10.2 Devise and implement a simulation algorithm for the Ising model of Example 10.2.

10.3 In the setup of Example 10.17:

(a) Evaluate the numerical value of the boundary constant ϵ derived from (10.5), given the data in Table 10.1.

(b) Establish the lower bound in (10.5). (*Hint*: Replace $\sum \lambda_i$ by 0 in the integrand [but not in the exponent] and do the integration.)

10.4 For the data of Table 10.1

(a) Estimate $\lambda_1, \ldots, \lambda_{10}$ using the model of Example 10.17.

(b) Estimate a and b in the loglinear model $\log \lambda = a + bt$, using a Gibbs sampler and the techniques of Example 2.26.

(c) Compare the results of the analyses.

10.5 (Besag 1994) Consider a distribution f on a finite state-space $\mathcal{X}$ of dimension k and conditional distributions $f_1, \ldots, f_k$ such that for every $(x, y) \in \mathcal{X}^2$, there exist an integer m and a sequence $x_0 = x, \ldots, x_m = y$ where x_i and x_{i+1} only differ in a single component and $f(x_i) > 0$.

(a) Show that this condition extends Hammersley–Clifford Theorem (Theorem 10.5) by deriving $f(y)$ as

$$f(y) = f(x) \prod_{i=0}^{m-1} \frac{f(x_{i+1})}{f(x_i)} .$$

(b) Deduce that irreducibility and ergodicity holds for the associated Gibbs sampler.

(c) Show that the same condition on a continuous state-space is not sufficient to ensure ergodicity of the Gibbs sampler. (*Hint:* See Example 10.7.)

10.6 Compare the usual demarginalization of the Student's t distribution discussed in Example 10.4 with an alternative using a slice sampler, by computing the empirical cdf at several points of interest for both approaches and the same number of simulations. (*Hint*: Use two uniform dummy variables.)

10.7 A bound similar to the one in Example 10.17, established in Problem 10.3, can be obtained for the kernel of Example 10.4; that is, show that

$$\begin{aligned} K(\theta, \theta') = & \int_0^\infty \sqrt{\frac{1+\eta}{2\pi}} \exp\left\{ -\left(\theta' - \frac{\theta_0 \eta}{1+\eta}\right)^2 \frac{1+\eta}{2} \right\} \left(\frac{1 + (\theta - \theta_0)^2}{2} \right)^\nu \\ & \times \frac{\eta^{\nu-1}}{\Gamma(\nu)} \exp\left\{ \frac{-\eta}{2} \left(1 + (\theta - \theta_0)^2\right) \right\} d\eta \\ \geq & \left[1 + (\theta' - \theta_0)^2\right]^{-\nu} \frac{e^{-(\theta')^2/2}}{\sqrt{2\pi}} . \end{aligned}$$

(*Hint*: Establish that

$$\exp\left\{ -\left(\theta' - \frac{\theta_0 \eta}{1+\eta}\right)^2 \frac{1+\eta}{2} \right\} \geq \exp\left\{ -\frac{1}{2}(\theta')^2 - \frac{\eta}{2}(\theta'^2 - 2\theta'\theta_0 + \theta_0^2) \right\}$$

and integrate.)

10.8 (Hobert et al. 1997) Consider a distribution f on $\mathbb{R}^k$ to be simulated by Gibbs sampling. A one-to-one transform Ψ on $\mathbb{R}^k$ is called a parameterization. The convergence of the Gibbs sampler can be jeopardized by a bad choice of parameterization.

(a) Considering that a Gibbs sampler can be formally associated with the full conditional distributions for every choice of a parameterization, show that there always exist parameterizations such that the Gibbs sampler fails to converge.
(b) The fact that the convergence depends on the parameterization vanishes if (i) the support of f is arcwise connected, that for every $(x, y) \in (\text{supp}(f))^2$, there exists a continuous function φ on $[0, 1]$ with $\varphi(0) = x$, $\varphi(1) = y$, and $\varphi([0, 1]) \subset \text{supp}(f)$ and (ii) the parameterizations are restricted to be continuous functions.
(c) Show that condition (i) in (b) is necessary for the above property to hold. (*Hint:* See Example 10.7.)
(d) Show that a rotation of $\pi/4$ of the coordinate axes eliminates the irreducibility problem in Example 10.7.

10.9 In the setup of Example 5.13, show that the Gibbs sampling simulation of the ordered normal means θ_{ij} can either be done in $I \times J$ conditional steps or in only two conditional steps. Conduct an experiment to compare both approaches.

10.10 The clinical mastitis data described in Example 10.15 is the number of cases observed in 127 herds of dairy cattle (the herds are adjusted for size). The data are given in Table 10.3.

0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
1	1	1	1	2	2	2	2	2	2	2	2	3	3	3
3	3	3	3	3	3	4	4	4	4	4	4	4	4	5
5	5	5	5	5	5	5	6	6	6	6	6	6	6	6
6	7	7	7	7	7	7	8	8	8	8	8	9	9	9
10	10	10	10	11	11	11	11	11	11	11	12	12	12	12
13	13	13	13	13	14	14	15	16	16	16	16	17	17	17
18	18	18	19	19	19	19	20	20	21	21	22	22	22	22
23	25	25	25	25	25	25								

Table 10.3. Occurrences of clinical mastitis in 127 herds of dairy cattle. (*Source:* Schukken et al. 1991.)

(a) Verify the conditional distributions for λ_i and β_i.
(b) For these data, implement the Gibbs sampler and plot the posterior density $\pi(\lambda_1|\mathbf{x}, \alpha)$. (Use the values $\alpha = .1, a = b = 1$.)
(c) Make histograms and monitor the convergence of λ_5, λ_{15}, and β_{15}.
(d) Investigate the sensitivity of your answer to the specification of α, a and b.

10.11 (a) For the Gibbs sampler of [A.40], if the Markov chain $(Y^{(t)})$ satisfies the positivity condition (see Definition 9.4), show that the conditional distributions $g(y_i|y_1, y_2, \ldots, y_{i-1}, y_{i+1}, \ldots, y_p)$ will not reduce the range of possible values of Y_i when compared with g.
(b) Prove Theorem 10.8.

10.12 Show that the Gibbs sampler of Algorithm [A.41] is reversible. (Consider a generalization of the proof of Algorithm [A.36].)

10.13 For the hybrid algorithm of [A.43], show that the resulting chain is a Markov chain and verify its stationary distribution. Is the chain reversible?

10.14 Show that the model (10.8) is not identifiable in (β, σ). (*Hint:* Show that it only depends on the ratio β/σ.)

10.15 If $K(x, x')$ is a Markov transition kernel with stationary distribution g, show that the Metropolis–Hastings algorithm where, at iteration t, the proposed value $y_t \sim K(x^{(t)}, y)$ is accepted with probability

$$\frac{f(y_t)}{f(x^{(t)})} \frac{g(x^{(t)})}{g(y_t)} \wedge 1 \,,$$

provides a valid MCMC algorithm for the stationary distribution f. (*Hint:* Use detailed balance.)

10.16 For the algorithm of Example 7.13, an obvious candidate to simulate $\pi(a|\mathbf{x}, \mathbf{y}, b)$ and $\pi(b|\mathbf{x}, \mathbf{y}, a)$ is the ARS algorithm of Section 2.4.2.

(a) Show that

$$\log \pi(a|\mathbf{x}, \mathbf{y}, b) = \sum_i y_i a - \sum_i y_i \log\left(1 + e^{a+bx_i}\right) - \sum_i \frac{e^{a+bx_i}}{1+e^{a+bx_i}} - \frac{a^2}{2\sigma^2},$$

and, thus,

$$\frac{\partial}{\partial a} \log \pi(a|\mathbf{x}, \mathbf{y}, b) = \sum_i y_i - \sum_i y_i \frac{e^{a+bx_i}}{1+e^{a+bx_i}} - \sum_i \frac{e^{a+bx_i}}{(1+e^{a+bx_i})^2} - \frac{a}{\sigma^2}$$

and

$$\begin{aligned} \frac{\partial^2}{\partial a^2} \log \pi(a|\mathbf{x}, \mathbf{y}, b) &= -\sum_i \frac{y_i e^{a+bx_i}}{(1+e^{a+bx_i})} - \sum_i \frac{e^{a+bx_i} - e^{2(a+bx_i)}}{(1+e^{a+bx_i})^3} - \sigma^{-2} \\ &= -\sum_i \frac{e^{a+bx_i}}{(1+e^{a+bx_i})^3} \left\{(1+y_i) - (1-y_i)e^{a+bx_i}\right\} - \sigma^{-2}. \end{aligned}$$

Argue that this last expression is not always negative, so the ARS algorithm cannot be applied.

(b) Show that

$$\begin{aligned} &\frac{\partial^2}{\partial b^2} \log \pi(b|\mathbf{x}, \mathbf{y}, a) \\ &\quad = -\sum_i \frac{e^{a+bx_i}}{(1+e^{a+bx_i})^3} \left\{(1+y_i) - (1-y_i)e^{a+bx_i}\right\} x_i^2 - \tau^{-2} \end{aligned}$$

and deduce there is also no log-concavity in the b direction.

(c) Even though distributions are not log-concave, they can be simulated with the ARMS algorithm [A.28]. Give details on how to do this.

10.17 Show that the Riemann sum method of Section 4.3 can be used in conjunction with Rao–Blackwellization to cover multidimensional settings. (*Note:* See Philippe and Robert 1998a for details.)

10.18 The *tobit* model is used in econometrics (see Tobin 1958). It is based on a transform of a normal variable $y_i^* \sim \mathcal{N}(x_i^t \beta, \sigma^2)$ $(i = 1, \ldots, n)$ by truncation

$$y_i = \max(y_i^*, 0).$$

Show that the following algorithm provides a valid approximation of the posterior distribution of (β, σ^2).

1. Simulate $y_i^* \sim \mathcal{N}_-(x_i^t\beta, \sigma^2, 0)$ if $y_i = 0$.
2. Simulate $(\beta, \sigma) \sim \pi(\beta, \sigma | y^*, x)$ with $[A_{31}]$

$$\pi(\beta, \sigma | y^*, x) \propto \sigma^{-n} \exp\left\{ -\sum_i (y_i^* - x_i^t\beta)^2 / 2\sigma^2 \right\} \pi(\beta, \sigma) .$$

10.19 Let (X_n) be a reversible stationary Markov chain (see Definition 6.44) with $\mathbb{E}[X_i] = 0$.

(a) Show that the distribution of $X_k | X_1$ is the same as the distribution of $X_1 | X_k$.

(b) Show that the covariance between alternate random variables is positive. More precisely, show that $\text{cov}(X_0, X_{2\nu}) = \mathbb{E}[\mathbb{E}(X_0|X_\nu)^2] > 0$.

(c) Show that the covariance of alternate random variables is decreasing; that is, show that $\text{cov}(X_0, X_{2\nu}) \geq \text{cov}(X_0, X_{2(\nu+1)})$. (*Hint*: Use the fact that

$$\mathbb{E}[\mathbb{E}(X_0|X_\nu)^2] = \text{var}[\mathbb{E}(X_0|X_\nu)] \geq \text{var}[\mathbb{E}\{\mathbb{E}(X_0|X_\nu)|X_{\nu+1}\}],$$

and show that this latter quantity is $\text{cov}(X_0, X_{2(\nu+1)})$.)

10.20 Show that a Gibbs sampling kernel with more than two full conditional steps cannot produce interleaving chains.

10.21 In the setup of Example 10.2:

(a) Consider a grid in $\mathbb{R}^2$ with the simple nearest-neighbor relation $\mathcal{N}$ (that is, $(i, j) \in \mathcal{N}$ if and only if $\min(|i_1 - j_1|, |i_2 - j_2|) = 1$). Show that a Gibbs sampling algorithm with only two conditional steps can be implemented in this case.

(b) Implement the Metropolization scheme of Liu (1995) discussed in Section 10.3.3 and compare the results with those of part (a).

10.22 (Gelfand et al. 1995) In the setup of Example 10.25, show that the original parameterization leads to the following correlations:

$$\rho_{\mu,\alpha_i} = \left(1 + \frac{I\sigma_y^2}{J\sigma_\alpha^2}\right)^{-1/2}, \qquad \rho_{\alpha_i,\alpha_j} = \left(1 + \frac{I\sigma_y^2}{J\sigma_\alpha^2}\right)^{-1},$$

for $i \neq j$.

10.23 (Continuation of Problem 10.22) For the hierarchical parameterization, show that the correlations are

$$\rho_{\mu,\eta_i} = \left(1 + \frac{IJ\sigma_\alpha^2}{\sigma_y^2}\right)^{-1/2}, \qquad \rho_{\eta_i,\eta_j} = \left(1 + \frac{IJ\sigma_\alpha^2}{\sigma_y^2}\right)^{-1},$$

for $i \neq j$.

10.24 In Example 10.29, we noted that a pair of conditional densities is *functionally compatible* if the ratio $f_1(x|y)/f_2(y|x) = h_1(x)/h_2(y)$, for some functions h_1 and h_2. This is a necessary condition for a joint density to exist, but not a sufficient condition. If such a joint density does exist, the pair f_1 and f_2 would be *compatible*.

(a) Formulate the definitions of compatible and functionally compatible for a set of densities $f_1, \ldots, f_m$.

(b) Show that if $f_1, \ldots, f_m$ are the full conditionals from a hierarchical model, they are functionally compatible.

(c) Prove the following theorem, due to Hobert and Casella (1998), which shows there cannot be any stationary probability distribution for the chain to converge to unless the densities are compatible.

Theorem 10.32. *Let $f_1, \ldots, f_m$ be a set of functionally compatible conditional densities on which a Gibbs sampler is based. The resulting Markov chain is positive recurrent if and only if $f_1, \ldots, f_m$ are compatible.*

(*Note:* Compatibility of a set of densities was investigated by Besag 1974, Arnold and Press 1989, and Gelman and Speed 1993.)

10.25 For the situation of Example 10.31,

(a) Show that the full "posterior distribution" of $(\beta, \sigma^2, \tau^2)$ is

$$\pi(\beta, \sigma^2, \tau^2|y) \propto \sigma^{-2-I}\, \tau^{-2-IJ} \exp\left\{-\frac{1}{2\tau^2}\sum_{i,j}(y_{ij}-\bar{y}_i)^2\right\}$$
$$\times \exp\left\{-\frac{J\,\sum_i(\bar{y}_i-\beta)^2}{2(\tau^2+J\sigma^2)}\right\}(J\tau^{-2}+\sigma^{-2})^{-I/2}\,.$$

(*Hint:* Integrate the (unobservable) random effects u_i.)

(b) Integrate out β and show that the marginal posterior density of $(\sigma^2, \tau^2|)$ is

$$\pi(\sigma^2, \tau^2|y) \propto \frac{\sigma^{-2-I}\,\tau^{-2-IJ}}{(J\tau^{-2}+\sigma^{-2})^{I/2}}\,(\tau^2+J\sigma^2)^{1/2}$$
$$\times \exp\left\{-\frac{1}{2\tau^2}\sum_{i,j}(y_{ij}-\bar{y}_i)^2 - \frac{J}{2(\tau^2+J\sigma^2)}\sum_i(\bar{y}_i-\bar{y})^2\right\}.$$

(c) Show that the full posterior is not integrable since, for $\tau \neq 0$, $\pi(\sigma^2, \tau^2|y)$ behaves like σ^{-2} in a neighborhood of 0.

10.26 (Chib 1995) Consider the approximation of the marginal density $m(x) = f(x|\theta)\pi(\theta)/\pi(\theta|x)$, where $\theta = (\theta_1, \ldots, \theta_B)$ and the full conditionals $\pi(\theta_r|x, \theta_s, s \neq r)$ are available.

(a) In the case $B = 2$, show that an appropriate estimate of the marginal loglikelihood is

$$\hat{\ell}(x) = \log\, f(x|\theta_1^*) + \log\, \pi_1(\theta_1^*) - \log\, \left\{\frac{1}{T}\sum_{t=1}\pi(\theta_1^*|x, \theta_2^{(t)})\right\},$$

where θ_1^* is an arbitrary point, assuming that $f(x|\theta) = f(x|\theta_1)$ and that $\pi_1(\theta_1)$ is available.

(b) In the general case, rewrite the posterior density as

$$\pi(\theta|x) = \pi_1(\theta_1|x)\pi_2(\theta_2|x, \theta_1)\cdots\pi_B(\theta_B|x, \theta_1, \ldots, \theta_{B-1})\,.$$

Show that an estimate of $\pi_r(\theta_r^*|x, \theta_1^*, \ldots, \theta_{r-1}^*)$ is

$$\hat{\pi}_r(\theta_r^*|x, \theta_s^*, s<r) = \frac{1}{T}\sum_{t=1}^{T}\pi(\theta_r^*|x, \theta_1^*, \ldots, \theta_{r-1}^*, \theta_{r+1}^{(t)}, \ldots, \theta_B^{(t)})\,,$$

where the $\theta_\ell^{(t)}$ $(\ell > r)$ are simulated from the full conditionals $\pi_\ell(\theta_\ell|\theta_1^*, \ldots, \theta_r^*, \theta_{r+1}, \ldots, \theta_B)$.

(c) Deduce that an estimate of the joint posterior density is $\prod_{r=1}^{B} \hat{\pi}_r(\theta_r^*|x, \theta_s^*, s < r)$ and that an estimate of the marginal loglikelihood is

$$\hat{\ell}(x) = \log\ f(x|\theta^*) + \log\ \pi(\theta^*) - \sum_{r=1}^{B} \log\ \hat{\pi}_r(\theta_r^*|x, \theta_s^*, s < r)$$

for an arbitrary value θ^*.

(d) Discuss the computational cost of this method as a function of B.

(e) Extend the method to the approximation of the predictive density $f(y|x) = \int f(y|\theta)\pi(\theta|x)d\theta$.

10.27 (Fishman 1996) Given a contingency table with cell sizes N_{ij}, row sums $N_{i\cdot}$, and column sums $N_{\cdot j}$, the chi squared statistics for independence is

$$\chi^2 = \sum_{(i,j)} \frac{(N_{ij} - N_{i\cdot}N_{\cdot j}/N)^2}{N_{i\cdot}N_{\cdot j}/N}.$$

Assuming fixed margins $N_{i\cdot}$ and $N_{\cdot j}$ ($i = 1, \ldots, I, j = 1, \ldots, J$), the goal is to simulate the distribution of χ^2. Design a Gibbs sampling experiment to simulate a contingency table under fixed margins. (*Hint:* Show that the vector to be simulated is of dimension $(I-1)(J-1)$.) (*Note:* Alternative methods are described by Aldous 1987 and Diaconis and Sturmfels 1998.)

10.28 The one-way random effects model is usually written

$$Y_{i,j} = \mu + \alpha_i + \varepsilon_{ij}, \quad \alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2), \quad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \quad j = 1, \ldots, n_i, \quad i = 1, \ldots, k\,.$$

(a) Show it can also be written as the hierarchical model

$$\begin{aligned} Y_{i,j}|\alpha_i &\sim \mathcal{N}(\mu + \alpha_i, \sigma^2) \\ \alpha_i &\sim \mathcal{N}(0, \sigma_\alpha^2)\,. \end{aligned}$$

(b) Show the joint density is

$$\left(\frac{1}{\sigma_\alpha^2}\right)^{k/2} \exp \frac{1}{2\sigma_\alpha^2} \sum_i \alpha_i^2 \left(\frac{1}{\sigma^2}\right)^{Nk/2} \exp \frac{1}{2\sigma_\alpha^2} \sum_{ij} (y_{ij} - \mu - \alpha_i)^2.$$

(*Note:* This is a complete data likelihood if we consider the α_i's to be missing data.)

(c) Show that the full conditionals are given by

$$\begin{aligned} \alpha_i|\mathbf{y}, \mu, \sigma_\alpha^2, \sigma^2 &\sim N\left(\frac{n_i\sigma_\alpha^2}{n_i\sigma_\alpha^2 + \sigma^2}(\bar{y}_i - \mu), \frac{\sigma_\alpha^2\sigma^2}{n_i\sigma_\alpha^2 + \sigma^2}\right), \\ \mu|\mathbf{y}, \alpha, \sigma_\alpha^2, \sigma^2 &\sim N\left(\bar{y} - \bar{\alpha}, \frac{\sigma^2}{Nk}\right), \\ \sigma_\alpha^2|\mathbf{y}, \alpha, \mu, \sigma^2 &\sim \mathcal{IG}\left(\frac{k}{2} - 1, \frac{1}{2}\sum_i \alpha_i^2\right), \\ \sigma^2|\mathbf{y}, \alpha, \mu, \sigma_\alpha^2 &\sim \mathcal{IG}\left(\frac{Nk}{2} - 1, \frac{1}{2}\sum_{ij}(y_{ij} - \mu - \alpha_i)^2\right), \end{aligned}$$

where $N = \sum_i n_i$ and $\mathcal{IG}(a, b)$ is the inverted gamma distribution (see Appendix A).

(d) The following data come from a large experiment to assess the precision of estimation of chemical content of turnip greens, where the leaves represent a random effect (Snedecor and Cochran 1971). Run a Gibbs sampler to estimate the variance components, plot the histograms and monitor the convergence of the chain.

Leaf	% Ca			
1	3.28	3.09	3.03	3.03
2	3.52	3.48	3.38	3.38
3	2.88	2.80	2.81	2.76
4	3.34	3.38	3.24	3.26

Table 10.4. Calcium concentration (%) in turnip greens (*Source:* Snedecor and Cochran 1971).

10.29 †[11](Gelfand et al. 1990) For a population of 30 rats, the weight y_{ij} of rat i is observed at age x_j and is associated with the model

$$Y_{ij} \sim \mathcal{N}(\alpha_i + \beta_i(x_j - \bar{x}), \sigma_c^2).$$

(a) Give the Gibbs sampler associated with this model and the prior

$$\alpha_i \sim \mathcal{N}(\alpha_c, \sigma_\alpha^2), \qquad \beta_i \sim \mathcal{N}(\beta_c, \sigma_\beta^2),$$

with almost flat hyperpriors

$$\alpha_c, \beta_c \sim \mathcal{N}(0, 10^4), \quad \sigma_c^{-2}, \sigma_\alpha^{-2}, \sigma_\beta^{-2} \sim \mathcal{G}a(10^{-3}, 10^{-3}).$$

(b) Assume now that

$$Y_{ij} \sim \mathcal{N}(\beta_{1i} + \beta_{2i} x_j, \sigma_c^2)\ ,$$

with $\beta_i = (\beta_{1i}, \beta_{2i}) \sim \mathcal{N}_2(\mu_\beta, \Omega_\beta)$. Using a Wishart hyperprior $\mathcal{W}(2, R)$ on Ω_β, with

$$R = \begin{pmatrix} 200 & 0 \\ 0 & 0.2 \end{pmatrix},$$

give the corresponding Gibbs sampler.

(c) Study whether the original assumption of independence between β_{1i} and β_{2i} holds.

10.30 †(Spiegelhalter et al. 1995a,b) Binomial observations

$$R_i \sim \mathcal{B}(n_i, p_i), \qquad i = 1, \ldots, 12,$$

correspond to mortality rates for cardiac surgery on babies in hospitals. When the failure probability p_i for hospital i is modeled by a random effect structure

$$\text{logit}(p_i) \sim \mathcal{N}(\mu, \tau^2),$$

[11] Problems with this dagger symbol are studied in detail in the `BUGS` manual of Spiegelhalter et al. (1995a,b) . Corresponding datasets can also be obtained from this software (see Note 10.6.2).

with almost flat hyperpriors

$$\mu \sim \mathcal{N}(0, 10^4), \quad \tau^{-2} \sim \mathcal{G}a(10^{-3}, 10^{-3}),$$

examine whether a Gibbs sampler can be implemented. (*Hint:* Consider the possible use of ARS as in Section 2.4.2.)

10.31 †(Boch and Aitkin 1981) Data from the Law School Aptitude Test (LSAT) corresponds to multiple-choice test answers $(y_{j1}, \ldots, y_{j5})$ in $\{0,1\}^5$. The y_{jk}'s are modeled as $\mathcal{B}(p_{jk})$, with $(j = 1, \ldots, 1000, k = 1, \ldots, 5)$

$$\text{logit}(p_{jk}) = \theta_j - \alpha_k, \qquad \theta_j \sim \mathcal{N}(0, \sigma^2).$$

(This is the *Rasch model.*) Using vague priors on the α_k's and σ^2, give the marginal distribution of the probability P_i to answer $i \in \{0,1\}^5$. (*Hint:* Show that P_i is the posterior expectation of the probability $P_{i|\theta}$ conditional on ability level.)

10.32 †(Dellaportas and Smith 1993) Observations t_i on survival time are related to covariates z_i by a Weibull model

$$T_i|z_i \sim \mathcal{W}e(r, \mu_i), \qquad \mu_i = \exp(\beta^t z_i),$$

with possible censoring. The prior distributions are

$$\beta_j \sim \mathcal{N}(0, 10^4), \qquad r \sim \mathcal{G}a(1, 10^{-4}).$$

(a) Construct the associated Gibbs sampler and derive the posterior expectation of the median, $\log(2\,\exp(-\beta^t z_i))^{1/r}$.
(b) Compare with an alternative implementation using a slice sampler [A.32]

10.33 †(Spiegelhalter et al. 1995a,b) In the study of the effect of a drug on a heart disease, the number of contractions per minute for patient i is recorded before treatment (x_i) and after treatment (y_i). The full model is $X_i \sim \mathcal{P}(\lambda_i)$ and for uncured patients, $Y_i \sim \mathcal{P}(\beta\lambda_i)$, whereas for cured patients, $y_i = 0$.
(a) Show that the conditional distribution of Y_i given $t_i = x_i + y_i$ is $\mathcal{B}(t_i, \beta/(1+\beta))$ for the uncured patients.
(b) Express the distribution of the Y_i's as a mixture model and derive the Gibbs sampler.

10.34 †(Breslow and Clayton 1993) In the modeling of breast cancer cases y_i according to age x_i and year of birth d_i, an exchangeable solution is

$$\begin{aligned} Y_i &\sim \mathcal{P}(\mu_i), \\ \log(\mu_i) &= \log(d_i) + \alpha_{x_i} + \beta_{d_i}, \\ \beta_k &\sim \mathcal{N}(0, \sigma^2). \end{aligned}$$

(a) Derive the Gibbs sampler associated with almost flat hyperpriors on the parameters α_j, β_k, and σ.
(b) Breslow and Clayton (1993) consider a dependent alternative where for $k = 3, \ldots, 11$ we have

$$\beta_k|\beta_1, \ldots, \beta_{k-1} \sim \mathcal{N}(2\beta_{k-1} - \beta_{k-2}, \sigma^2), \tag{10.16}$$

while $\beta_1, \beta_2 \sim \mathcal{N}(0, 10^5\sigma^2)$. Construct the associated Gibbs sampler and compare with the previous results.

(c) An alternative representation of (10.16) is

$$\beta_k|\beta_j, j \neq k \sim \mathcal{N}(\bar{\beta}_k, n_k\sigma^2).$$

Determine the value of $\bar{\beta}_k$ and compare the associated Gibbs sampler with the previous implementation.

10.35 †(Dobson 1983) The effect of a pesticide is tested against its concentration x_i on n_i beetles, $R_i \sim \mathcal{B}(n_i, p_i)$ of which are killed. Three generalized linear models are in competition:

$$\begin{aligned} p_i &= \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}, \\ p_i &= \Phi(\exp(\alpha + \beta x_i)), \\ p_i &= 1 - \exp(-\exp(\alpha + \beta x_i)) \,; \end{aligned}$$

that is, the logit, probit, and log-log models, respectively. For each of these models, construct a Gibbs sampler and compute the expected posterior *deviance*; that is, the posterior expectation of

$$D = 2\left(\sum_{i=1}^{n} \hat{\ell}_i - \sum_{i=1}^{n} \ell_i\right),$$

where

$$\ell_i = r_i \log(p_i) + (n_i - p_i)\log(1 - p_i), \qquad \hat{\ell}_i = \max_{p_i} \ell_i.$$

10.36 †(Spiegelhalter et al. 1995a,b) Consider a standard Bayesian ANOVA model ($i = 1, \ldots, 4, j = 1, \ldots, 5$)

$$\begin{aligned} Y_{ij} &\sim \mathcal{N}(\mu_{ij}, \sigma^2), \\ \mu_{ij} &= \alpha_i + \beta_j, \\ \alpha_i &\sim \mathcal{N}(0, \sigma_\alpha^2), \\ \beta_j &\sim \mathcal{N}(0, \sigma_\beta^2), \\ \sigma^{-2} &\sim \mathcal{G}a(a, b), \end{aligned}$$

with $\sigma_\alpha^2 = \sigma_\beta^2 = 5$ and $a = 0$, $b = 1$. Gelfand et al. (1992) impose the constraints $\alpha_1 > \ldots > \alpha_4$ and $\beta_1 < \cdots < \beta_3 > \cdots > \beta_5$.

(a) Give the Gibbs sampler for this model. (*Hint:* Use the optimal truncated normal Accept–Reject algorithm of Example 2.20.)

(b) Change the parameterization of the model as ($i = 2, \ldots, 4, j = 1, 2, 4, 5$)

$$\begin{aligned} \alpha_i &= \alpha_{i-1} + \epsilon_i, &\qquad \epsilon_i > 0, \\ \beta_j &= \beta_3 - \eta_j, &\qquad \eta_j > 0, \end{aligned}$$

and modify the prior distribution in

$$\alpha_1 \sim \mathcal{N}(0, \sigma_\alpha^2), \quad \beta_3 \sim \mathcal{N}(0, \sigma_\alpha^2), \quad \epsilon_i, \eta_j \sim \mathcal{G}a(0, 1).$$

Check whether the posterior distribution is well defined and compare the performances of the corresponding Gibbs sampler with the previous implementation.

10.37 In the setup of Note 10.6.3:

(a) Evaluate the time requirements for the computation of the exact formulas for the $\mu_i - \mu_{i+1}$.
(b) Devise an experiment to test the maximal value of $x_{(n)}$ which can be processed on your computer.
(c) In cases when the exact value can be computed, study the convergence properties of the corresponding Gibbs sampler.

10.38 In the setup of Note 10.6.3, we define the *canonical moments* of a distribution and show that they can be used as a representation of this distribution.
(a) Show that the two first moments μ_1 and μ_2 are related by the following two inequalities:

$$\mu_1^2 \le \mu_2 \le \mu_1 \ ,$$

and that the sequence (μ_k) is monotonically decreasing to 0.
(b) Consider a kth-degree polynomial

$$P_k(x) = \sum_{i=0}^{k} a_i x^i .$$

Deduce from

$$\int_0^1 P_k^2(x) dG(x)\, dx \ge 0$$

that

$$a^t C_k a \ge 0, \qquad \forall a \in \mathbb{R}^{k+1}, \tag{10.17}$$

where

$$C_k = \begin{pmatrix} 1 & \mu_1 & \mu_2 & \dots & \mu_k \\ \mu_1 & \mu_2 & \mu_3 & \dots & \mu_{k+1} \\ \dots & \dots & \dots & \dots & \dots \\ \mu_k & \mu_{k+1} & & \dots & \mu_{2k} \end{pmatrix}$$

and $a^t = (a_0, a_1, \dots, a_k)$.
(c) Show that for every distribution g, the moments μ_k satisfy

$$\begin{vmatrix} 1 & \mu_1 & \mu_2 & \dots & \mu_k \\ \mu_1 & \mu_2 & \mu_3 & \dots & \mu_{k+1} \\ \dots & \dots & \dots & \dots & \dots \\ \mu_k & \mu_{k+1} & & \dots & \mu_{2k} \end{vmatrix} \ge 0.$$

(*Hint:* Interpret this as a property of C_k.)
(d) Using inequalities similar to (10.17) for the polynomials $t(1-t)P_k^2(t)$, $tP_k^2(t)$, and $(1-t)P_k^2(t)$, derive the following inequalities on the moments of G:

$$\begin{vmatrix} \mu_1 - \mu_2 & \mu_2 - \mu_3 & \dots & \mu_{k-1} - \mu_k \\ \mu_2 - \mu_3 & \mu_3 - \mu_4 & \dots & \mu_k - \mu_{k+1} \\ \dots & \dots & \dots & \dots \\ \mu_{k-1} - \mu_k & & \dots & \mu_{2k-1} - \mu_{2k} \end{vmatrix} \ge 0 \, ,$$

$$\begin{vmatrix} \mu_1 & \mu_2 & \dots & \mu_k \\ \mu_2 & \mu_3 & \dots & \mu_{k+1} \\ \dots & \dots & \dots & \dots \\ \mu_k & \mu_{k+1} & \dots & \mu_{2k-1} \end{vmatrix} \geq 0\,,$$

$$\begin{vmatrix} 1-\mu_1 & \mu_1-\mu_2 & \dots & \mu_{k-1}-\mu_k \\ \mu_1-\mu_2 & \mu_2-\mu_3 & \dots & \mu_k-\mu_{k+1} \\ \dots & \dots & \dots & \dots \\ \mu_{k-1}-\mu_k & & \dots & \mu_{2k-2}-\mu_{2k-1} \end{vmatrix} \geq 0\,.$$

(e) Show that the bounds in parts (c) and (d) induce a lower (upper) bound $\underline{c}_{2k}$ ($\bar{c}_{2k}$) on μ_{2k} and that part (c) [(d)] induces a lower (upper) bound $\underline{c}_{2k-1}$ ($\bar{c}_{2k-1}$) on μ_{2k-1}.

(f) Defining p_k as

$$p_k = \frac{c_k - \underline{c}_k}{\bar{c}_k - \underline{c}_k},$$

show that the relation between $(p_1, ..., p_n)$ and $(\mu_1, ..., \mu_n)$ is one-to-one for every n and that the p_i are independent.

(g) Show that the inverse transform is given by the following recursive formulas. Define

$$q_i = 1 - p_i, \quad \zeta_1 = p_1, \quad \zeta_i = p_i q_{i-1} \qquad (i \geq 2).$$

then

$$\begin{cases} S_{1,k} = \zeta_1 \ + \cdots + \zeta_k & (k \geq 1), \\ S_{j,k} = \sum_{i=1}^{k-j+1} \zeta_i S_{j-1,i+j-1} & (j \geq 2), \\ c_n = S_{n,n}. \end{cases}$$

(*Note:* See Dette and Studden 1997 for details and useful complements on canonical moments.)

10.39 The modification of [A.38] corresponding to the reparameterization discussed in Example 10.26 only involves a change in the generation of the parameters. In the case $k = 2$, show that it is given by

`Simulate`

$$p|\mu,\theta,\sigma,\tau \sim \mathcal{B}e(n_1+1, n_2+1);$$

$$\theta|\mu,\sigma,\tau,p \sim \mathcal{N}\left(\frac{(\overline{x}_2-\mu)}{\tau}\left[1+\frac{\sigma^2\zeta^{-2}}{n_2}\right]^{-1}, \frac{\sigma^2}{n_2+\sigma^2\xi^{-2}}\right);$$

$$\mu|\theta,\sigma,\tau,,p \sim \mathcal{N}\left(\frac{n_1\overline{x}_1+\sigma^{-2}n_2(\overline{x}_2-\tau\theta)}{n_1+n_2\sigma^{-2}}, \frac{\tau^2\sigma^2}{n_2\sigma^2+n_2}\right);$$

$$\sigma^{-2}|\mu,\theta,\tau,p \sim \mathcal{G}a\left(\frac{n_2-1}{2}, \frac{s_2^2+n_2(\overline{x}_2-\mu-\tau\theta)^2}{2\tau^2}\right)\mathbb{I}_{\sigma^2<1}; \tag{10.18}$$

$$\tau^{-2}|\mu,\theta,\sigma,p \sim \mathcal{G}a\left(\frac{n}{2}, \frac{1}{2}\left(s_1^2+n_1(\overline{x}_1-\mu)^2+\frac{s_2^2}{\sigma^2}+\frac{n_2(\overline{x}_2-\mu)^2}{n_2\zeta^2+\sigma^2}\right)\right).$$

10.40 (Gruet et al. 1999) As in Section 10.4.1, consider a reparameterization of a mixture of exponential distributions, $\sum_{j=1}^k p_j \ \mathcal{E}xp(\lambda_j)$.

(a) Use the identifiability constraint, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$, to write the mixture with the "cascade" parameterization

$$p\,\mathcal{E}xp(\lambda) + \sum_{j=1}^{k-1} (1-p)\cdots q_j\;\mathcal{E}xp(\lambda\sigma_1\cdots\sigma_j)\,,$$

with $q_{k-1} = 1$, $\sigma_1 \leq 1, \ldots, \sigma_{k-1} \leq 1$.

(b) For the prior distribution $\pi(\lambda) = \frac{1}{\lambda}$, p, $q_j \sim \mathcal{U}_{[0,1]}$, $\sigma_j \sim \mathcal{U}_{[0,1]}$, show that the corresponding posterior distribution is always proper.

(c) For the prior of part (b), show that $[A.34]$ leads to the following algorithm:

Algorithm A.47 –Exponential Mixtures–

2. Generate

$$\begin{aligned}
\lambda &\sim \mathcal{G}a\left(n, n_0\overline{x}_0 + \sigma_1 n_1 \overline{x}_1 + \cdots + \sigma_1\cdots\sigma_{k-1}n_{k-1}\overline{x}_{k-1}\right),\\
\sigma_1 &\sim \mathcal{G}a\left(n_1 + \cdots + n_{k-1}, \lambda\left\{n_1\overline{x}_1 + \sigma_2 n_2\overline{x}_2 + \cdots + \right.\right.\\
&\qquad\qquad \left.\left.\sigma_2\cdots\sigma_{k-1}n_{k-1}\overline{x}_{k-1}\right\}\right)\mathbb{I}_{\sigma_1\leq 1},\\
&\;\vdots\\
\sigma_{k-1} &\sim \mathcal{G}a\left(n_{k-1}, \lambda n_{k-1}\overline{x}_{k-1}\right)\mathbb{I}_{\sigma_{k-1}\leq 1},\\
p &\sim \mathcal{B}e(n_0+1, n-n_0+1),\\
&\;\vdots\\
q_{k-2} &\sim \mathcal{B}e(n_{k-2}+1, n_{k-1}+1),
\end{aligned}$$

where $n_0, n_1, \ldots, n_{k-1}$ denote the size of subsamples allocated to the components $\mathcal{E}xp(\lambda)$, $\mathcal{E}xp(\lambda\sigma_1)$, ..., $\mathcal{E}xp(\lambda\sigma_1\ldots\sigma_{k-1})$ and $n_0\overline{x}_0, n_1\overline{x}_1, \ldots, n_{k-1}\overline{x}_{k-1}$ are the sums of the observations allocated to these components.

10.6 Notes

10.6.1 A Bit of Background

Although somewhat removed from statistical inference in the classical sense and based on earlier techniques used in statistical Physics, the landmark paper by Geman and Geman (1984) brought Gibbs sampling into the arena of statistical application. This paper is also responsible for the name *Gibbs sampling*, because it implemented this method for the Bayesian study of *Gibbs random fields* which, in turn, derive their name from the physicist Josiah Willard Gibbs (1839–1903). This original implementation of the Gibbs sampler was applied to a discrete image processing problem and did not involve completion.

The work of Geman and Geman (1984) built on that of Metropolis et al. (1953) and Hastings (1970) and his student Peskun (1973), influenced Gelfand and Smith (1990) to write a paper that sparked new interest in Bayesian methods, statistical computing, algorithms, and stochastic processes through the use of computing algorithms such as the Gibbs sampler and the Metropolis–Hastings algorithm. It is interesting to see, in retrospect, that earlier papers had proposed similar solutions but did not find the same response from the statistical community. Among these, one may quote Besag (1974, 1986), Besag and Clifford (1989), Broniatowski et al. (1984), Qian and Titterington (1990), and Tanner and Wong (1987).