^2\}\exp-\{(\theta_i-\mu_\theta)^2/2\sigma_\theta^2\}\exp-\{(\delta_i-\mu_\delta)^2/2\sigma_\delta^2\}\right\}$$

$$\prod_{\ell_i=0}\exp-\{(\xi_i-\mu_P)^2/2\sigma_P^2\}\prod_{\ell_i=1}\exp-\{(\xi_i-\mu_D)^2/2\sigma_D^2\}$$

$$\sigma_c^{-\sum_i J_i^c-1}\sigma_a^{-\sum_i J_i^a-1}\sigma_t^{-\sum_i J_i^t-1}(\sigma_\theta\sigma_\delta)^{-I-1}\sigma_D^{-I_D-1}\sigma_P^{-I_P-1}. \quad \ldots(2.2)$$

It does not integrate into explicit formulas for the marginal distributions of the parameters of interest and therefore cannot lead to closed form expressions for the posterior expectations of these parameters (which are the usual Bayesian estimators). However, if (2.2) can be simulated, these posterior expectations can be approximated by Monte-Carlo techniques, e.g. by using the averages of the simulated sample. In particular, if the fully conditional posterior distributions of the parameters are available for the distribution (2.2), we can take advantage of the *Gibbs sampling* Monte-Carlo method which provides a Markov chain with stationary distribution (2.2) by simulating successively and iteratively from these fully conditional distributions, as described by Gelfand and Smith (1990) and Casella and George (1992).

More precisely, consider the averages

$$\overline{x}_i=\frac{1}{J_i^c}\sum_{i=1}^{J_i^c}x_{ij},\quad \overline{y}_i=\frac{1}{J_i^a}\sum_{i=1}^{J_i^a}y_{ij},\quad \overline{z}_i=\frac{1}{J_i^t}\sum_{i=1}^{J_i^t}z_{ij}$$

and

$$\overline{\theta}=\frac{1}{I}\sum_{i=1}^{I}\theta_i,\qquad \overline{\delta}=\frac{1}{I}\sum_{i=1}^{I}\delta_i.$$

Then the full conditional distributions are indeed explicit and can be simulated

from, since they are given by $(1 \leq i \leq I)$

$$\theta_i \sim \mathcal{N}\left(\frac{\sigma_\theta^{-2}\mu_\theta + J_i^c\sigma_c^{-2}\overline{x}_i + J_i^a\sigma_a^{-2}(\overline{y}_i - \delta_i) + J_i^t\sigma_t^{-2}(\overline{z}_i - \delta_i - \xi_i)}{\sigma_\theta^{-2} + J_i^c\sigma_c^{-2} + J_i^a\sigma_a^{-2} + J_i^t\sigma_t^{-2}}, \right.$$
$$\left. (\sigma_\theta^{-2} + J_i^c\sigma_c^{-2} + J_i^a\sigma_a^{-2} + J_i^t\sigma_t^{-2})^{-1}\right) \quad \ldots(2.3)$$

$$\delta_i \sim \mathcal{N}\left(\frac{\sigma_\delta^{-2}\mu_\delta + J_i^a\sigma_a^{-2}(\overline{y}_i - \theta_i) + J_i^t\sigma_t^{-2}(\overline{z}_i - \theta_i - \xi_i)}{\sigma_\delta^{-2} + J_i^a\sigma_a^{-2} + J_i^t\sigma_t^{-2}}, \right.$$
$$\left. (\sigma_\delta^{-2} + J_i^a\sigma_a^{-2} + J_i^t\sigma_t^{-2})^{-1}\right) \quad \ldots(2.4)$$

$$\xi_i \sim \mathcal{N}\left(\frac{\sigma_D^{-2\ell_i}\sigma_P^{-2(1-\ell_i)}\mu_D^{\ell_i}\mu_P^{1-\ell_i} + J_i^t\sigma_t^{-2}(\overline{z}_i - \theta_i - \delta_i)}{\sigma_D^{-2\ell_i}\sigma_P^{-2(1-\ell_i)} + J_i^t\sigma_t^{-2}}, \right.$$
$$\left. (\sigma_D^{-2\ell_i}\sigma_P^{-2(1-\ell_i)} + J_i^t\sigma_t^{-2})^{-1}\right) \quad \ldots(2.5)$$

$$\mu_\theta \sim \mathcal{N}(\overline{\theta}, \sigma_\theta^2/I), \qquad \mu_\delta \sim \mathcal{N}(\overline{\delta}, \sigma_\delta^2/I), \quad \ldots(2.6)$$

$$\mu_P \sim \mathcal{N}\left(\sum_{\ell_i=0} \xi_i/I_P, \sigma_P^2/I_P\right), \qquad \mu_D \sim \mathcal{N}\left(\sum_{\ell_i=1} \xi_i/I_D, \sigma_D^2/I_D\right), \quad \ldots(2.7)$$

$$\sigma_c^{-2} \sim \mathcal{G}a\left(\sum_i \frac{J_i^c}{2}, \sum_{i,j} \frac{(x_{ij} - \theta_i)^2}{2}\right), \quad \ldots(2.8)$$

$$\sigma_a^{-2} \sim \mathcal{G}a\left(\sum_i \frac{J_i^a}{2}, \sum_{i,j} \frac{(y_{ij} - \theta_i - \delta_i)^2}{2}\right), \quad \ldots(2.9)$$

$$\sigma_t^{-2} \sim \mathcal{G}a\left(\sum_i \frac{J_i^t}{2}, \sum_{i,j} \frac{(z_{ij} - \theta_i - \delta_i - \xi_i)^2}{2}\right), \quad \ldots(2.10)$$

$$\sigma_\theta^{-2} \sim \mathcal{G}a\left(\frac{I}{2}, \sum_i \frac{(\theta_i - \mu_\theta)^2}{2}\right), \quad \sigma_\delta^{-2} \sim \mathcal{G}a\left(\frac{I}{2}, \sum_i \frac{(\delta_i - \mu_\delta)^2}{2}\right), \quad \ldots(2.11)$$

$$\sigma_P^{-2} \sim \mathcal{G}a\left(\frac{I_P}{2}, \sum_{\ell_i=0} \frac{(\xi_i - \mu_P)^2}{2}\right), \quad \sigma_D^{-2} \sim \mathcal{G}a\left(\frac{I_D}{2}, \sum_{\ell_i=1} \frac{(\xi_i - \mu_D)^2}{2}\right). \quad \ldots(2.12)$$

An iterative simulation from the distributions (2.3)–(2.12) using the current values of the other parameters produces a Markov chain with stationary distribution (2.2), as shown in Tierney (1994), and Robert (1994, 1996). In other words, if we start with the values $(\theta_i^{(0)}, \delta_i^{(0)}, \ldots, \sigma_P^{(0)}, \sigma_D^{(0)})$ and simulate $\theta_i^{(1)}$ from (2.3), i.e. from

$$\mathcal{N}\left(\frac{\sigma_\theta^{(0)-2}\mu_\theta^{(0)} + J_i^c\sigma_c^{(0)-2}\overline{x}_i + J_i^a\sigma_a^{(0)-2}(\overline{y}_i^{(0)} - \delta_i^{(0)}) + J_i^t\sigma_t^{(0)-2}(\overline{z}_i^{(0)} - \delta_i^{(0)} - \xi_i^{(0)})}{\sigma_\theta^{(0)-2} + J_i^c\sigma_c^{(0)-2} + J_i^a\sigma_a^{(0)-2} + J_i^t\sigma_t^{(0)-2}}, \right.$$
$$\left. \left(\sigma_\theta^{(0)-2} + J_i^c\sigma_c^{(0)-2} + J_i^a\sigma_a^{(0)-2} + J_i^t\sigma_t^{(0)-2}\right)^{-1}\right)$$

then $\delta_i^{(1)}$ from (2.4) conditionally on $(\theta_i^{(1)}, \xi_i^{(0)}, \ldots, \sigma_D^{(0)})$, up to $\sigma_D^{(1)}$ from (2.12),

i.e.

$$\sigma_D^{(1)^{-2}} \sim \mathcal{G}a\left(I_D/2, \sum_{\ell_i=1}(\xi_i^{(1)} - \mu_D^{(1)})^2/2\right),$$

the Markov chain $(\theta_i^{(n)}, \delta_i^{(n)}, \ldots, \sigma_P^{(n)}, \sigma_D^{(n)})$ is ergodic with stationary distribution (2.2). A major difficulty with Markov Chain Monte-Carlo methods is the control of the convergence of the chain since (2.2) is only the asymptotic distribution of $(\theta_i^{(n)}, \delta_i^{(n)}, \ldots, \sigma_P^{(n)}, \sigma_D^{(n)})$. However, in simple settings such as ours, convergence is quite straightforward and can be assessed by simple graphical tools such as the monitoring of the average of the different quantities of interest.

A common problem in experiments such as the one described here is due to *missing data*, namely, there is often some uncertainty about the statistical treatment of the rats for which one of the three stages (control, intoxication, treatment) is missing. Since such cases still bring some information about the observed stages of the experiment, they should not be discarded. The Bayesian processing of these partial observations is quite straightforward. The posterior distribution (2.2) indeed allows for the cases $J_i^c = 0$, $J_i^a = 0$ or $J_i^t = 0$, as long as one of the three sample sizes is different from 0. If $J_i^t = 0$, ξ_i does not appear in the parameter vector. Similarly, if $J_i^a = J_i^t = 0$, ξ_i and δ_i disappear from the parameter vector. In the conditional distributions used in the Gibbs sampling algorithm, (2.3)–(2.6) and (2.8)–(2.11) then remain the same (including the missing cases for (2.3) or (2.4)); in (2.7) and (2.12), I_P and I_D have to be interpreted as the numbers of rats for which a placebo and a drug treatment have been observed, respectively.

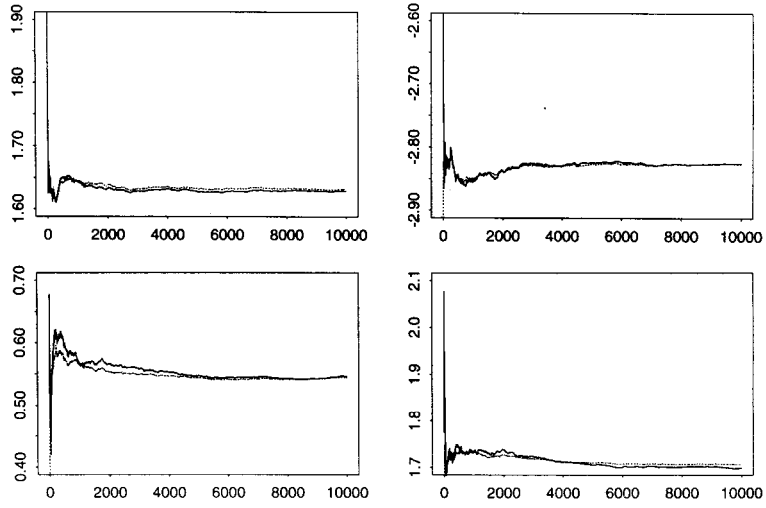

Figure 2.1. Convergence curves for μ_θ *(top left)*, μ_δ *(top right)*, μ_P *(bott. left)* and μ_D. The dotted curves are the partial Rao-Blackwell averages.

As we are able to simulate from (2.2), we can approximate the posterior distributions of the different parameters of the model and, in particular, of the four mean effects, μ_θ, μ_δ, μ_P and μ_D. For the data at hand, we provide in Fig. 2.1 the convergence graphs of the posterior expectations of these four means, the abscissa being the number of iterations k and the ordinate the partial average

$$\frac{1}{k}\sum_{m=1}^{k}\omega_m, \qquad \ldots(2.13)$$

where ω_m represents the m-th simulation of one of the four means above according to (2.3)–(2.12). The dotted curve in these graphs is an alternative approximation called *Rao-Blackwellisation* by reference to the Rao-Blackwell theorem (see Gelfand and Smith (1990), Robert (1994) or Casella and Robert (1996) for details) since it substitutes ω_m for the conditional expectation $\mathbb{E}^\pi[\omega|\tilde{\omega}_m]$, where $\tilde{\omega}_m$ represents the vector of all the parameters except ω. For instance, the running term for μ_θ is then

$$\frac{1}{I}\sum_{i=1}^{I}\theta_i^{(m)}, \qquad \ldots(2.14)$$

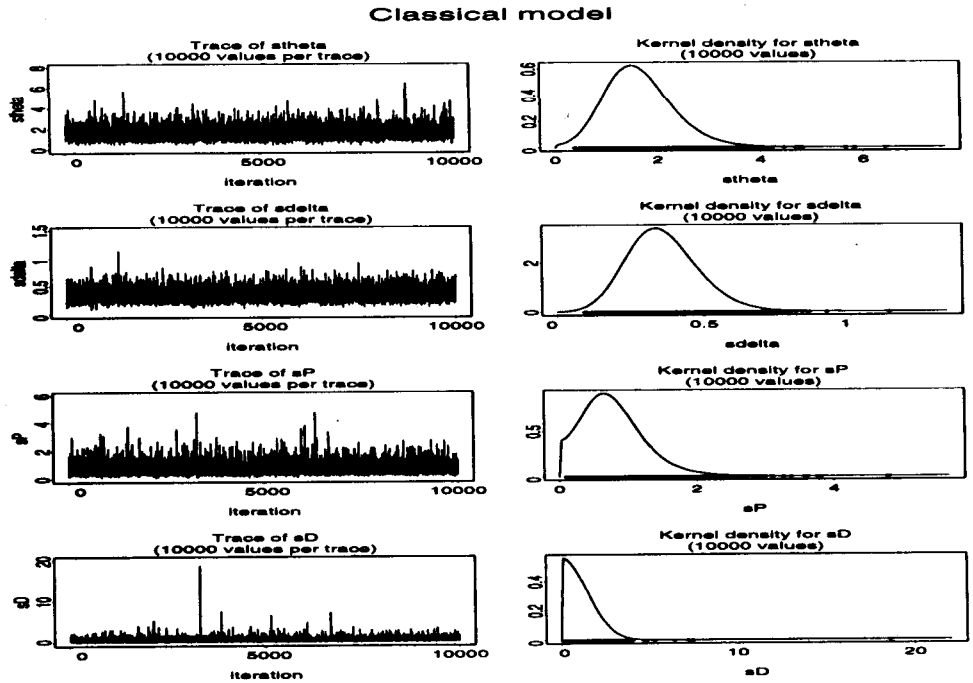

Figure 2.2. Output plots and density estimation for the variances σ_θ (stheta), σ_δ (sdelta), σ_P (sP), σ_D (sD) (obtained by CODA).

where $\theta_i^{(m)}$ denotes the m-th iteration simulation of θ_i. Since both quantities converge to the Bayes estimator $\mathbb{E}^\pi[\omega|Data]$, the similarity of both curves is a partial assessment of convergence, indicating that 10,000 iterations of the Gibbs sampler are enough to insure stability of the averages (2.13) and (2.14).

Other diagnostics like those implemented in CODA of Best, Cowles and Vines (1995) quite similar indications of convergence. Besides the output plot in Figure 2.2, Raftery and Lewis' (1992) give convergence time around 4000, Geweke's (1992) Z-score is under 2 for all parameters but one, and Heidelberger and Welch's (1983) stationarity test passes all parameters.

The mean effects are then estimated by $\mathbb{E}^\pi[\mu_\theta|Data] = 1.63$, $\mathbb{E}^\pi[\mu_\delta|Data] = -2.82$, $\mathbb{E}^\pi[\mu_P|Data] = 0.53$ and $\mathbb{E}^\pi[\mu_D|Data] = 1.39$. Figure 2.2 provides in addition the approximate posterior distributions of the four means, namely the histograms of the Gibbs samples.

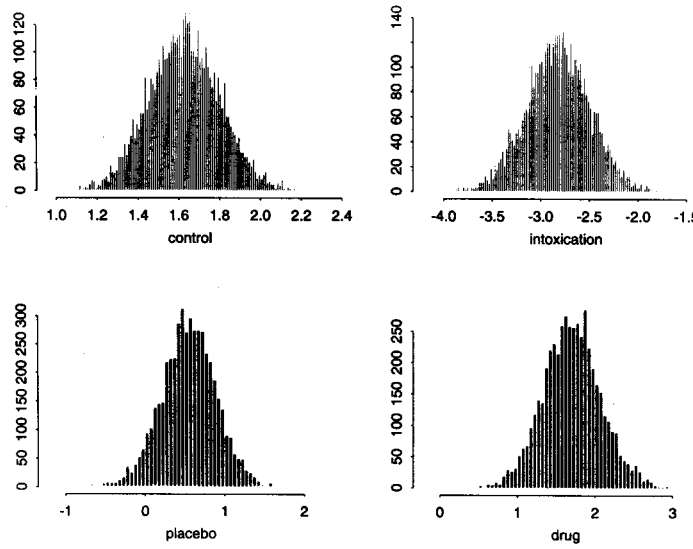

Figure 2.3 Histograms of the Gibbs samples for μ_θ, μ_δ, μ_P and μ_D.

The questions of interest being the true effects of the intoxication and of both treatments, we now focus on the comparisons of μ_δ, μ_D, μ_P, and $\mu_D - \mu_P$ with 0. From a Bayesian perspective, these comparisons can be operated through tests of whether $0 > \mu_\delta$, $\mu_D > 0$, etc. A second solution is to use the above simulations to derive 95% confidence intervals for the differences μ_δ, μ_D, μ_P and $\mu_D - \mu_P$, and then to check whether 0 belongs to these intervals. Table 2.1 provides the posterior probabilities that the effects are significant as well as these four posterior confidence intervals for the data at hand. This allows us to conclude that intoxication, drug and placebo effects are significant, although at a lesser degree for the placebo, but also that the drug effect significantly differs from the placebo effect, thus that the drug is contributing to an improvement of

the gaze stability of the rats if not to a complete restoration of the normal level, since $P(\mu_\delta + \mu_D > 0|Data) = 0.017$.

These results mainly demonstrate the antagonistic effects of successive high and very small concentrations of a picrotoxin solution on gaze stability by successive impairment and improvement of this stability, and also indicates a possible reduction of the concentration of the toxic substance in the brain due to the placebo, which would then explain for the significant placebo effect.

Table 2.1 POSTERIOR PROBABILITIES OF SIGNIFICANCE AND CONFIDENCE INTERVALS FOR THE MEAN EFFECTS.

	μ_δ	μ_D	μ_P	$\mu_D - \mu_P$
Probability	1.00	0.9998	0.94	0.985
Confidence	[-3.48,-2.17]	[0.94,2.50]	[-0.17,1.24]	[0.14,2.20]

3. Modelling Heterogeneity Through a Mixture Structure

The choice of the normal distributions in (1.1)–(1.3) is quite arbitrary and it may well happen that these distributions do not correctly fit the data at hand. This is particularly the case with heterogeneous populations where individuals responses can be separated in several categories. In our example, the rats' reaction to the intoxication did indeed appear to be heterogeneous, the histogram of the intoxication effect y_{ij} being quite widespread. Moreover, the frequent inter-individual variability to intoxication (whether it is alcohol, benzene or picrotoxin) could be related to different levels of the speed of the metabolic desintoxication mechanism of genetic origin. For instance, curare intoxication can be counteracted by either slow-acetyl-prone or fast-acetyl-prone enzymes (see Larrey, 1986). A logical alternative to the normal distribution is then to consider local normality, i.e. normality for the homogeneous components of the population; this is equivalent to propose a mixture structure with several normal components for the δ_i's. See West (1992) and Robert (1996) for details on mixture modelling. Additional mixture structures could be proposed at the control and treatment stages but, in our particular setup, the intoxication mixture appears to be sufficient to capture the heterogeneity of the population. We consider here only a two components mixture structure as in (1.4).

An interesting feature of this analysis, besides providing a more accurate representation of the modeled phenomenon and allowing to assess more thoroughly the effect of the drug, is that the mixture modelling takes place at a latent (i.e. unobserved) level of the model since we only observe the y_i. Due to the flexibility allowed by the Bayesian approach and the Gibbs sampling simulation techniques, inference on this extension is still feasible.

For stability reasons exposed in Mengersen and Robert (1996), we substitute to (1.4) the parameterization

$$\delta_i \sim p\mathcal{N}(\mu_\delta, \tau_\delta^2) + (1-p)\mathcal{N}(\mu_\delta + \tau_\delta\epsilon_\delta, \tau_\delta^2\sigma_\delta^2), \qquad \ldots(3.1)$$

which links both components by common location and scale parameters μ_δ and τ_δ. (In the setup of the experiment, the second component represents the rats with a reaction to the drug different from the main group.) Furthermore, we can impose $\sigma_\delta < 1$ by identifiability arguments. The parameterization of (3.1) also allows for a noninformative prior, given by

$$\pi(p, \mu_\delta, \tau_\delta, \epsilon_\delta, \sigma_\delta) \propto \frac{1}{\tau_\delta}\mathbb{I}_{[0,1]}(p)\mathbb{I}_{[0,1]}(\sigma_\delta)\exp(-\epsilon_\delta^2/2),$$

the normal distribution $\mathcal{N}(0,1)$ on ϵ_δ being rather insensitive to a change in its variance (see Mengersen and Robert, 1996). This new distribution on the δ_i's obviously complicates the posterior distribution and, like (2.2), does not lead to explicit expressions. However, the posterior distribution can again be simulated via Gibbs sampling techniques. The conditional distributions of the θ_i's, the ξ_i's, μ_θ, μ_P, μ_D, σ_c, σ_a, σ_t, σ_θ, σ_P and σ_D in (2.3)–(2.12) remain the same, while the δ_i's are now distributed as

$$\begin{aligned}
\pi(\delta_i) \quad &\propto p(\tau_\delta^{-2} + J_i^a\sigma_a^{-2} + J_i^t\sigma_t^{-2})^{-1/2} \\
&\quad \exp\left(-\tfrac{1}{2}(\tau_\delta^{-2} + J_i^a\sigma_a^{-2} + J_i^t\sigma_t^{-2})^{-1}\left\{J_i^a\sigma_a^{-2}\tau_\delta^{-2}(\mu_\delta - \overline{y}_i + \theta_i)^2\right.\right. \\
&\qquad \left.\left. + J_i^t\sigma_t^{-2}\tau_\delta^{-2}(\mu_\delta - \overline{z}_i + \theta_i + \xi_i)^2 + J_i^a\sigma_a^{-2}J_i^t\sigma_t^{-2}(\overline{y}_i - \overline{z}_i + \xi_i)^2\right\}\right) \\
&\quad \mathcal{N}\left(\tfrac{\tau_\delta^{-2}\mu_\delta + J_i^a\sigma_a^{-2}(\overline{y}_i - \theta_i) + J_i^t\sigma_t^{-2}(\overline{z}_i - \theta_i - \xi_i)}{\tau_\delta^{-2} + J_i^a\sigma_a^{-2} + J_i^t\sigma_t^{-2}}, (\tau_\delta^{-2} + J_i^a\sigma_a^{-2} + J_i^t\sigma_t^{-2})^{-1}\right) \\
&+ (1-p)(\tau_\delta^{-2}\sigma_\delta^{-2} + J_i^a\sigma_a^{-2} + J_i^t\sigma_t^{-2})^{-1/2} \\
&\quad \exp\left(-\tfrac{1}{2}(\tau_\delta^{-2}\sigma_\delta^{-2} + J_i^a\sigma_a^{-2} + J_i^t\sigma_t^{-2})^{-1}\left\{J_i^a\sigma_a^{-2}\tau_\delta^{-2}\sigma_\delta^{-2}(\mu_\delta + \tau_\delta\epsilon_\delta - \overline{y}_i + \theta_i)^2\right.\right. \\
&\qquad \left.\left. + J_i^t\sigma_t^{-2}\tau_\delta^{-2}\sigma_\delta^{-2}(\mu_\delta + \tau_\delta\epsilon_\delta - \overline{z}_i + \theta_i + \xi_i)^2 + J_i^a\sigma_a^{-2}J_i^t\sigma_t^{-2}(\overline{y}_i - \overline{z}_i + \xi_i)^2\right\}\right) \\
&\quad \mathcal{N}\left(\tfrac{\tau_\delta^{-2}\sigma_\delta^{-2}(\mu_\delta + \tau_\delta\epsilon_\delta) + J_i^a\sigma_a^{-2}(\overline{y}_i - \theta_i) + J_i^t\sigma_t^{-2}(\overline{z}_i - \theta_i - \xi_i)}{\tau_\delta^{-2}\sigma_\delta^{-2} + J_i^a\sigma_a^{-2} + J_i^t\sigma_t^{-2}},\right. \\
&\qquad \left.(\tau_\delta^{-2}\sigma_\delta^{-2} + J_i^a\sigma_a^{-2} + J_i^t\sigma_t^{-2})^{-1}\right).
\end{aligned}$$

The posterior (conditional) distribution of δ_i is therefore a mixture of two normal distributions with weights proportional to the two factors in front of the normal

distributions and its simulation is straightforward. This mixture actually reduces to the prior distribution (3.1) if there is no observation at both the intoxication and the treatment levels for the i-th rat, although δ_i is not part of the parameter vector in this case.

The other modification to the conditional distributions (2.3)–(2.12) is the addition of the conditional distributions of the hyperparameters $p, \mu_\delta, \tau_\delta, \epsilon_\delta$ and σ_δ. However, the hierarchical structure of the model induces very straightforward changes since the posterior distribution of these parameters is only based upon the simulated sample of the δ_i's. Therefore, we can reproduce the estimation technique derived in Mengersen and Robert (1996). ¿From a Gibbs sampling perspective, it is enough to say that the iteration step for the simulation of $(p, \mu_\delta, \tau_\delta, \epsilon_\delta, \sigma_\delta)$ is made of two parts, a data completion stage and a simulation stage, as follows:

1. Separate the sample of the δ_i's in two subsamples corresponding to both components, δ_i being allocated to the first component with probability

$$\frac{p\exp\{-(\delta_i-\mu_\delta)^2/2\tau_\delta^2\}}{p\exp\{-(\delta_i-\mu_\delta)^2/2\tau_\delta^2\}+(1-p)\sigma_\delta^{-1}\exp\{-(\delta_i-\mu_\delta-\tau_\delta\epsilon_\delta)^2/2\tau_\delta^2\sigma_\delta^2\}}.$$

Compute the averages and sums of squares for each subsample, $\bar{\delta}_1$, $s_{\delta 1}^2$ and $\bar{\delta}_2$, $s_{\delta 2}^2$, respectively, I_1 and I_2 denoting the sizes of each subsample—with $I_2 = I - I_1$ if there is no missing observation.

2. Simulate from

$$\begin{aligned}
p &\sim \mathcal{B}e(I_1+1, I_2+1);\\
\epsilon_\delta &\sim \mathcal{N}\left(\tfrac{(\bar{\delta}_2-\mu_\delta)}{\tau_\delta}\left[1+\tfrac{\sigma_\delta^2}{I_2}\right]^{-1}, \tfrac{\sigma_\delta^2}{I_2+\sigma_\delta^2}\right);\\
\mu_\delta &\sim \mathcal{N}\left(\tfrac{I_1\bar{\delta}_1+\sigma_\delta^{-2}I_2(\bar{\delta}_2-\tau_\delta\epsilon_\delta)}{I_1+I_2\sigma_\delta^{-2}}, \tfrac{\tau_\delta^2\sigma_\delta^2}{I_1\sigma_\delta^2+I_2}\right);\\
\tau_\delta^{-2} &\sim \mathcal{G}a\left(\tfrac{I_1+I_2}{2}, \tfrac{1}{2}(s_{\delta 1}^2+I_1(\bar{\delta}_1-\mu_\delta)^2)+\tfrac{s_{\delta 2}^2}{2\sigma_\delta^2}+\tfrac{I_2(\bar{\delta}_2-\mu_\delta)^2}{2(\sigma_\delta^2+I_2)}\right);\\
\sigma_\delta^{-2} &\sim \mathcal{G}a\left(\tfrac{I_2-1}{2}, \tfrac{s_{\delta 2}^2+I_2(\bar{\delta}_2-\mu_\delta-\tau_\delta\epsilon_\delta)^2}{2\tau_\delta^2}\right)\mathbb{I}_{\sigma_\delta^{-2}>1};
\end{aligned}$$

the last distribution denoting a *truncated gamma distribution.* When $I_2 = 0$, this distribution reduces to the prior uniform $\mathcal{U}_{[0,1]}$ distribution on σ and, otherwise, it can be simulated using an accept-reject algorithm described in Philippe (1997).

These additional steps slightly complicate the implementation of the Gibbs sampler but mainly slow down convergence, since convergence to the stationary distribution is not as immediate in the case of the mixture distributions. In particular, as described in Mengersen and Robert (1996), the identifiability restriction $\sigma_\delta < 1$ may cause the algorithm to be trapped in some regions of the parameter space and it is preferable to run the algorithm first *without* the

restriction on σ_δ in order to create good starting conditions.

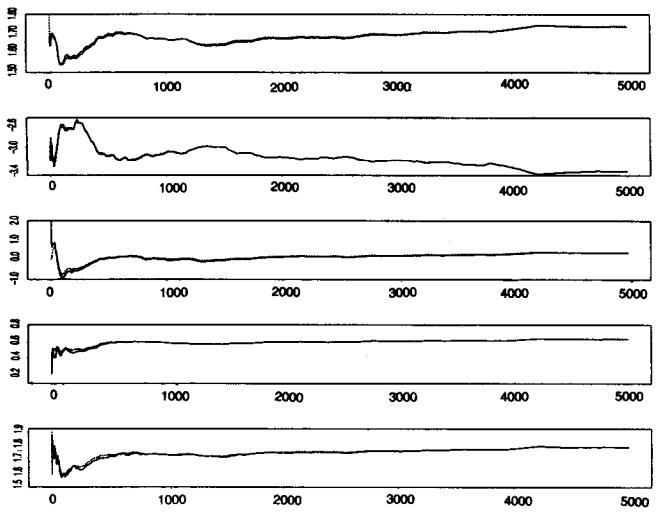

Figure 3.1. Convergence curves for μ_θ, μ_δ, ϵ_δ, μ_P and μ_D. The dotted curves are the partial Rao-Blackwell averages.

The unconstrained prior distribution on σ_δ is then

$$\sigma_\delta \sim \frac{1}{2}\mathcal{U}_{[0,1]} + \frac{1}{2}\mathcal{P}a(2,1),$$

leading to a conditional posterior distribution which can be easily generated by an accept-reject algorithm (see Mengersen and Robert, 1996).

Again, CODA diagnostics can be invoked in this case to assess convergence, besides the agreement between regular and Rao-Blackwell averages as in Figure 3.1. The number of iterations is higher than in the classical case, but 50,000 iterations lead to positive signals from most criteria, except for Raftery and Lewis' (1992) diagnostic which requires a much higher number of iterations. (The autocorrelations are also larger in this case.) The histogram approximations of the posterior distributions of μ_θ, μ_δ, $\mu_\theta + \epsilon_\delta\tau_\delta$, μ_P and μ_D are shown in Figure 3.2.

The analysis of the sample produced by the modified Gibbs sampling is akin to the one presented in Section 2. The estimators of the parameters of the mixture (3.1) are

$$p^\pi = 0.49, \quad \mu_\delta^\pi = -3.74, \quad \epsilon_\delta^\pi = 0.52, \quad \tau_\delta^\pi = 1.73, \quad \sigma_\delta^\pi = 0.72;$$

this shows that the mixture modelling is indeed appropriate for the dataset since the estimated mixture does not collapse into a single component. The other mean effects are $\mathbb{E}^\pi[\mu_\theta|Data] = 1.79$, $\mathbb{E}^\pi[\mu_P|Data] = 0.66$ and $\mathbb{E}^\pi[\mu_D|Data] = 1.80$.

The overall mean for the intoxication effect is then $\mathbb{E}^{\pi}[\mu_\delta + (1-p)\epsilon_\delta\tau_\delta|Data] = -3.28$ and the four mean effect estimators are thus slightly different from their counterparts obtained in Section 2. Note that the Gibbs sampler provides in addition the posterior probabilities for a given rat to belong to one of the two homogeneous populations. For instance, rat 1 has a posterior probability of 0.75 to be in the first component of (3.1), while this probability is 0.39 for rat 21.

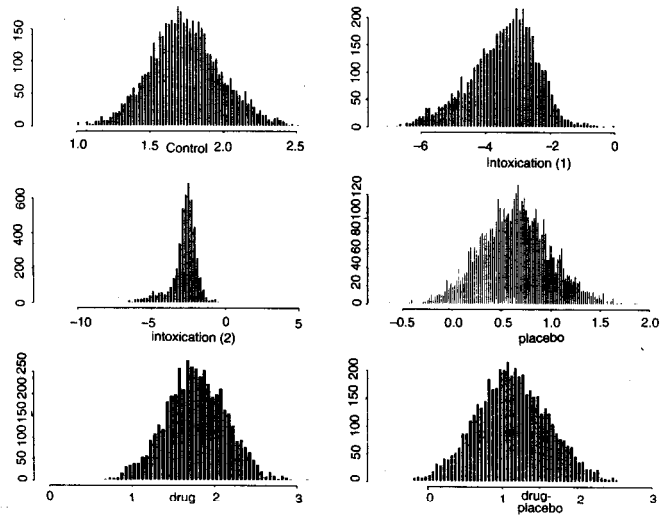

Figure 3.2. Histograms for the samples of μ_θ, μ_δ, $\mu_\delta + \tau_\delta\epsilon_\delta$, μ_P, μ_D and $\mu_D - \mu_P$ (from left to right and top to bottom).

The effect of the intoxication can similarly be compared to the control level through 95% confidence intervals for μ_δ and $\mu_\delta + \tau_\delta\epsilon_\delta$ and the effect of the treatment through the 95% confidence intervals for μ_D, μ_P and $\mu_D - \mu_P$, given in Table 3.1.

Table 3.1 POSTERIOR PROBABILITIES OF SIGNIFICANCE AND CONFIDENCE INTERVALS IN THE MIXTURE CASE FOR THE MEAN EFFECTS.

	μ_δ	$\mu_\delta + \tau_\delta\epsilon_\delta$	μ_D	μ_P	$\mu_D - \mu_P$
Prob.	0.997	0.988	1.0	0.936	0.988
Conf.	[-5.87,-1.51]	[-5.82,-0.94]	[0.985,2.57]	[0.095,1.41]	[0.14,2.17]

When compared with the analysis of Section 2, this modelling thus exhibits both a stronger intoxication effect and a larger drug enhanced recovery effect, while the placebo effect is numerically stable and the intoxication plus drug effect is almost identical to the normal analysis. This reinforces the previous analysis

in favor of the actual improvement brought by the drug, as opposed to the placebo effect. Obviously this setup extends to the cases commonly encountered in pathology, therapy and clinical trials, where intoxication is replaced by a disease whose strength may vary in a few identified but not recognizable classes. In settings where the strength class may change within a same individual along time as in long stepwise diseases, the mixture modelling should be replaced by an hidden Markov chain structure (see Robert *et al*, 1993).

Acknowledgments. The authors thank the Laboratoires Boiron, Mont-Saint-Aignan, and the Région de Haute-Normandie for their financial support of the biological experiments.

References

BERGER, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York.

BERNARDO, J.M. AND SMITH, A.F.M. (1994). *Bayesian Theory*, John Wiley, New York.

BEST, N.G., COWLES, M.K. AND VINES, K. (1995). 'CODA: Convergence Diagnosis and Output Analysis software for Gibbs sampling output, Version 0.30'. *Tech. Report*, MRC Biostatistics Unit, Univ. of Cambridge.

CASELLA, G. AND GEORGE, E.I. (1992). An introduction to Gibbs sampling. *The American Statistician* **46**, 167–174.

CASELLA, G. AND ROBERT, C.P. (1996). Rao-Blackwellisation of sampling schemes. *Biometrika* **83**(1), 81–94.

GELFAND, A.E. AND SMITH, A.F.M. (1990). Sampling based approaches to calculating marginal densities. *Jour. Amer. Statist. Assoc.* **85**, 398–409.

GEWEKE, J. (1993). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion). In *Bayesian Statistics 4*, J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (Eds.). Oxford University Press, Oxford, 169–193.

HEIDELBERGER, P. AND WELCH, P.D. (1983). A spectral method for confidence interval generation and run-length control in simulation. *Commun. ACM* **24**, 233–245.

LARREY, D. (1986). 'Polymorphisme du métabolisme hépatique des médicaments'. *Médecine–Sciences II* **7**, 364.

MENGERSEN, K.L. AND ROBERT, C.P. (1996). Testing for mixtures: a Bayesian entropic approach. In *Bayesian Statistics 5* (J.O. Berger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith, eds.) Oxford University Press, 255–276.

PHILIPPE, A. (1997). Simulation of right- and left-truncated Gamma distributions by mixtures. *Statistics and Computing*, **7**, 173–181.

RAFTERY, A.E. AND LEWIS, S. (1992). How many iterations in the Gibbs sampler? In *Bayesian Statistics 4* (J.O. Berger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith, eds.) Oxford University Press, 763–773.

REBER, A., SARRAU, J.M., CARNET, J., MAGNIN, M. AND STELZ, T. (1991). Horizontal optokinetic nystagmus in unilaterally enucleated pigmented rats: role of the pretectal commissural fibers. *J. Comp. Neurol.* **313**, 604–612.

ROBERT, C.P. (1994). *The Bayesian Choice: A Decision-Theoretic Motivation*, Springer-Verlag, New York.

– – – – (1996). Inference in mixture models. In *Practical MCMC* (W. Gilks, S. Richardson and D. Spiegelhalter, eds.). Chapman and Hall, London, 441–464.

ROBERT, C.P., DIEBOLT, J. AND CELEUX, G. (1993). Bayesian estimation of Hidden Markov Models: a stochastic implementation. *Statistics and Probability Letters* **16**(1), 77–83.

SARRAU, J.M., REBER, A., MAZARI, B. AND CARNET, J. (1989). Capteur angulaire de mouvements occulaires. In *Conférence Internationale sur les Capteurs*, Apist et Ciame (eds.) Paris, 55–62.

TANNER, M. (1993). *Tools for Statistical Inference*, Springer-Verlag, New York.

TIERNEY, L. (1994). Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.* **22**, 1701–1786.

YE, K. AND BERGER, J.O. (1991). Noninformative priors for inferences in exponential regression models. *Biometrika* **78**, 645–656.

WEST, M. (1992). Modelling with mixtures. In *Bayesian Statistics 4* (J.O. Berger, J.M. Bernardo, A.P. Dawid and A.F.M. Smith, eds.) Oxford University Press,, 503–525.

CHRISTIAN P. ROBERT
LABORATOIRE DE STATISTIQUE
CREST, INSEE
75675 PARIS CEDEX 14, FRANCE
e-mail : robert@ensae.fr

ANNIE REBER
LABORATOIRE DE NEUROSCIENCES ET
ENVIRONNEMENT, UNIVERSITÉ DE ROUEN,
76821 MONT SAINT AIGNAN CEDEX
FRANCE