

Estimate density with Gibbs Sampling

Wei Liu

May 4, 2009

Abstract

Implement Gibbs Sampling in R; Draw samples from bivariate distribution with known conditional distribution but with marginal distribution unknown. Compute the variance of the sample mean, sample variance of the samples generated by Gibbs sampling. Estimate probability density function by Gibbs sampling and compare three methods.¹.

1 Gibbs Sampling and Monte-Carlo Estimate of the variance

As a variant of Metropolis-Hastings sampling, Gibbs Sampling are often used for multivariate distribution sampling, where we know conditional distribution but do not know the joint distribution. Often the conditional distribution is in simple form like χ^2 or normal distribution, so we can have all random variable fixed but only one, and draw samples from the conditional distribution $f(x_i|x_1, x_2 \dots x_{i-1}, x_{i+1} \dots x_d)$.

One can use a single long Gibbs sequence. After some ‘burn-in’ period (will take about this later), say 2,000 sampling from x_1 to x_d , where $d = 2000$, the Markov-Chain is believed to be stationary, and we accept all m samples from x_{d+1} to x_{d+m} . The samples may still be dependent, but the empirical distribution of these samples converges to $f(x)$ [1]. In my test I used this method (as method 2 in the slides) to generate samples, and estimate pdf using these samples.

Or we can use multiple Gibbs sequences. For each single Gibbs, we only preserve the final observation of the samples. Because different Gibbs sequences are independent, the samples are accordingly, independent. This method usually takes longer time than single Gibbs sequence, as each sequence need a ‘burn-in’ time. It could be run in parallel to improve the performance. I used this method (as method 1 in the slides) to generate samples and estimate the pdf.

In the experiment of this part, I implement the Gibbs sampling on a bivariate normal distribution with mean μ_1, μ_2 and variance σ_1^2, σ_2^2 , and correlation ρ . And[3]

$$f(x_1|x_2) \sim N\left(\mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2 - \mu_2), (1 - \rho^2)\sigma_1^2\right) \quad (1)$$

$$f(x_2|x_1) \sim N\left(\mu_2 + \frac{\rho\sigma_2}{\sigma_1}(x_1 - \mu_1), (1 - \rho^2)\sigma_2^2\right) \quad (2)$$

Because the sample mean and variance is also random variable and may be inconsistent, I also used monte-carlo method to compute the variance of the sample mean and sample variance. Besides the 2 Gibbs sampling method above, I also draw the sample with another variant of Gibbs Sampling (method 3), and a regular R’s method (the `mvrnorm` in MASS library) for comparison purpose. The results is in table 1.

From talbe 1, it is hard to see which one is better. As these statistics is also random variables, I used Monte-carlo to repeat Gibbs sampling 50 times and compute the variance of these sample statistics, and the results in table 1

¹part of this document came from the slides of the presentation, which I also uploaded with name as ‘final.pdf’.

	μ	σ^2	ρ
multiple Gibbs	-0.022	1.008	-0.74
Single Gibbs with sampling interval	0.025	1.000	-0.76
Single Gibbs with continous sampling	-0.115	1.046	-0.73
R method	0.009	0.947	-0.75
true	0	1	-0.75

Table 1: Sample mean μ , sample variance σ^2 and sample correlation ρ computed from the samples generated by four methods. Single Gibbs with sampling interval means after the ‘burn-in’, we pick out one sample every r th samples, to eliminate the correlation between adjacent samples. R method refer to the *mvnorm* function in R’s MASS lib. The methods is, unfortunately unknown.

	Var(μ)	Var(σ^2)	Var(ρ)
multiple Gibbs	0.009	0.018	0.002
Single Gibbs with sampling interval	0.010	0.021	0.001
Single Gibbs with continous sampling	0.030	0.041	0.005
R method	0.009	0.021	0.002

Table 2: variance of the sample mean, sample variance and correlation between two random variables

From table 1 we can see among the three Gibbs smapling method, multiple Gibbs sequence have least variance on the μ , σ^2 and ρ . This is reasonable because multiple Gibbs have indepent samples, while the other two methods do not. The R’s method have similar variance with multiple Gibbs.

2 Estimate pdf with Gibbs Sampling

Gibbs sampling can be used to estimate the density distribution. Here we do not use the samples $x_1 \dots x_m$ with kernel density estimation. Instead, we can make use of the conditional densities (which we assume is already known). Take bivariates as example, we can use the average of $f(x|y = y^{(i)})$ as a approximation of $f(x)$. Here we use $y^{(i)}$ as the i th observation of random variable y . And the formula is

$$\widehat{f(x)} = \frac{1}{m} \sum_{i=1}^m f(x|y = y^{(i)}) \quad (3)$$

The theory behind this is the marginal distribution can be seen as the expected value of conditional distribution, i.e.

$$f(x) = \int f(x, y) dy = \int f(x|y) \cdot f(y) dy \quad (4)$$

In the experiment of this part, I compare the density functions recovered from samples generated by three emthods (didn’t try the Gibbs sampling with continous sample because of its low performance)

- 1: Generate *i.i.d* samples $x_2(i)$ by equation 1 and 2 from m Gibbs sequences, and compute $f(x_1)$ based on conditional distribution by formula 3.
- 2: Generate samples $x_1^{(i)}$ (correlated) from one long Gibbs sequences by extracting every r th observation, and compute $f(x_1)$ by kernel density estimation. Here I choose r as 1000, which is long enough to eliminate most of the dependence between $x^{(i)}$ and $x^{(i+1)}$.
- 3: Generate samples by *mvnorm* in R (MASS library), and compute $f(x_1)$ by kernel density estimation. For ernel density estimate, I just use the *density* function in R with the default Gaussian kernel.

The estimated pdf is in figure 2. We can see the multiple Gibbs sampling have better estimation than the other two. The overlappd estimated pdf is in 2

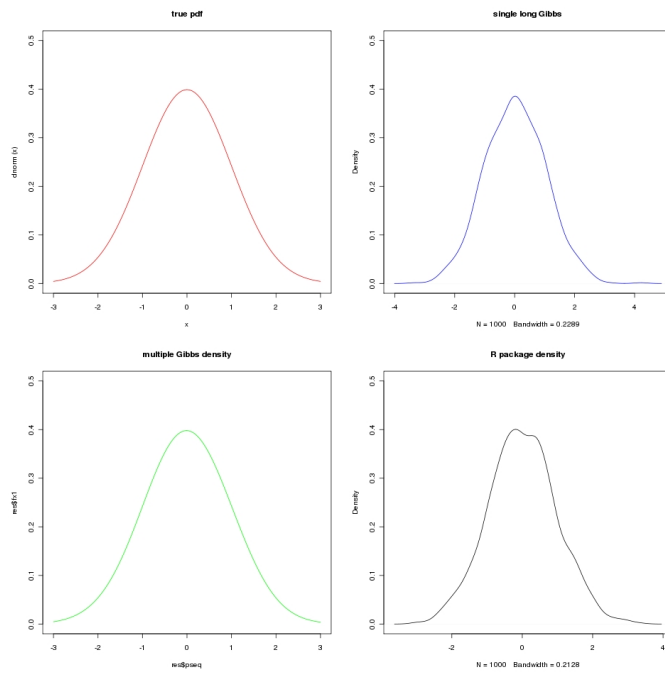


Figure 1: estimated density function by 3 methods: Multiple Gibbs, single Gibbs, and regular R method

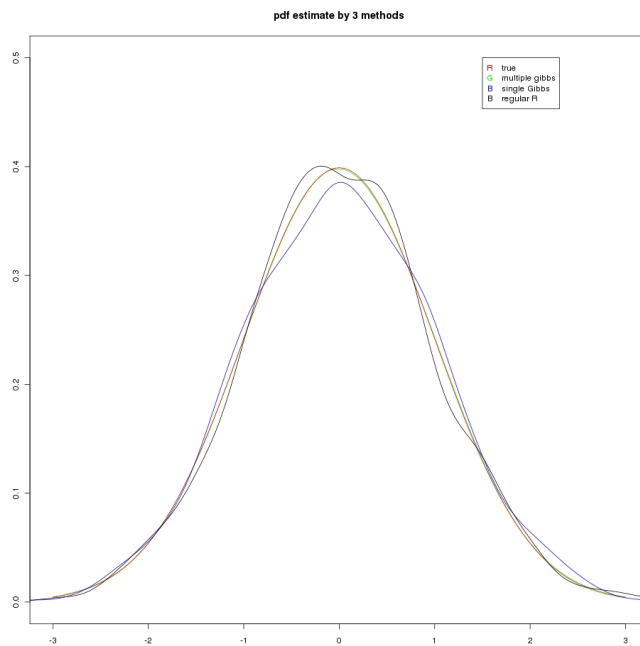


Figure 2: Overlapped density function by 3 methods: Multiple Gibbs (in red color), single Gibbs (in blue color), and regular R method (in black color). The true pdf is also given in green color.

3 Discussion

There are two explanation that Gibbs sampling with conditional distribution is better. One is it makes use of the conditional distribution, which is a analytical form and have more information than just samples. The explanation can also be found in Rao-backwell theorem[2].

A Comments on the R code

The R function `var_est` is the main function for estimating the variance of the sample mean, variance and correlation. It calls `multigibbs`, `longgibbs` and `naivegibbs` function to generate the following three Gibbs sequence: 1) Multiple Gibbs sequence of method one, 2) Single long Gibbs of method 2 (with gaps between samples), and 3) the single Gibbs with continous sampling. The `var_est` also call `regmethod`, which use R's package MASS to generate samples.

The function `pdf_est` is the main function for estimating pdf.

References

- [1] George Casella and Edward I. George, *Explaining the gibbs sampler*, The American Statistician **46** (1992), no. 3, 167–174.
- [2] Alan E. Gelfand and Adrian F. M. Smith, *Sampling-based approaches to calculating marginal densities*, Journal of the American Statistical Association **85** (1990), no. 410, 398–409.
- [3] Maria L. Rizzo, *Statistical computing with R*, Chapman & Hall/CRC, Boca Raton, FL, 2008, ISBN 1-584-88545-9.