Bayesian estimation of transition probabilities from repeated cross sections

Ben Pelzer*

*Research Technical Department, University of Nijmegen,
P.O. Box 9104, 6500 HE Nijmegen, The Netherlands*

Rob Eisinga

*Department of Social Science Research Methods,
University of Nijmegen, P.O. Box 9104,
6500 HE Nijmegen, The Netherlands*

This paper discusses some simple practical advantages of Markov chain Monte Carlo (MCMC) methods in estimating entry and exit transition probabilities from repeated independent surveys. Simulated data are used to illustrate the usefulness of MCMC methods when the likelihood function has multiple local maxima. Actual data on the evaluation of an HIV prevention intervention program among drug users are used to demonstrate the advantage of using prior information to enhance parameter identificaiton. The latter example also demonstrates an important strength of the MCMC approach, namely the ability to make inferences on arbitrary functions of model parameters.

Key Words and Phrases: Bayesian analysis, MCMC, Markov transition models, repeated cross sections.

1   Introduction

PAAP (2002) has shown that MCMC methods are not just a new set of techniques that exploit modern computing technology. Rather, they allow researchers to work with statistical models (and data) previously considered intractable. These include models with dynamics in latent variables, hierarchical, mixture, item-response and nonresponse models and combinations of these model types (see CONGDON 2001). While the main advantage is estimation in complex models, Bayesian simulation has also some less sweeping but useful aspects. This short communication is concerned with the problem of estimating binary transition probabilities from independent repeated cross-sectional (RCS) data and aims to demonstrate some practical advantages of Bayesian statistics based on the following issues.

*b.pelzer@maw.kun.nl

Any model that can be estimated by maximum likelihood can obviously also be estimated by Bayesian simulation. However, when the likelihood function is asymmetric or has multiple local maxima, evaluating the likelihood only around the global maximum, as in ordinary maximum likelihood estimation (MLE), may produce inaccurate information about the distributions of the parameters. Bayesian simulation has an advantage in these circumstances because it is not concerned with finding the parameter values for which the likelihood reaches the global maximum. It is primarily concerned with generating samples from the posterior distribution of the parameters given both the data and a prior density and this distribution may be asymmetric and multimodal. Simulated data will be used to illustrate this. Also, identification may be less of a problem in Bayesian analysis compared with classic approaches such as MLE. While unidentified parameters cannot be estimated in MLE, in the Bayesian approach it is possible to use an 'informative' prior that can provide identification. Our example below is concerned with a simple type of Bayesian data combination, in which the posterior determined from a small sized panel data set is used as the prior for a subsequent analysis of repeated cross-sectional data to yield a set of identified parameters. Finally, MCMC offers the opportunity to make inferences on arbitrary functions of model parameters. We will use this ability to derive samples from the posterior distribution of entry and exit transition probabilities in RCS data.

2   Estimating binary transitions from RCS data

We will first briefly present the model we use to estimate transition probabilities from repeated cross sections. Consider a two-state Markov matrix of transition rates in which the cell probabilities sum to unity across rows. For this 2×2 table, we define the following three terms, were Y_{it} denotes the value of the binary random variable Y for observation i at time point t: $p_{it} = P(Y_{it} = 1)$, $\mu_{it} = P(Y_{it} = 1 | Y_{it-1} = 0)$, and $\lambda_{it} = P(Y_{it} = 0 | Y_{it-1} = 1)$. These probabilities give rise to the equation $p_{it} = \mu_{it}(1 - p_{it-1}) + (1 - \lambda_{it})p_{it-1} = \mu_{it} + \eta_{it}p_{it-1}$, where $\eta_{it} = 1 - \lambda_{it} - \mu_{it}$. If we let the initial probability $p_{i0} = 0$ (or $t \rightarrow \infty$), it is straightforward to show that the reduced form for p_{it} is

$$p_{it} = \mu_{it} + \sum_{\tau=1}^{t-1} \left(\mu_{i\tau} \prod_{s=\tau+1}^{t} \eta_{is} \right).$$

To estimate this equation with repeated independent cross-sectional data, current and backcasted values of time-invariant and time-varying covariates X_{it} (i.e., $X_{it}, X_{it-1}, \ldots, X_{i1}$) are employed to generate backward predictions of the transition probabilities ($\mu_{it}, \mu_{it-1}, \ldots, \mu_{i1}$ and $\lambda_{it}, \lambda_{it-1}, \ldots, \lambda_{i2}$) and thereby of the marginal probabilities ($p_{it}, p_{it-1}, \ldots, p_{i1}$). The transition probabilites themselves are specified as $\mu_{it} = F(X_{it}\beta)$ and $\lambda_{it} = 1 - F(X_{it}\beta^*)$, where F is the logistic link function

and β and β^* are two potentially different sets of parameters associated with two potentially different sets of covariates X_{it}. To incorporate 'non-backcastable' variables (i.e., time-dependent covariates for which past histories are unknown) into the model, two different sets of parameters are estimated for both μ_{it} and λ_{it}: one for the current transition probability estimates and a separate one for the preceding estimates. If we define Z_{it} as a vector of non-backcastable variables and ζ as the associated parameter vector representing the effect on μ_{it}, we can write $\text{logit}(\mu_{it}) = X_{it}\beta^{**} + Z_{it}\zeta$ for cross section t, and $\text{logit}(\mu_{it}) = X_{it}\beta$ for the cross sections $1, \ldots, t-1$. In our applications below we assume that $\beta^{**} = \beta$. Also, we define the first observed outcome of the process, $P(Y_{i1} = 1)$, to equal the state probability p_{i1} (rather then the transition probability μ_{i1}) and assume that the Y_{i1}'s are random variables with a probability distribution $\text{Prob}(Y_{i1} = 1) = F(X_{it}\delta)$, where F is the logistic function and δ a set of parameters to be estimated. ML estimates of β, β^* and can be obtained by maximizing the log likelihood $LL = \sum_{t=1}^{T} \sum_{i=1}^{n_t} \ell\ell_{it} = \sum_{t=1}^{T} \sum_{i=1}^{n_t} [y_{it}\log(p_{it}) + (1 - y_{it})\log(1 - p_{it})]$ with respect to the parameters, where n_t is the number of observations of cross section t and T is the number of cross sections. Fisher's method-of-scoring may be used for maximum likelihood estimation. If we suppress the subscript i and define $p_0 = 0$, the first order partial derivatives of $\ell\ell$ with respect to the parameters β and β^* are

$$\frac{\partial \ell\ell}{\partial \beta} = \frac{\partial \ell\ell}{\partial p_t} \cdot \frac{\partial p_t}{\partial \beta} = \frac{y_t - p_t}{p_t(1 - p_t)} \cdot \left(\frac{\partial p_{t-1}}{\partial \beta} \eta_t + \frac{\partial \mu_t}{\partial \beta}(1 - p_{t-1}) \right)$$

and

$$\frac{\partial \ell\ell}{\partial \beta^*} = \frac{\partial \ell\ell}{\partial p_t} \cdot \frac{\partial p_t}{\partial \beta^*} = \frac{y_t - p_t}{p_t(1 - p_t)} \cdot \left(\frac{\partial p_{t-1}}{\partial \beta^*} \eta_t - \frac{\partial \lambda_t}{\partial \beta^*} p_{t-1} \right),$$

where $\partial \mu_t / \partial \beta = x_t \mu_t (1 - \mu_t)$ and $\partial \lambda_t / \partial \beta^* = -x_t \lambda_t (1 - \lambda_t)$. Further details about the model are provided by MOFFITT (1993) and PELZER, EISINGA and FRANSES (2001, 2002). In the examples below, the ML estimates were used as starting values of the Markov chain to reduce the period required for burning-in the sampler.

3 Multimodal likelihood function and Bayesian simulation

The likelihood function can have multiple local maxima with some distributions and models and assuring oneself that a local maximum is indeed the global maximum can be computationally difficult or intensive. Also, if the likelihood function is not well behaved around its maximum, standard errors produced by MLE can lead to unreliable inferences. Markov chain algorithms for sampling from the posterior offer a more complete picture of the uncertainty in the estimation of the unknown parameters. We will illustrate this with a simulated data set. For this simulation, we generated data for $T = 5$ cross sections with $n_t = 2,500$ observations each, using the following equations and parameter values:

$$p_{i1} = \mu_{i1} \qquad\qquad\qquad\qquad\qquad \text{for } t = 1$$
$$p_{it} = \mu_i(1 - p_{it-1}) + (1 - \lambda_i)p_{it-1} \qquad\qquad \text{for } t = 2, \ldots, 5$$

$$\text{logit}(\mu_{i1}) = \beta_1 + \beta_2 X_i \qquad\qquad \beta_1 = -0.69 \quad \beta_2 = 0.25$$
$$\text{logit}(\mu_i) = \beta_3 + \beta_4 X_i \qquad\qquad \beta_3 = -1.09 \quad \beta_4 = 0.25$$
$$\text{logit}(\lambda_i) = -\beta_1^* - \beta_2^* X_i \qquad\qquad \beta_1^* = 0 \qquad\quad \beta_2^* = 0.75.$$

The X_i values were drawn from the standard normal distribution and subsequently rounded to the nearest integer. The values ranged from -4 to $+4$, with about 38% of the observations having zero values. Note that the X_i values were fixed over time. Also note that the transition probabilities μ_i and λ_i were taken to be time-constant. The Y_{it} values were sampled from a Bernoulli (p_{it}) distribution, $t = 1, \ldots, 5$.

The intercept values β_1, β_3 and β_1^* were selected so that for observations with zero X_i values the marginal probabilities equal $p_{i1} = p_{it} = \mu_i/(\mu_i + \lambda_i)$, which is 0.334. This steady state condition is reflected in the marginal distribution of the simulated Y_{it}, the proportions of $Y_{it} = 1$ being 0.34, 0.35, 0.36, 0.37, and 0.34 for the respective cross sections. In a steady state condition, different sets of parameter estimates for β and β^* may yield an (almost) identical maximum likelihood, especially if the covariate X_i has weak effects. If X_i has no effect at all, we may as well remove it from the model. However, for a model with intercept parameters only, infinitely many estimates satisfy $p_{i1} = p_{it} = \mu_i/(\mu_i + \lambda_i)$ and thus produce an identical maximum likelihood. The covariate X_i reduces the infinitely many ML estimates to a single one, but there still may be many sets of point estimates that yield nearly similar maximum likelihood. MCMC techniques, which seek to characterize the posterior distribution of the regression parameters, can be usefully applied here.

The Metropolis algorithm is often used to generate samples from the posteriors (TANNER 1996). As is well known, this scheme is potentially inefficient when confronted with posteriors with multiple peaks, especially if they are well separated. Multimodal target distributions (especially if the starting values trap us near one of the modes) lead to a poorly mixing chain that stays in small regions of the parameter space for long periods of time. The result is that a very large number of random draws is needed to locate the modes. An algorithm that is more efficient in these circumstances is parallel tempering (LIU 2000). Parallel tempering uses a number of chains to traverse the full parameter space, each chain being updated M times by the Metropolis algorithm. After M updates, a swap of the states of two randomly chosen adjacent chains is proposed and this swap is accepted with a particular probability. This swapping mechanism enables parallel chains to explore the entire parameter space, jumping over 'narrow' bridges with low likelihood, from one modal area to another one. To promote a visit of all the modes, the Markov chains can be 'heated' to different 'temperatures': a 'hot' chain is, when going through the M Metropolis updates, more willing to accept parameter proposals with a low likelihood than a

'cold' chain. Heating chains is especially effective when the modes are well separated. If the modes are close, heating the chains may be superfluous.

We started the analysis with ML estimation, using Fisher-scoring and the true parameter values as starting values. The resulting parameter estimates were $\hat{\beta}_1 = -0.70$, $\hat{\beta}_2 = 0.27$, $\hat{\beta}_3 = -1.04$, $\hat{\beta}_4 = 0.28$, $\hat{\beta}_1^* = -0.05$, $\hat{\beta}_2^* = 0.77$, and the corresponding log likelihood was -7729.72. These estimates are close to the true values used in simulating the data. We subsequently performed a number of MCMC analyses using different heating schemes. The trace plots of $\beta_3, \beta_4, \beta_1^*$ and β_2^* all showed two sample bands, indicating that the posterior distributions are bimodal. However, the unheated chain appeared to mix poorly. The 'hotter' the heated chains, the poorer the mixing of the unheated chain. We therefore decided to use multiple unheated chains in our simulation. The final analysis employed 20 unheated chains with uninformative priors. Five million samples were run, discarding 50,000 samples for initial settling. Figure 1 plots β_4 against sample iteration number (the other parameters are not displayed as their traces are very similar).

Figure 1 displays two well separated sample bands. The upper mode is located near a value of approximately 0.90 and the lower mode is close to 0.25, i.e., the true value of β_4. The posterior probability distributions of the μ_i and λ_i parameters are shown in Figure 2. As can be seen, the distributions are all bimodal. Also note that for all the parameters, the true values are located near the modes with the lowest density. To verify additionally that the likelihood has two different modes, ML estimation was performed using the modes with the highest density as starting values. The resulting parameter estimates were $\hat{\beta}_1 = -0.69$, $\hat{\beta}_2 = 0.27$, $\hat{\beta}_3 = -2.30$, $\hat{\beta}_4 = 0.92$, $\hat{\beta}_1^* = 1.50$, $\hat{\beta}_2^* = -0.02$. These estimates correspond with the high-density

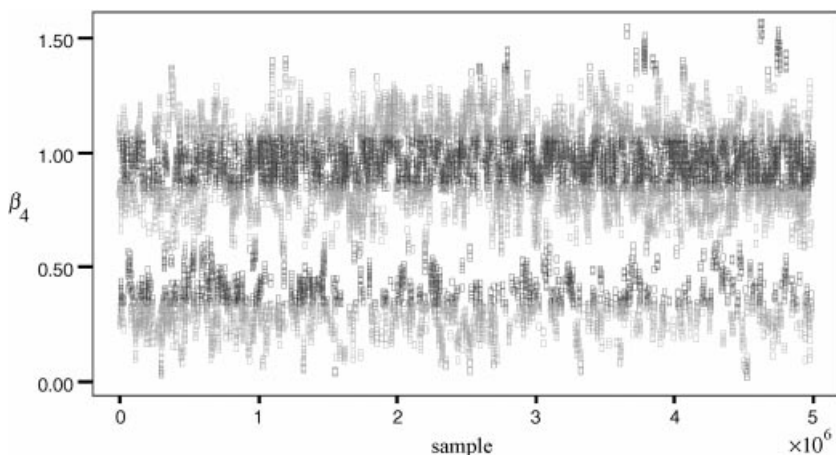
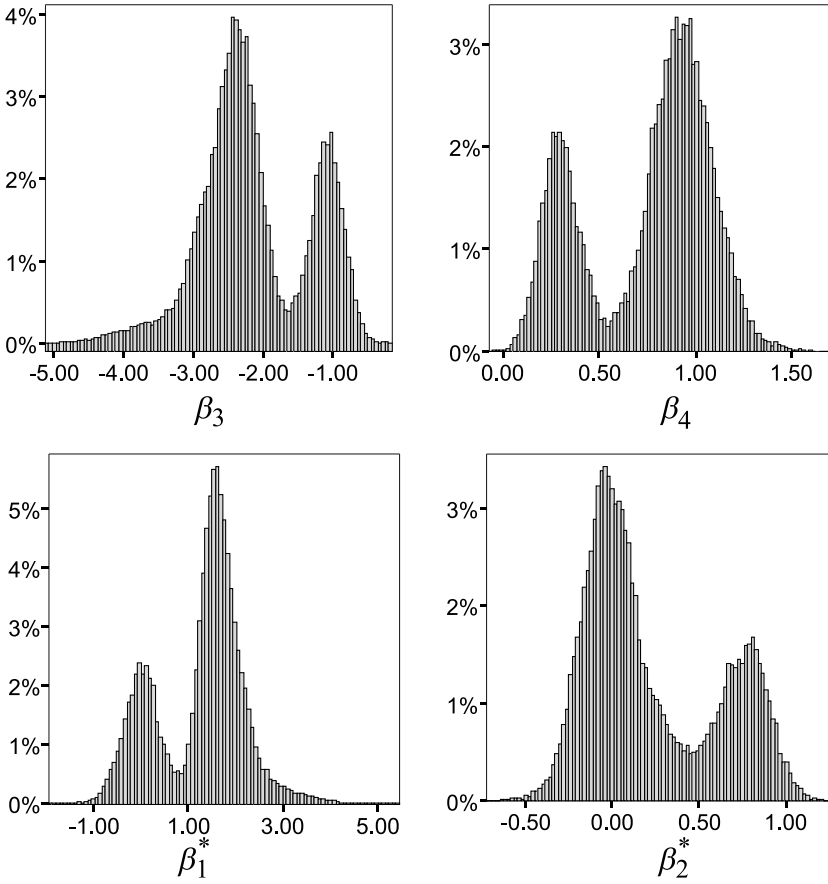Fig. 1. Trace of β_4 (for visual clarity, every 100th sample from the chain is displayed).

Fig. 2.   Posterior distributions of β_3, β_4, β_1^*, and β_2^*.

modes in Figure 2. The log likelihood obtained was -7728.47; hence slightly smaller than the log likelihood of the previous analysis.

   These results indicate that models and data with multimodal posteriors may easily cause the unwary ML user to get misleading results. A properly implemented MCMC method will produce the entire parameter distribution and thus reveal asymmetric or multimodal posteriors. In addition, under ML estimation we would compute the mode of the log likelihood function and use the local curvature to construct confidence intervals. Consider how odd it would be to use this procedure here. Since standard confidence intervals step on to some fixed distance from the mean and assume a normal parameter density, they completely ignore potentially multimodal or asymmetric features of the distribution. An advantage of Bayesian simulation is that it aims to recover the posterior density without the assumption of normality.

4 Bayesian data combination

Likelihood-based estimation can be troublesome when the parameters are barely identified or unidentified. In practice, however, additional knowledge may exist about the parameters. This information can, when incorporated in a Bayesian analysis as an informative prior, help to produce uniquely defined estimates. In the example below previously estimated model parameters computed from a different data set are combined with new observations to yield an updated set of identified parameters.

Table 1. Repeated cross-section and partial-transition data.*

Repeated cross-section data ($n = 1{,}337$)																
$u =$	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
Area $=$	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
Sex $=$	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
Talk $=$	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
Likely $=$	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

time (t)																
1	11	28	12	18	5	29	6	15	7	46	9	29	5	18	4	13
2	14	28	15	32	7	29	11	29	8	54	12	23	6	40	6	19
3	7	31	3	20	2	34	4	9	5	40	6	35	1	31	1	12
4	10	38	6	23	7	35	5	14	2	33	6	24	0	32	7	22
5	9	36	7	22	2	36	4	11	2	34	4	16	3	36	0	22

Partial-transition data ($n = 215$)

time ($t, t+1$)	$u =$	1	1	1	1	2	2	2	2	1	1	1	1	2	2	2	2
	$u' =$	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1, 2		1	0	1	1	1	2	0	0	2	1	0	0	0	11	0	0
2, 3		1	1	0	0	0	1	0	0	0	0	0	0	1	10	0	2
3, 4		0	0	0	0	0	4	0	2	0	1	0	0	0	4	0	4
4, 5		0	1	0	0	1	5	2	3	0	1	0	0	0	3	0	1

	$u =$	3	3	3	3	4	4	4	4	3	3	3	3	4	4	4	4
	$u' =$	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1, 2		1	0	0	0	0	2	1	3	0	0	1	0	0	2	0	1
2, 3		0	0	0	1	0	0	0	2	0	2	0	0	0	1	1	4
3, 4		0	0	0	0	0	1	0	5	0	1	0	0	0	3	0	3
4, 5		0	0	1	0	1	3	0	0	0	0	0	0	0	2	1	2

	$u =$	5	5	5	5	6	6	6	6	5	5	5	5	6	6	6	6
	$u' =$	5	6	7	8	5	6	7	8	5	6	7	8	5	6	7	8
1, 2		1	0	0	0	0	3	0	0	0	0	0	0	2	6	1	1
2, 3		0	1	0	0	1	6	0	0	0	1	0	0	0	9	0	0
3, 4		0	1	0	0	0	4	0	1	0	0	0	0	0	7	0	1
4, 5		0	1	0	0	1	6	1	2	0	0	0	0	0	10	0	1

	$u =$	7	7	7	7	8	8	8	8	7	7	7	7	8	8	8	8
	$u' =$	5	6	7	8	5	6	7	8	5	6	7	8	5	6	7	8
1, 2		0	0	1	1	0	2	0	2	0	0	0	0	0	1	0	1
2, 3		1	0	1	0	0	1	1	3	0	0	0	0	0	3	0	1
3, 4		0	0	0	1	0	1	1	3	0	0	0	0	0	1	0	1
4, 5		0	0	1	1	0	1	0	1	0	0	0	1	0	1	0	1

Source: Reprinted with permission of the International Biometric Society from HAWKINS and HAN (2000).
*The index u is used to present the partial-transition data economically.

Table 1 is based on data presented by HAWKINS and HAN (2000) taken from an evaluation study of an HIV prevention–intervention program among drug injectors attempting to modify high-risk behaviors such as sharing unbleached needles to inject drugs. The study consisted of repeated independent surveys conducted at five consecutive time-points in two geographical areas, i.e., an intervention area which underwent various intervention efforts and a comparison area which underwent no intervention. The variable of interest was knowledge of the risk of the transmission of HIV through sharing unclean needles, as measured by responses to the question "How likely is it that you will get AIDS if you share, but don't clean with bleach, drug needles?" The responses of the 1,337 drug users were one of two categories of LIKELY (0 = not likely, 1 = very likely). Explanatory variables include the time-constant covariates AREA (0 = comparison, 1 = intervention) and SEX (0 = male, 1 = female) and the time-varying covariate TALK (0 = no, 1 = yes). The latter variable records responses to the question "In the last 2 months, has anyone talked to you about AIDS, HIV, or cleaning needles with bleach?" In addition to the independent cross-sectional data, shown in the top part of Table 1, the study also collected partial-transition data (for pairs of consecutive waves) from a small sample of 215 drug users. The partial-transition data obtained by haphazard recaptures are shown in the bottom part of Table 1.

The repeated cross-sectional data alone can be used to estimate relatively simple transition models such as those with both time-constant intercepts and time-constant covariate effects. But models that drop this assumption are likely to produce problems of overparameterization. That is, it seems impossible to estimate more complex models without the partial-transition data. Several methods can be used to combine the two types of data. The procedure pursued here is based on the idea that the partial-transition data set provides useful auxiliary information about the behavior of the parameters in the repeated cross-sectional context. We therefore first analysed the partial-transition data separately using the Metropolis sampler with a non-informative prior for the regression parameters. The non-informative prior was approximated by a normal distribution with zero mean and variance 10^6. The means and the variance–covariance matrix of the estimated model parameters were thereupon transferred into the analysis of the repeated cross-sectional data. That is, they were used to construct a multivariate normal prior. Without this prior the problem would be overparameterized and the parameters would be unidentifiable. The regression parameters at $t = 1$ were assumed to follow independent normal distributions with zero mean and variance 10^6 (i.e., diffuse or non-informative priors). In the Markov chain sampling, we run the Metropolis algorithm 100,000 times excluding an initial burn-in of 5,000 samples. The posterior estimates are shown in Table 2.

One notes from this table that the entry decisions are affected by SEX. That is, the transition probabilities from the 'unlikely' to the 'very likely' response are higher among females than they are among males. Both AREA and TALK affect the probability of staying in the 'very likely' category (i.e., the $(1 - \lambda_t)$ transition). Hence

Table 2. Metropolis sampler posterior estimates.*

	$\delta(p_{t=1})$	t	$\beta(\mu_t)$	t	$\beta^*(1-\lambda_t)$
Area	0.504 (0.287) [−0.052, 1.071]	2–5	0.661 (0.734) [−0.669, 2.195]	2–5	0.667 (0.337) [0.025, 1.350]
Sex	0.186 (0.302) [−0.390, 0.798]	2–5	1.430 (0.633) [0.206, 2.696]	2–5	−0.082 (0.333) [−0.777, 0.555]
Talk	0.490 (0.300) [−0.089, 1.085]	2–5	−0.449 (0.693) [−1.929, 0.811]	2–5	1.133 (0.329) [0.505, 1.799]
Constant	0.690 (0.272) [0.148, 1.222]	2	−0.686 (0.794) [−2.279, 0.846]	2	1.046 (0.425) [0.257, 1.919]
		3–5	1.123 (0.682) [−0.097, 2.571]	3	1.436 (0.449) [0.621, 2.378]
				4	0.975 (0.363) [0.299, 1.727]
				5	1.190 (0.367) [0.499, 1.937]

*The mean of the last 100,000 samples is reported as the point estimate. The standard deviation is reported in parenthesis and the limits of the 95% credibility interval in brackets.

these probabilities are higher in the intervention area and among those reporting TALK = 'yes'.

Of course, other methods could be used to analyse these data. One is simply to pool the two data sets and to analyse the combined data using either maximum likelihood or Bayesian analysis with uninformative priors. When the same model is specified under these approaches, estimates from the ML and Bayesian procedures are close (even with these relatively small samples) and they would converge asymptotically. However, sometimes external constraints prohibit explicit data pooling. In many instances previous observations will not be available and even if they were, estimating the whole data might be so time-consuming that shortcut procedures using only the new data, and the estimates of the old as priors, would be appealing. Also, in many research problems data acquisition and data evaluation proceed in stages. Bayesian updating–the transfer of previously estimated model parameters to a new context–can reduce the need for a large data collection in the next stage.

A final issue we would like to address is that MCMC can be employed to obtain inference reaching beyond point estimates and approximate standard errors. A particular strength of the Markov chain Monte Carlo approach is the ability to make inferences on arbitrary functions of model parameters. Moreover, anything we wish to know about this function can be discovered up to any degree of accuracy via random sampling from the density distribution. We may, for example, obtain a sample from the posterior distribution of the mean entry and exit transition probabilities.
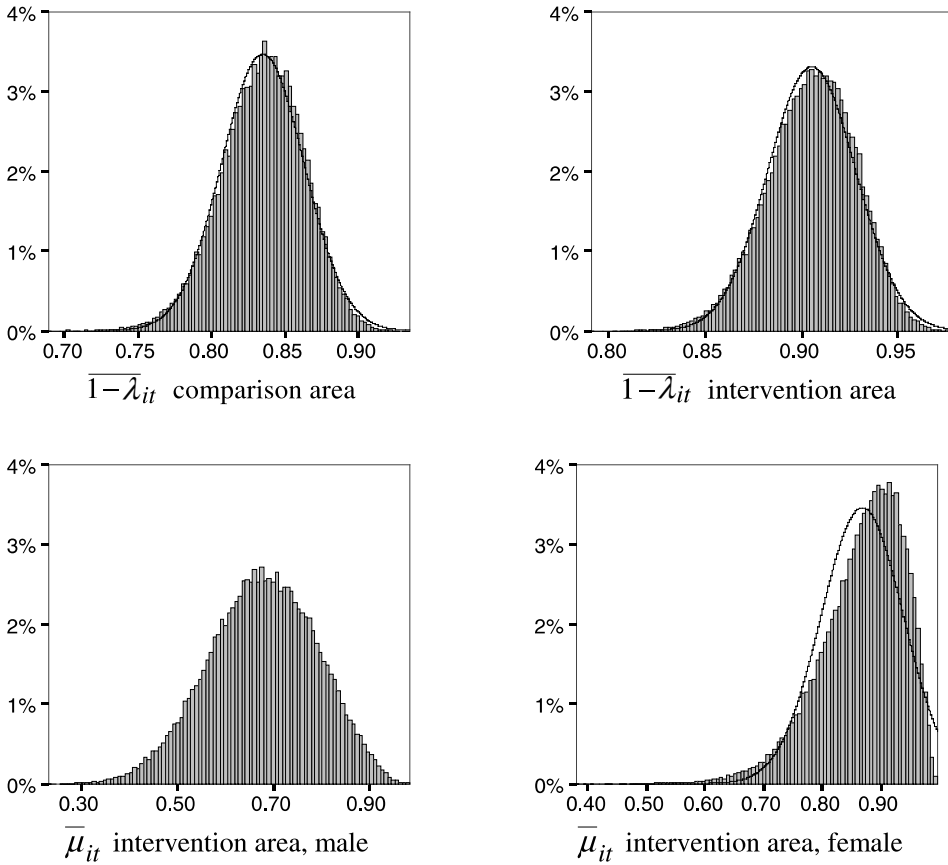
Fig. 3. Posterior distributions of average transition probabilities in comparison and intervention area (normal curve superimposed).

The top part of Figure 3 shows the mean posterior $(1 - \lambda_t)$ transition probabilities for the two study areas. The distributions illustrate the beneficial effect of the intervention program on $(1 - \lambda_t)$. The two densities presented in the bottom part of Figure 3 display the mean μ_t for males and females in the experimental area. These figures indicate considerable gender differences and they also show that the distribution for females is asymmetric.

5   In conclusion

We have presented two simple examples to illustrate the strengths of modern tools for Bayesian simulation. A straightforward advantage of the MCMC approach is that it provides estimates when traditional maximum likelihood struggles. Bayesian

simulation recovers the posterior precisely, without any need to rely on assumptions about the shape of the likelihood function. This feature may help one to arrive at a deeper understanding of the problem of interest.

References

CONGDON, P. (2001), *Bayesian statistical modelling*, Wiley, Chichester.

HAWKINS, D. L. and C. P. HAN (2000), Estimating transition probabilities from aggregate samples plus partial transition data, *Biometrics* **56**, 848–854.

LIU, J. S. (2001), *Monte Carlo strategies in statistical computing*, Springer, New York.

MOFFITT, R. (1993), Identification and estimation of dynamic models with a time series of repeated cross-sections, *Journal of Econometrics* **59**, 99–123.

PAAP, R. (2002), What are the advantages of MCMC based inference in latent variable models? *Statistica Neerlandica* **56**, 2–22.

PELZER, B., R. EISINGA and PH. H. FRANSES (2001), Estimating transition probabilites from a time series of independent cross sections, *Statistica Neerlandica* **55**, 249–262.

PELZER, B., R. EISINGA and PH. H. FRANSES (2001), *Inferring transition probabilities from repeated cross sections: a cross-level inference approach to US presidential voting*. Econometric Institute Report 2001-21, Erasmus University Rotterdam.

TANNER, M. (1996), *Tools for statistical inference*, Springer, New York.