

Compare posteriors

Phyllis

11/16/2020

For the running example with in Section 3 where a population of $N = 100$ heterogeneous agents that interact through a fully connected network for $T = 90$ steps, we want to compare the approximate posterior samples from three algorithms: (1) pmcmc, (2) single-site gibbs and (3) blocked gibbs updates with $B = 5$ agents. The simulated dataset and the algorithms have been detailed in the paper. Since this is a simulated time series and we have access to the data generating parameters (dgp), the parameters have been initialized from the dgp.

We might add blocked gibbs updates with $B = 1$ to the comparision.

```
load("data_run_gibbs.RData");
singlesite_accept_chain_dgpinit <- accept_chain;
singlesite_lpost_chain_dgpinit <- lpost_chain;
singlesite_param_chain_dgpinit <- chain;

load("data_run_block_gibbs.RData");
block_accept_chain_dgpinit <- accept_chain;
block_lpost_chain_dgpinit <- lpost_chain;
block_param_chain_dgpinit <- chain;

load("data_run_pmcmc_final.RData");
pmcmc_accept_chain_dgpinit <- accept_chain;
pmcmc_lpost_chain_dgpinit <- lpost_chain;
pmcmc_param_chain_dgpinit <- chain;

load("data_run_pmcmc_randominitial.RData");
pmcmc_accept_chain_randinit <- accept_chain;
pmcmc_lpost_chain_randinit <- lpost_chain;
pmcmc_param_chain_randinit <- chain;

load("data_run_gibbs_randominit.RData");
singlesite_accept_chain_randinit <- accept_chain;
singlesite_lpost_chain_randinit <- lpost_chain;
singlesite_param_chain_randinit <- chain;

load("data_run_block_gibbs_randominit.RData");
block_accept_chain_randinit <- accept_chain;
block_lpost_chain_randinit <- lpost_chain;
block_param_chain_randinit <- chain;

load("data_run_block_gibbs_badinit.RData");
block_accept_chain_tailinit <- accept_chain;
block_lpost_chain_tailinit <- lpost_chain;
```

```

block_param_chain_tailinit <- chain;

nburn <- burn_in;
nmcmc <- num_mcmc;

set.seed(2020);
rand_init_param <- c(rnorm(6, sd = 3), runif(1)); ### RANDOM INIT
tail_init_param <- c(-10,4,-6,6,6,6,0.95); ## a bad point

```

The average acceptance probabilities (of the parameters) are:

```

##      pmcmc_dgp singlesite_dgp    block5_dgp      pmcmc_rand
## 0.336152      0.223438      0.123505      0.336000
## singlesite_rand    block5_rand
## 0.278514      0.120276

```

Posterior mean

```

##          method        b01        b02        b11
## 1      dgp -4.59511985013459          0         -1
## 2      pmcmc_dgp -4.92370875731419 -0.160108793568598 1.03701986897339
## 3  singlesite_dgp -4.62888999720101 -0.111268059048342 -1.15418657944351
## 4      block5_dgp -4.82958191002586 -0.278393727904335 1.44613953599017
## 5      pmcmc_rand -5.02623569848901 -0.100216333058715 0.808377190042494
## 6 singlesite_rand -4.98212416019782 -0.395597857602707 0.0546401022083614
## 7      block5_rand -5.27043772677146 0.183427556311835 0.836745001236306
## 8      block5_tail -4.93247180064674 0.312175584564511 1.48032032689414
##          bl2        bg1        bg2
## 1          2         -1         -1
## 2 -0.1106318112684 0.471423344039665 -0.227310007174236
## 3 2.29804906914684          -1.47433919753 -1.21344195889324
## 4 0.0327017198744929 -0.389157237863423 0.040926248758967
## 5 0.636911063116238 0.696782749412329 -0.453071380063606
## 6 -1.05651442661052 -0.0733257575018876 3.1263544578279
## 7 0.0637358881354053 -0.589268912691186 0.0417643765919665
## 8 0.0344559721527489 -0.488490648389561 0.0333368824318559
##          rho
## 1          0.8
## 2 0.840970635135669
## 3 0.706281321508852
## 4 0.674918606759719
## 5 0.831468293315779
## 6 0.806069869900597
## 7 0.666210939857251
## 8 0.664808916065744

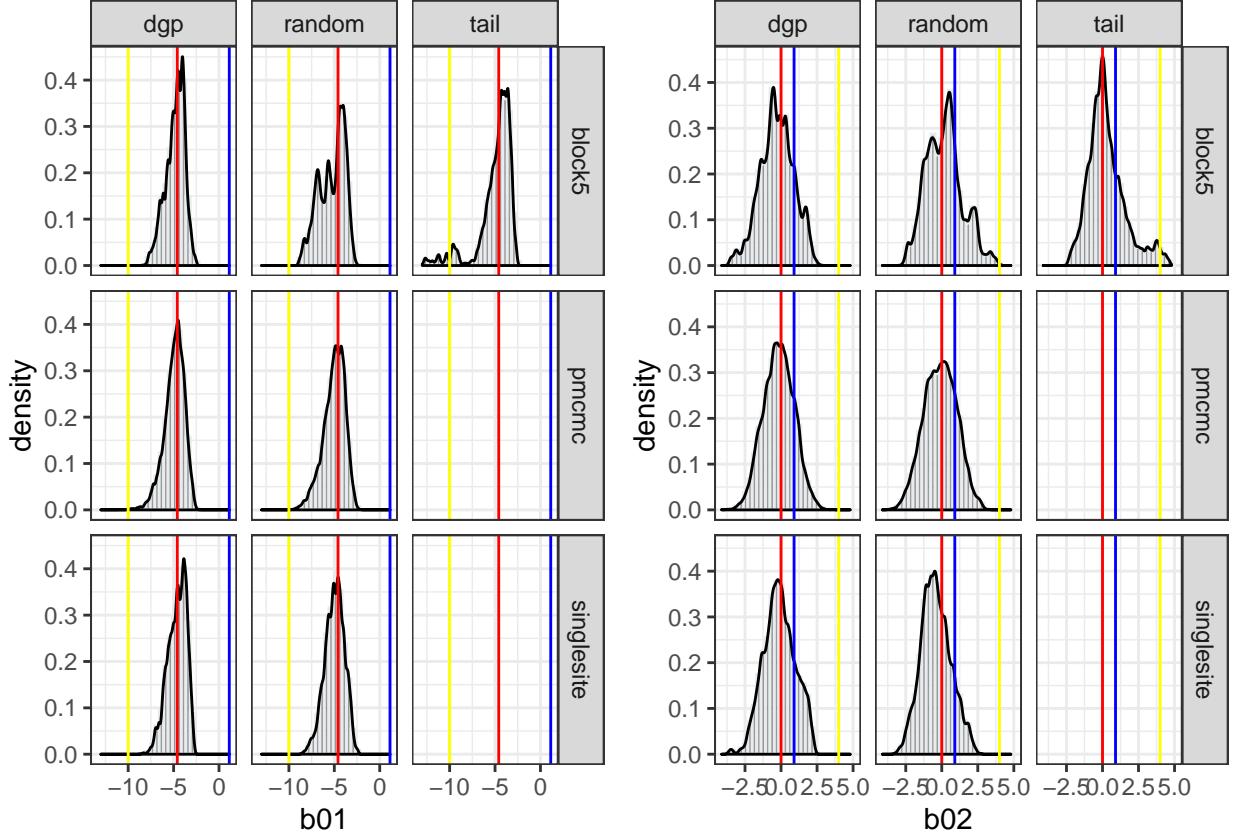
```

We can see that the singlesite Gibbs sampler recovers the signs of the parameters, although it is very likely that this happens due to initializing the chains from dgp and poor mixing of the singlesite Gibbs sampler and that it does not imply that singlesite Gibbs is superior to the other two methods.

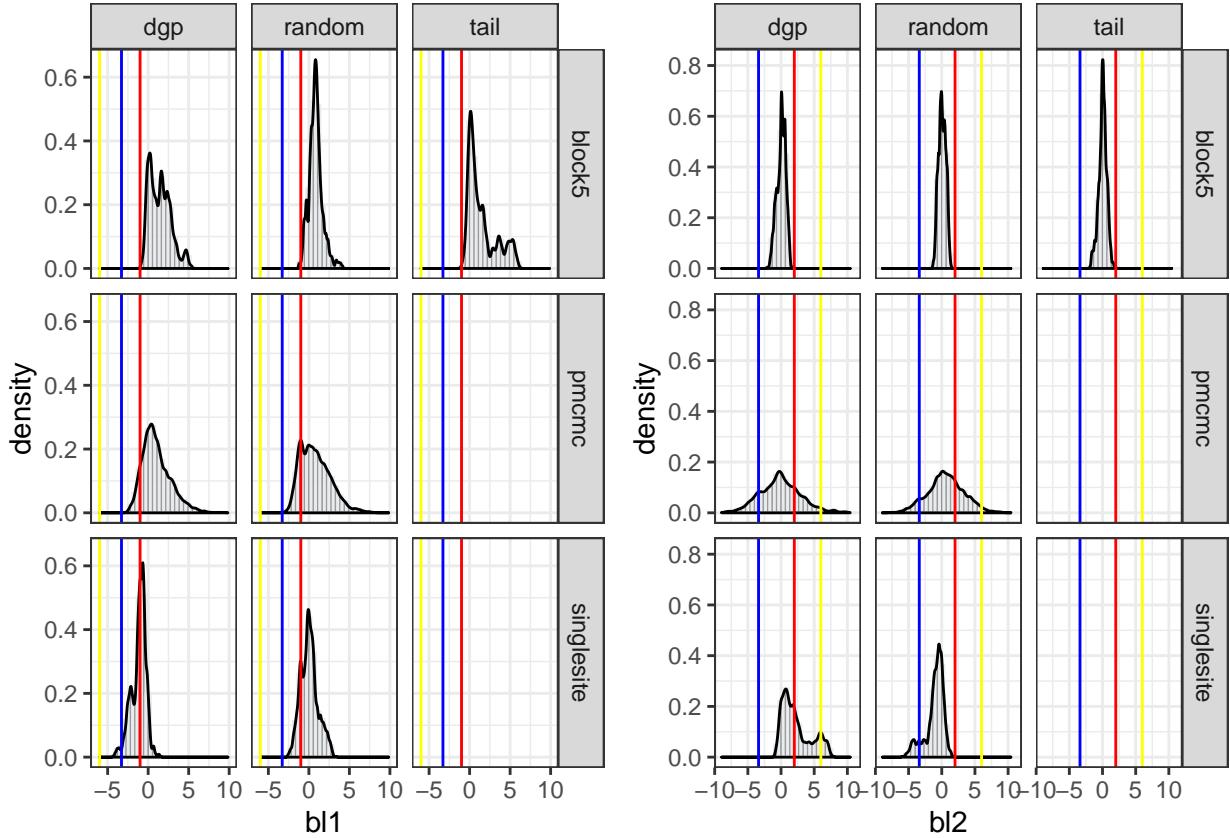
Posterior histograms

For all the 7 parameters, the posterior samples from PMCMC are more spread out. The posterior histograms for parameters β_0^1 and β_0^2 using the three methods look similar. The posterior histograms for the other parameters, $\beta_\lambda^{\{1,2\}}$, $\beta_\gamma^{\{1,2\}}$ and ρ are very different.

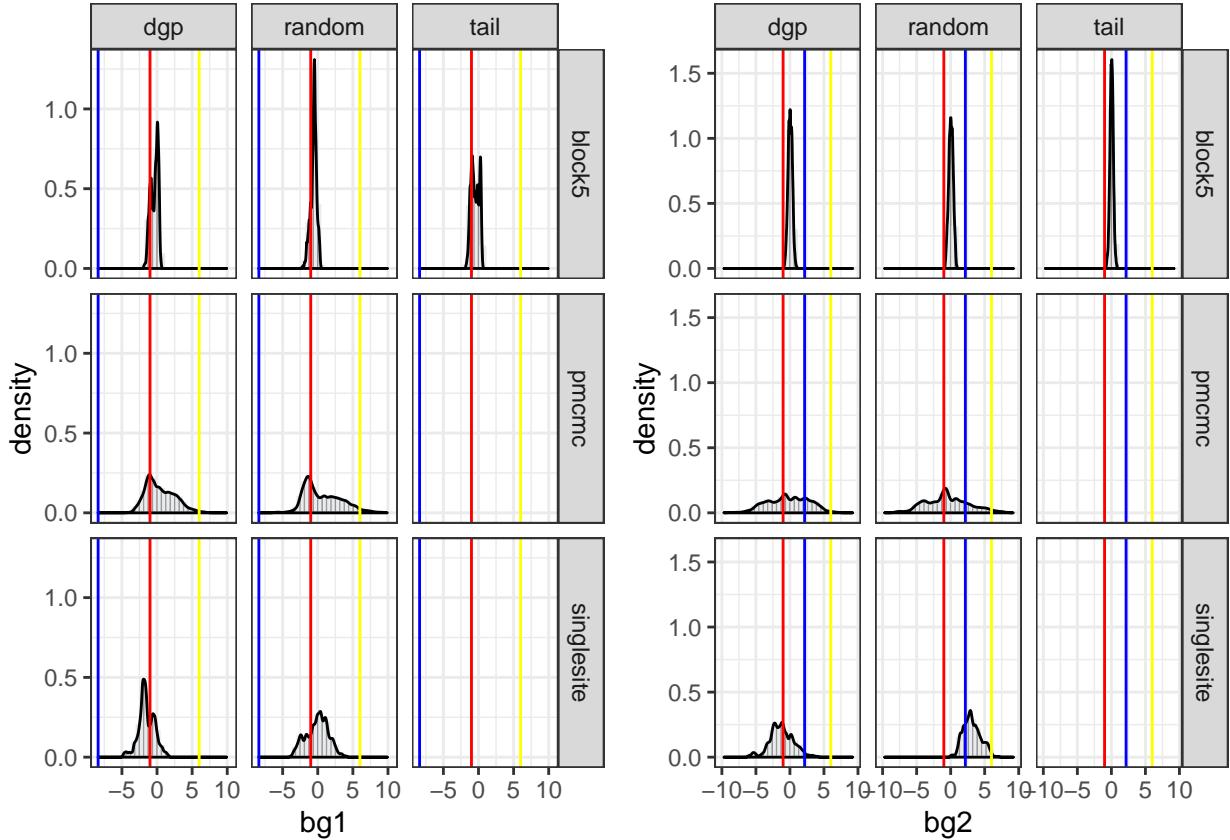
In this document, the red lines are dgp and the blue lines are the random initialization for the samplers.



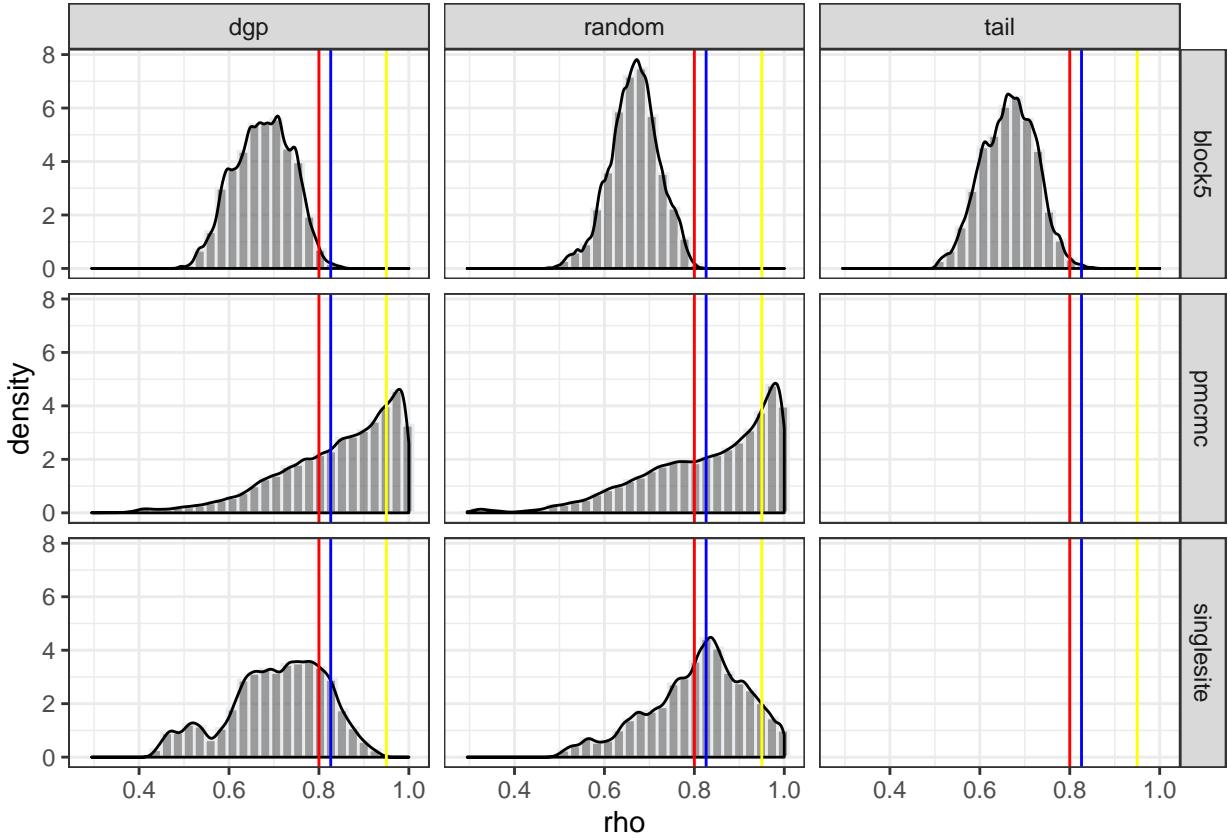
For the parameter $\beta_0^{1,2}$, the posterior from PMCMC is unimodal and centers around the DGP. For β_0^1 (with DGP -4.6), the posteriors from both Gibbs strategies are left-skewed and maximized around -4. The posteriors of β_0^2 seem to be symmetric around the true value 0 for all three samplers.



The posterior of $\beta_\lambda^{1,2}$ from PMCMC are unimodal with MAP around $\beta_\lambda^{\{1,2\}} = 0$, when the DGP values are -1 and 2 . The posteriors from Gibbs samplers are, however, multimodal. Single-site Gibbs yields samples with smaller β_λ^1 and larger β_λ^2 . But blocked Gibbs updates gives more samplers that have larger β_λ^1 and smaller β_λ^2 . The Gibbs samplers seems to not explore the sample space as much as the PMCMC.



For the parameters $\beta_\gamma^{\{1,2\}}$, we can see that samples from the block Gibbs sampler is very concentrated and it does not visit the parameter space with $\beta_\gamma^2 < \beta_\gamma^{2,*} = -1$. The posterior from pmcmc seems to be unimodal and the MAP is close to the DGP, but the posteriors from Gibbs samplers looks multimodal.

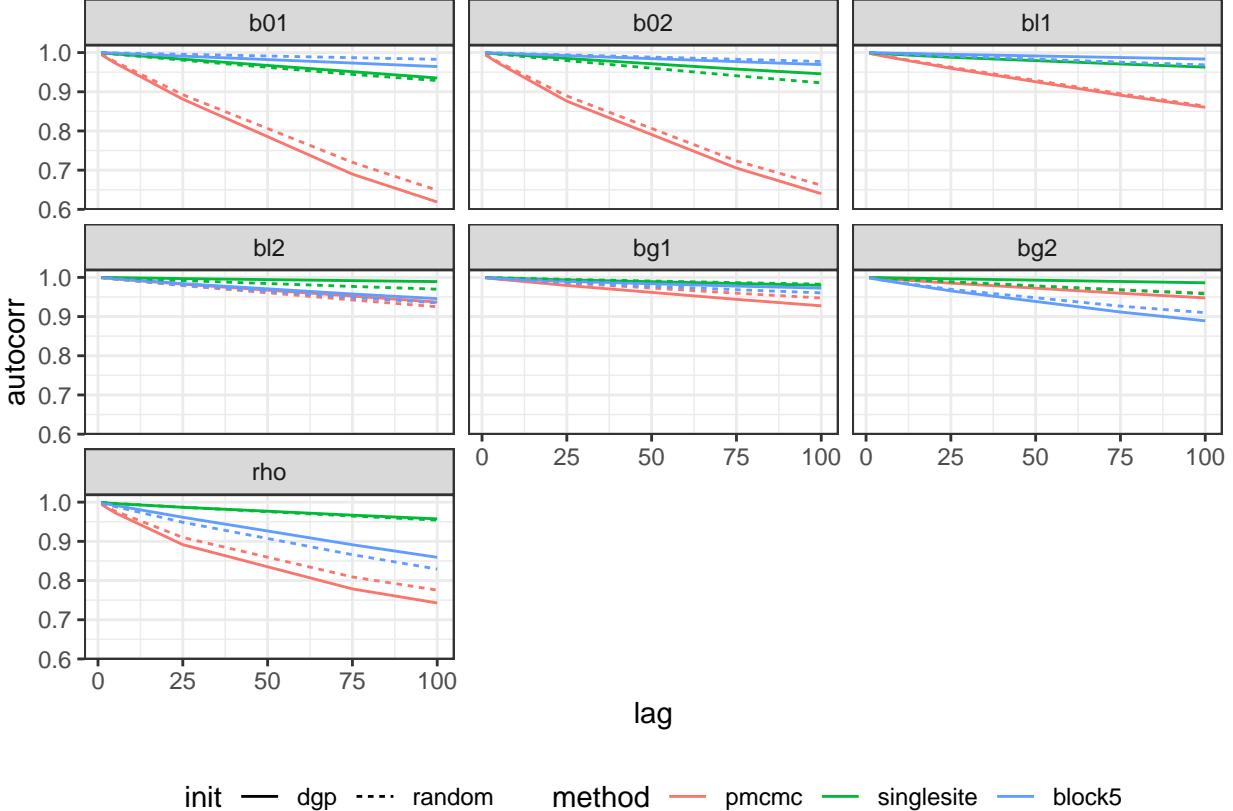


Finally, the posterior of ρ from PMCMC shows a mode at $\rho \approx 1$ and the density monotonically increase as ρ becomes larger. However, the posteriors from Gibbs samplers rarely visits values of $\rho > 0.8$. As we have seen in the experiments on particle filters, the variance of $\log \hat{Z}(\rho)$ is smaller as ρ gets closer to 1, which implies that the PMCMC becomes more similar to the idealized MCMC algorithm, as if we can compute exactly the marginal likelihood. I wonder if the shape of PMCMC samples is a result of the small variance of $\log \hat{Z}(\rho)$.

Auto-correlation

In order to access mixing, we can inspect the auto correlation plots of certain parameters (or functions of parameters).

```
p_acf <- ggplot(acf_df, aes(x = lag, y = autocorr)) + geom_line(aes(linetype = init, colour = method))
p_acf
```

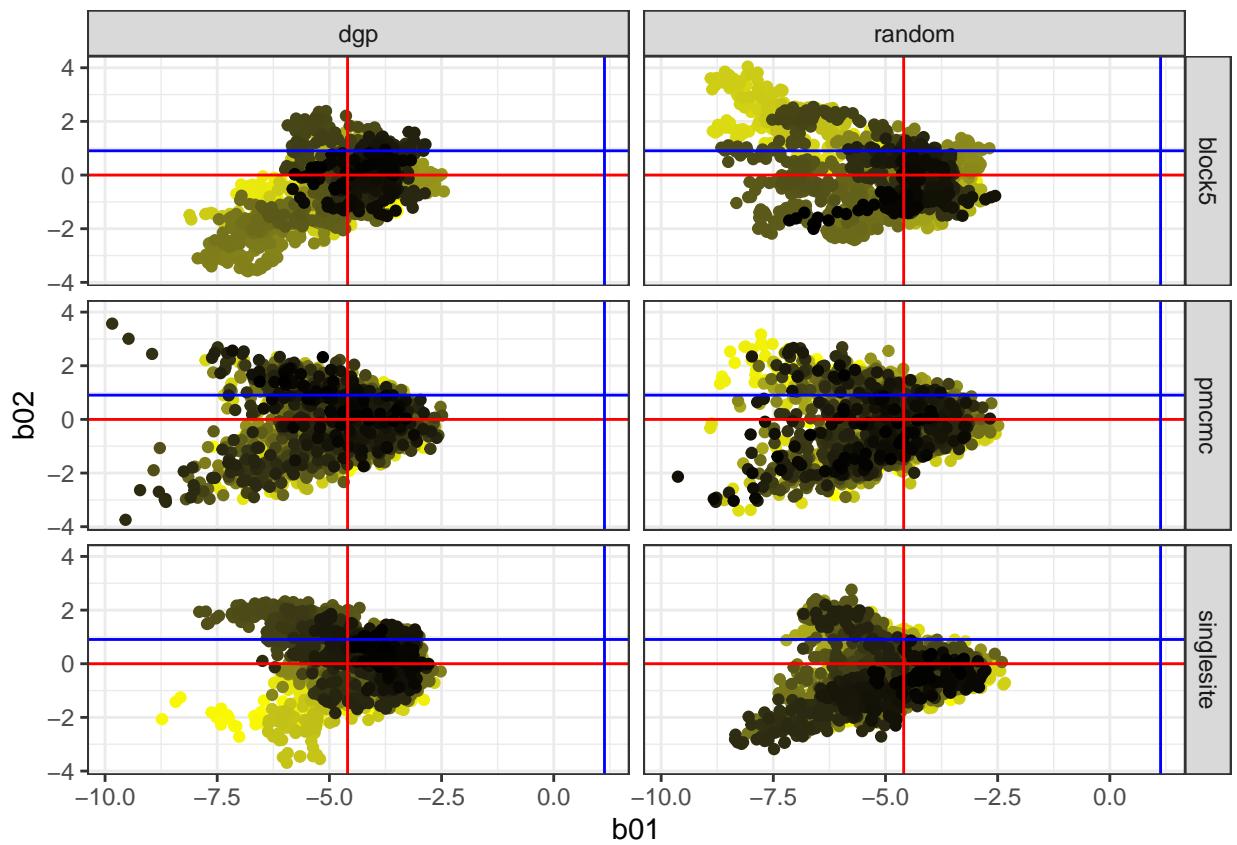


The samplers are very sticky: for β_γ^2 , with lag = 100, the smallest the auto correlation with lag = 100 iterations is 0.88. For single-site gibbs, the auto correlation of β_γ^2 with lag of 100 iterations is 0.986.

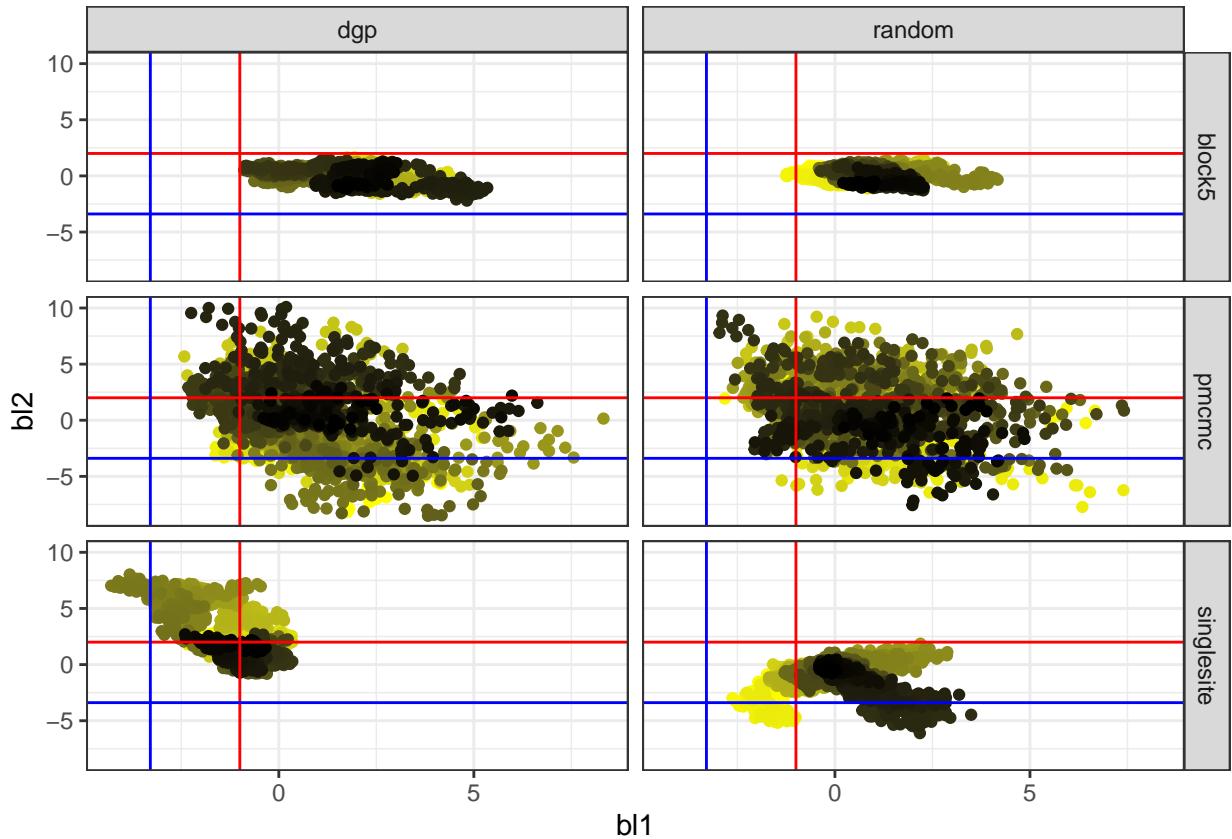
For all the parameters except β_γ^2 , we can see that pmcmc mixes faster than block Gibbs, which mixes better than single-site Gibbs. Despite the smallest auto correlation for β_γ^2 reported by the blocked Gibbs sampler, when we inspect the histograms and pair-wise marginals, we can see the β_γ^2 has been trapped inside a local mode by the blocked Gibbs sampler.

Pair-wise correlations

It is also interesting to examine the posterior correlation between pairs of parameters, for example (β_0^1, β_0^2) , $(\beta_\lambda^1, \beta_\lambda^1)$ and $(\beta_\gamma^1, \beta_\gamma^2)$. We can plot the pairwise marginals here. For the purpose of visualization, I have thinned the samples by displaying a sample in every 50 iterations. The coloring of the samplers indicates the iteration at which the sample is obtained.

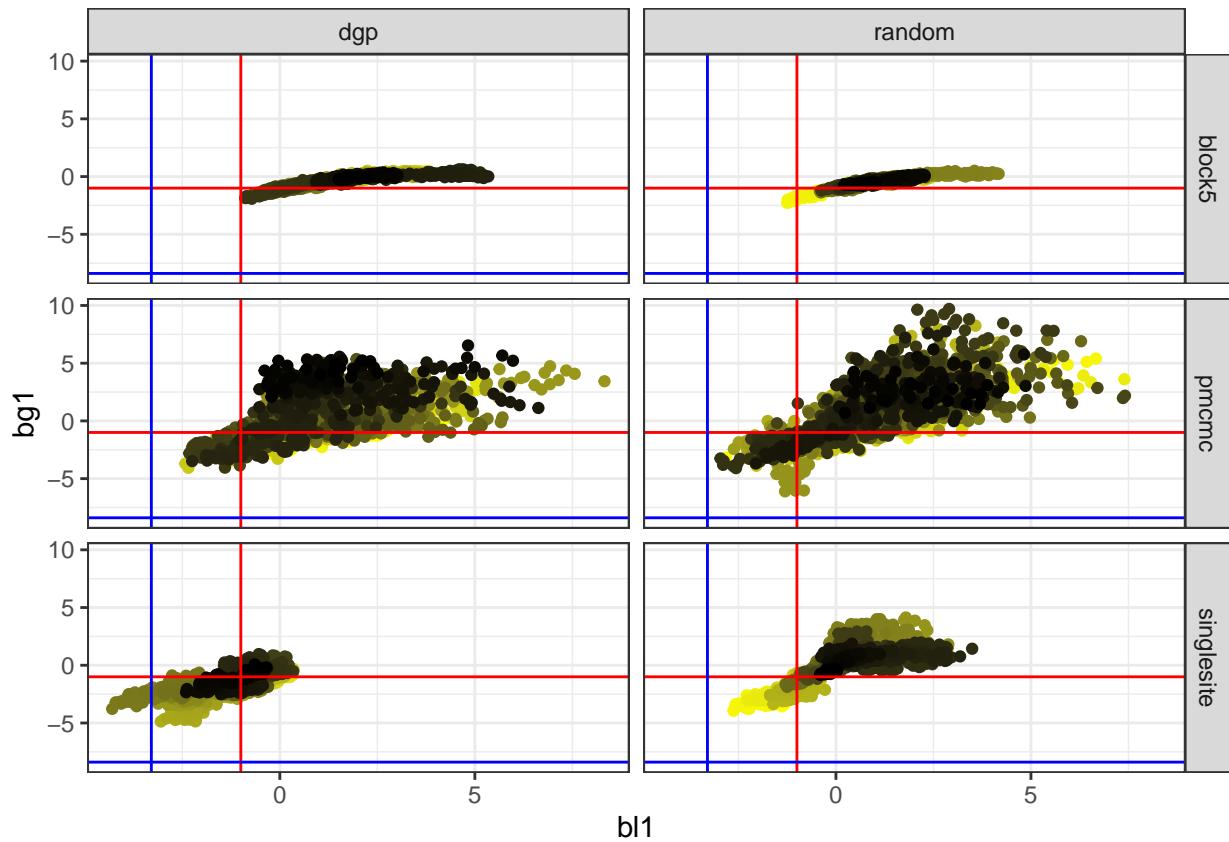
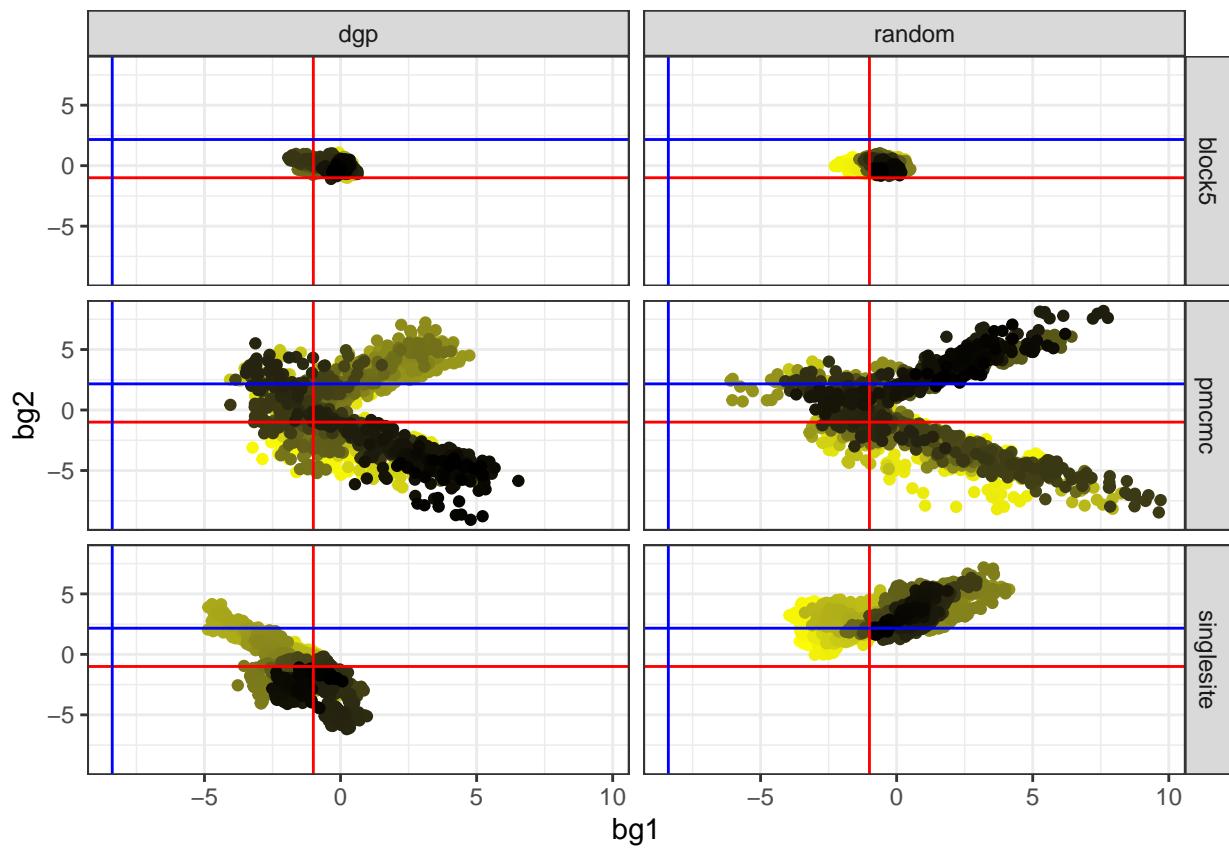


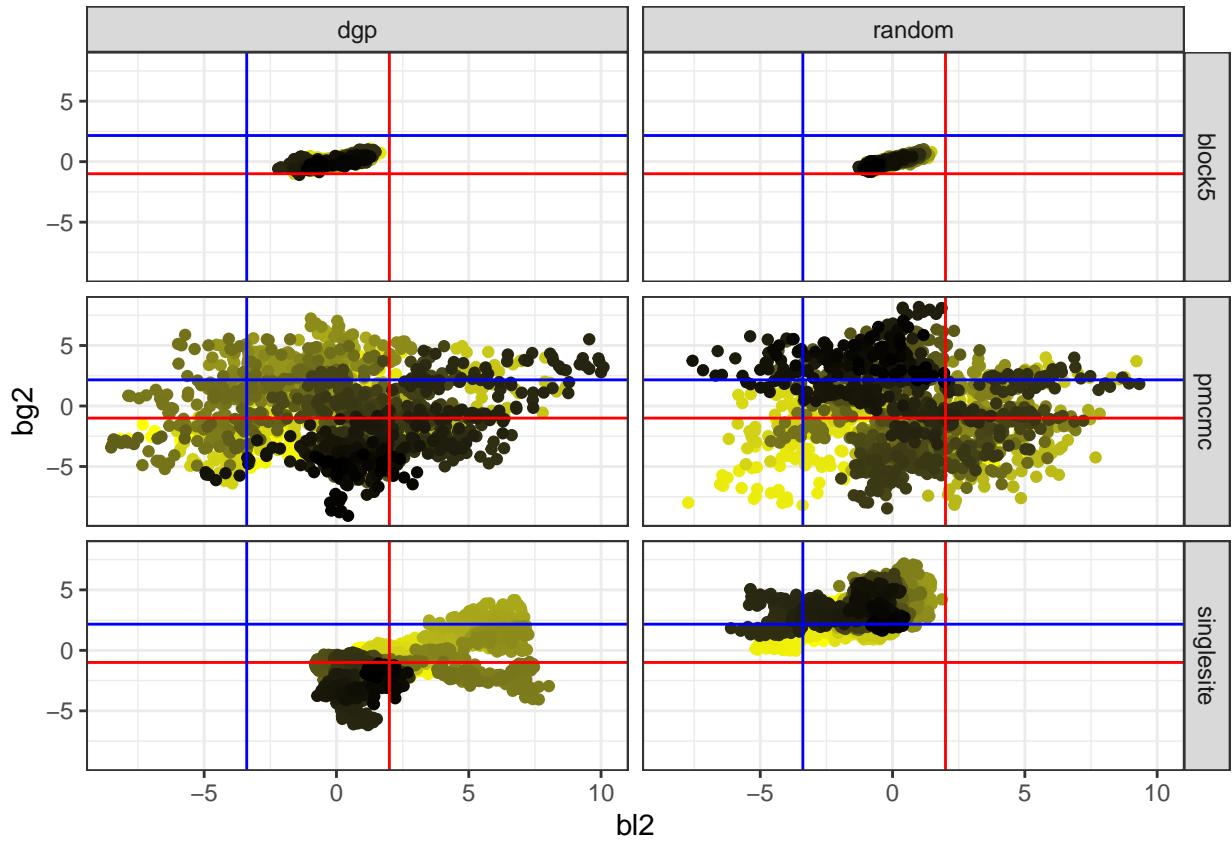
As we have seen in the marginal histograms, the posterior of β_0 agrees from the three algorithms.



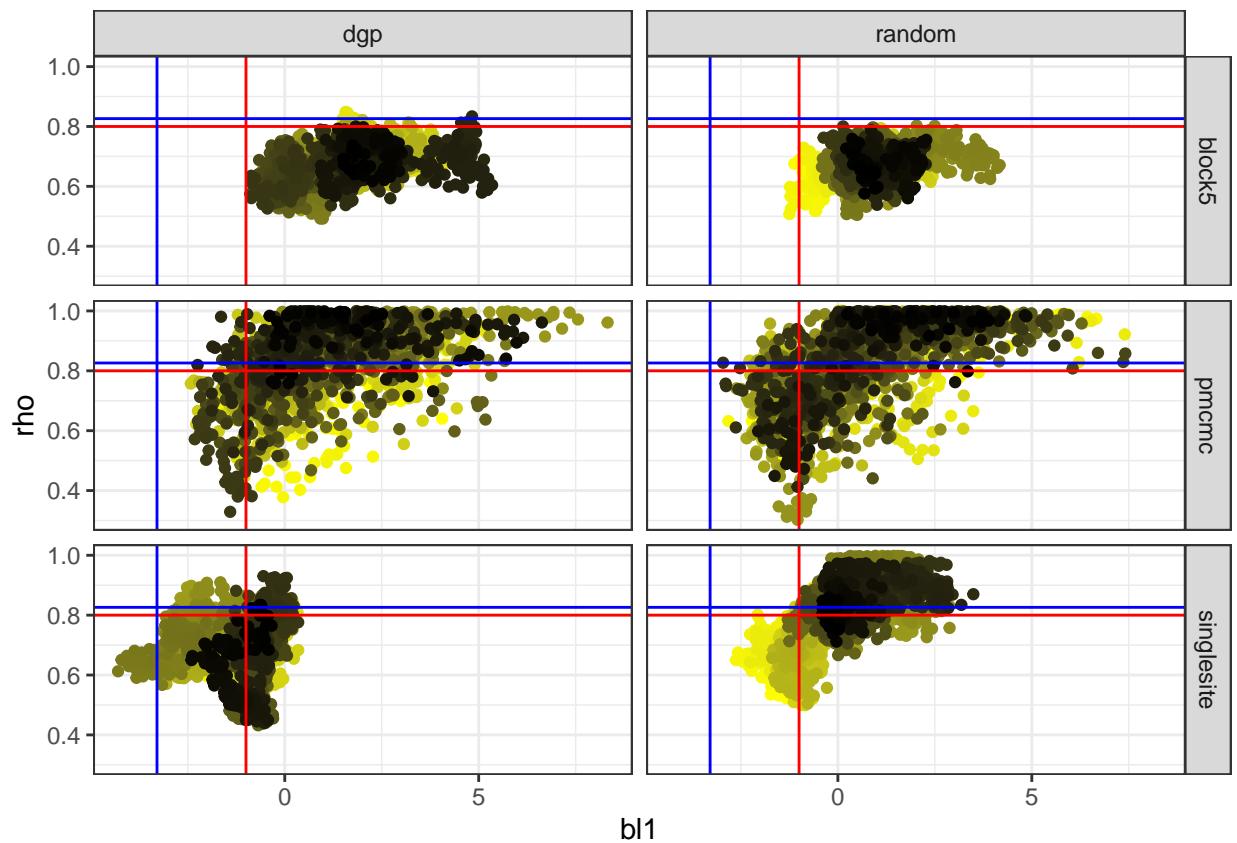
These pairwise scatter plots reveals a curious phenomenal unseen from the single parameter histograms. In particular, single-site Gibbs never visits an area with $\beta_\lambda^1 \approx -2$ and $3 < \beta_\lambda^2 < 5$. There is a white ‘hole’ in that space. Moreover, all three plots exhibit a negative correlation between β_λ^1 and β_λ^2 , despite the Gibbs samplers failing to visit parts of the parameter space. The correlations are printed below.

```
## [the correlation between beta-lambda-1&2 is: -0.3652402 -0.625556 -0.3716069 -0.05397888 -0.02424587
```





The posterior samples from pmcmc reminds us of the heatmaps that we have seen in the particle filtering experiments. The butterfly shape is reminiscent of the ridges from the log-likelihood contour maps. The samples from Gibbs samplers, however, seem to be trapped in some local modes of the posterior due to the irregular shape of the (complete) likelihood.



Finally we decide to look at the pairwise marginal distributions on (β_λ^1, ρ) .

```
## [the correlation between beta-lambda-1 and rho is: 0.4412121 0.1482701 0.4780492 0.5960698 0.6673217
```

Although it is unclear to me why the baseline infection rate and the reporting rate should be positively related conditioning on the data, this positive correlation seems to be consistent among the three samplers.