A General Framework for Updating Belief Distributions

(Bissiri et al. (2016))

Standard Bayesian Approach

General posterior

Validity of General posterior

Calibration

Example

Standard Bayes

$X_n = x_1, \ldots, x_n$ random sample generated from $f(x|\theta)$

prior $\pi(\theta)$

$$\pi(\theta | X_n) \propto \pi(\theta) \prod_{i=1}^{n} f(x_i | \theta)$$

Challenge 1. need to specify $f(x|\theta)$

2. we might want to exclude parameters we're not interested in

If there is a framework of using general loss functions to convey Bayesian inference

e.g. method of moment

General Posterior

ℓ : loss function for a parameter θ

argmin $\ell(X_n : \theta)$

$$\pi(\theta \mid X_n) \propto \pi(\theta) \times \exp\left(-w\, \ell(X_n, \theta)\right)$$

w : ~~a~~ learning rate

↳ controls the posterior of the uncertainty

$w = 1$

negative log-likelihood $\Big\}$ → Standard Bayes

what are differences?

What happen if introduce non-informative prior for all the nuisance parameter?

difference with Lasso ?

Validity of General Posterior

$\ell(\cdot,)$
$L(\cdot,\cdot)$

explore ~~asymptotics for large~~ n

~ how to update prior belief $\pi(\theta)$ to get posterior belief $\pi(\theta \mid X_n)$?

ν : probability measure on space of θ

$\pi(\theta)$ ~ What is the "optimal" posterior $\hat{\nu}$?

$$\hat{\nu} = \arg\min_{\nu} L(\nu ; \pi, x)$$

$L(\nu : \pi, x)$: loss function on the space of probability measures on θ - space

① L

$$\pi(\theta \mid x) = \psi\{\ell(\theta, x), \pi(\theta)\}$$

$$\psi[I(\theta, x_2), \psi\{I(\theta, x_1), \pi(\theta)\}] = \psi\{I(\theta, x_1) + I(\theta, x_2), \pi(\theta)\}$$

② Additivity

$$L(r; \pi, x) \equiv h_1(r, x) + h_2(r, \pi)$$

represents "fidelity" to data

prior

coherence

- $h_2 \rightarrow$ KL divergence ← why? Intuition?

- h_1

$$\iint \ell(\theta, x)\, dF_0(x)\, r_1(d\theta) \leq \iint \ell(\theta, x)\, dF_0(x)\, r_2(d\theta)$$

prefer r_1 to r_2

$$h_1(r, x_n) = \int \ell(\theta, x_n)\, r(\theta)\, d\theta$$

$$\hat{r}(\theta) = \frac{\exp(-\ell(\theta, x_n))\,\pi(\theta)}{\int \exp(-\ell(\theta, x_n))\,\pi(\theta)\, d\theta}$$

Example : Survival Analysis

Example : Clustering

In standard Bayesian approach,

$$f(x \mid c) = \sum_{j=1}^{K} p_j \underbrace{f_j(x \mid c_j)}$$

c_j : parameters associated with the jth cluster

They

$$\ell(S, x_1, \dots, x_n) = w \sum_{C_k \in S} \sum_{i j \in C_k} (x_{ij} - \bar{x}_{C_k})^2$$

→ The posterior $\quad p(S \mid x) \propto \pi(S) \exp\{-\ell(S, x)\}$

① $R := \dfrac{p(S \mid x)}{p(S' \mid x)}$

② $1 + f_\alpha^*(k-1) / (n-k)$

③ $w := -\log(R)\, \ell(S, x) \dfrac{f_\alpha^*(k-1)}{n-k}$

Next week

Types of loss function

Calibration

Illustration

General forms of Information

Why $h_2 \to KL$?

go over one specific example.

when can do diff (standard Bayes. / General

① clustering

$$\sum \sum (x_{ij} - \bar{x}_{c_a})$$

why not θ_a ?

Example: Clustering

- Data

Co	de Name	0	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68
AL	Alabama	35	21	24	8	22	31	27	48	14	13	13	18	19	35	39	42	70	14
AR	Arkansas	35	40	37	20	28	39	29	39	13	18	21	30	21	44	46	43	44	31
DE	Delaware	54	54	52	33	50	56	58	65	51	43	45	45	50	52	55	49	39	45
FL	Florida	19	21	22	8	18	31	28	57	25	24	26	30	34	55	57	52	48	41
GA	Georgia	29	18	31	4	7	29	18	43	8	13	15	18	18	30	33	37	54	30
KY	Kentucky	49	47	48	25	47	49	49	59	40	40	42	43	41	50	54	54	36	44
LA	Louisiana	21	10	12	5	7	31	20	24	7	11	14	19	17	47	53	29	57	23
MD	Maryland	52	49	49	24	45	55	45	57	36	37	41	48	49	55	60	46	35	42
MS	Mississippi	10	5	7	2	5	14	8	18	4	3	4	6	3	40	24	25	87	14
MO	Missouri	46	50	49	30	47	55	50	56	35	38	48	48	42	51	50	50	36	45
NC	North Carolina	45	40	46	12	42	43	55	29	29	27	26	33	33	46	49	48	44	40
SC	South Carolin	7	5	6	1	2	4	2	9	2	1	4	4	4	49	25	49	59	39
TN	Tennessee	45	53	46	24	43	51	44	54	32	31	33	39	37	50	49	53	44	38
TX	Texas	31	22	22	9	17	24	20	52	11	12	19	17	25	53	55	49	37	40
VA	Virginia	44	37	38	17	32	38	33	54	30	29	32	37	41	56	55	52	46	43
WV	West Virginia	54	55	53	21	49	55	49	58	44	39	43	45	42	48	47	54	32	40

Figure: Voting of Southern states

Example: Clustering

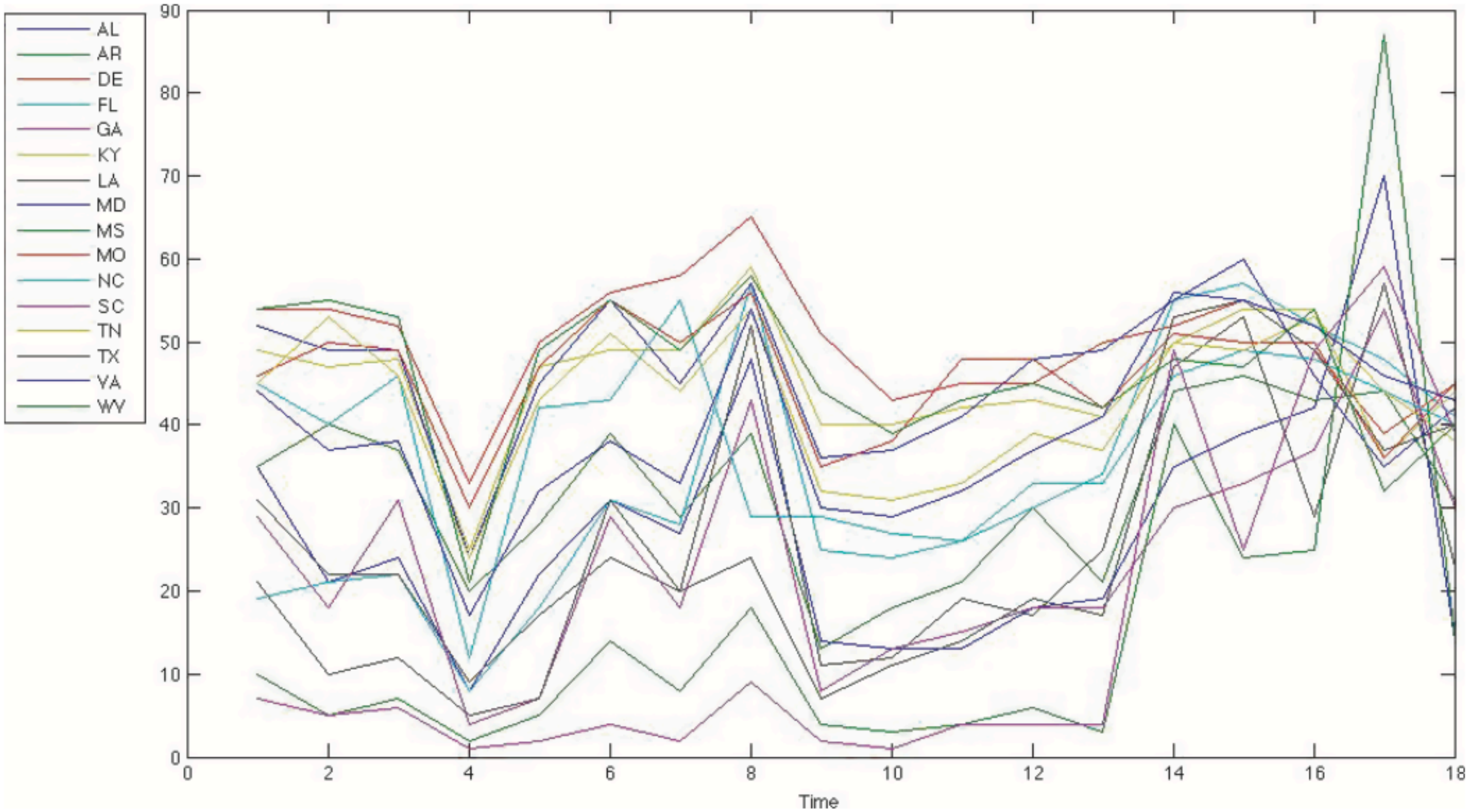

Fig. 4. Voting of southern states, illustrating the percentage of the Republican vote for Presidential elections every 4 years beginning in 1900: AL, Alabama; AR, Arkansas; DE, Delaware; FL, Florida; GA, Georgia; KY, Kentucky; LA, Louisiana; MD, Maryland; MS, Mississippi; MO, Missouri; NC, North Carolina; SC, South Carolina; TN, Tennessee; TX, Texas; VA, Virginia; WV, West Virginia

Figure: Voting of Southern states

Example: Clustering

Table 1. Average loss of partitions across MCMC samples (and log-posterior probabilities in parentheses)†

Number of state clusters k_s	*Average loss $\times 10^4$ for the following numbers of change points in time k_t (groups $= k_t + 1$)*		
	$k_t = 0$	$k_t = 1$	$k_t = 2$
1	7.98 (−14.49)	6.82 (−14.34)	6.72 (−14.73)
2	5.36 (−13.69)	5.13 (−13.65)	3.19 (−13.58)
3	5.09 (−13.64)	3.92 (−13.38)	2.36 (−*13.28*)
4	4.99 (−13.91)	3.32 (−13.50)	2.02 (−13.41)

†The average loss is $T^{-1} \Sigma_{i=1}^{T} l(S_i, x)$ with $S_i \sim \pi(S|x, k_s, k_t)$, where k_s denotes the number of clusters of states and k_t denotes the number of time series change points. Log-posterior-probabilities are shown in parentheses using a Poisson(3) and Poisson(2) prior on the number of groups and number of time clusters $k_t + 1$. The maximum posterior clustering is shown in italics.

Example: Clustering

- If I use k-means clustering without change points for $k = 1, \ldots, 4$ and calculate $\sum_{C_k \in S} \sum_{ij \in C_k} (x_{ij} - \bar{x}_{C_k})^2$, I got

	[,1]	[,2]	[,3]	[,4]
k	1.000000	2.000000	3.000000	4.000000
los_vec	1.245333	7.981199	5.289999	4.854897

Figure:

Example: Clustering

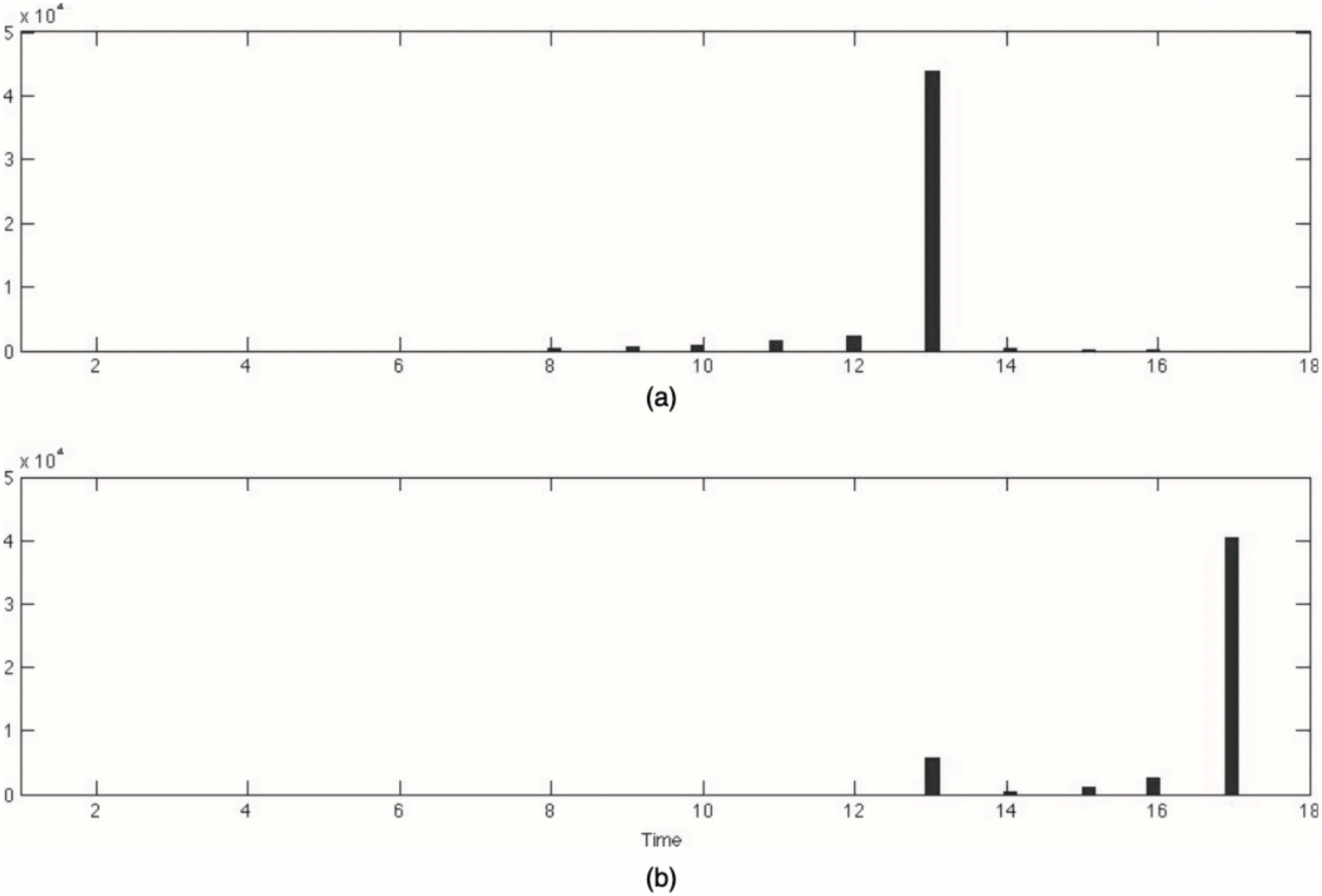

Fig. 5. Time change point locations for the two-change-point, $k_t = 2$, model and $k_s = 3$ groups: (a) change point 1; (b) change point 2

Figure:

Example: Clustering

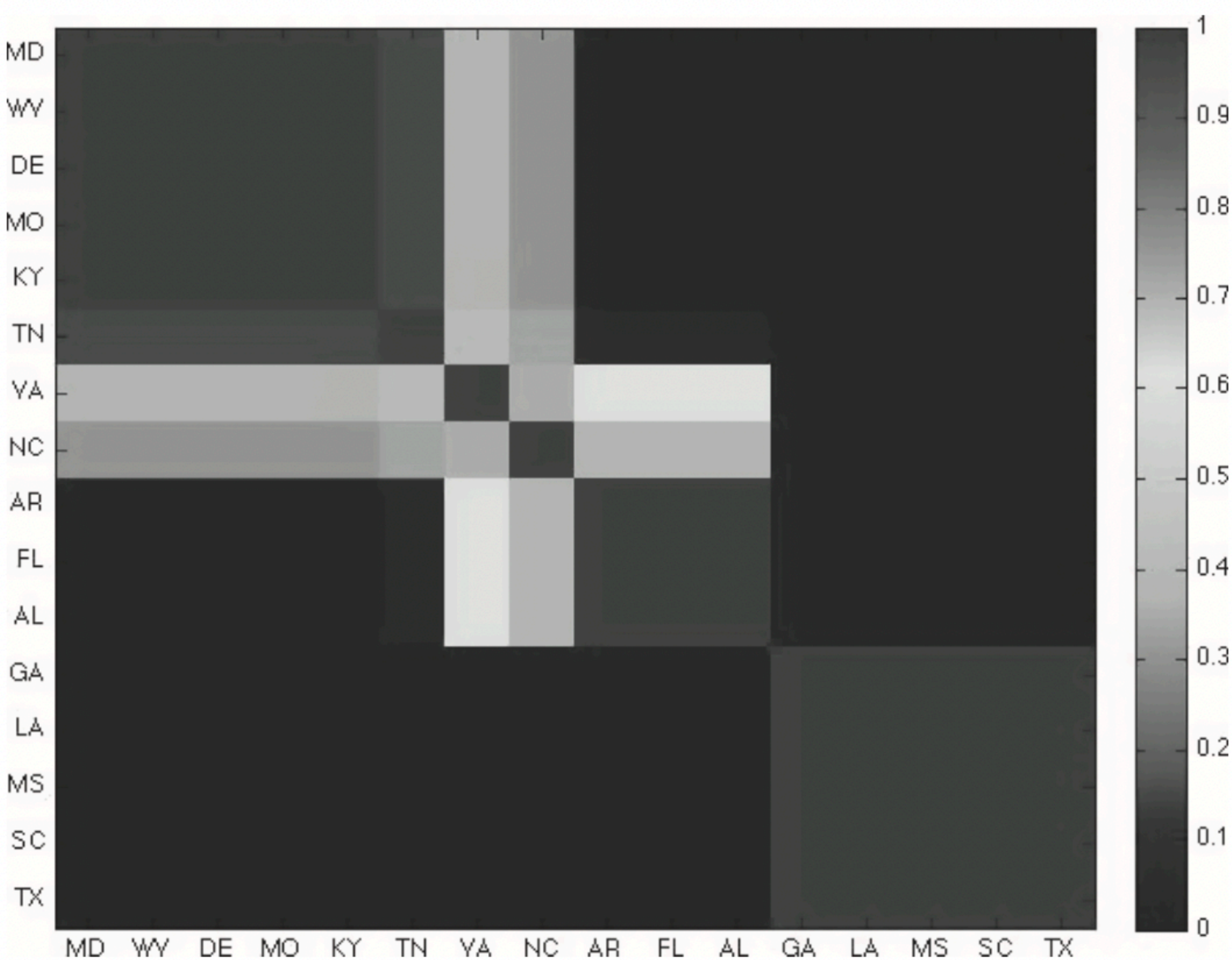

Fig. 6. Pairwise co-clustering probabilities across three groups and two time change points: AL, Alabama; AR, Arkansas; DE, Delaware; FL, Florida; GA, Georgia; KY, Kentucky; LA, Louisiana; MD, Maryland; MS, Mississippi; MO, Missouri; NC, North Carolina; SC, South Carolina; TN, Tennessee; TX, Texas; VA, Virginia; WV, West Virginia

References I

- P. G. Bissiri., C. C. Holmes & S. G. Walker (2016). A general framework for updating belief distributions. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 78(5), 1103-1130.