Estimate density with Gibbs Sampling

Wei Liu

Scientific Computing and Imaging Institute
University of Utah

weiliu@sci.utah.edu

April 30, 2009

Gibbs sampling refresh

- Introduced by Geman and Geman[3] as a special case of Metropolis-Hastings sampling.
- the random sample is always accepted (i..e $\alpha = 1$).
- Often used to generate multivariate distribution.
- Assume all conditional distributions are available and easy to generate samples from them.
- By sampling from conditioinal distribution, Gibbs sampling get samples from marginal distribution, without having to calculate the density itself. Conditional distribution $p(x|y)$ is easier to compute than integration of joint density $\int p(x,y)dy$

Prodedures of Gibbs Sampling

Consider byvariate random variable (x, y). Assume pdf $f(x|y)$ and $f(y|x)$ are known and can easily generate samples, but the density $f(x)$ and $f(y)$ may be hard to get.

Step 1: Initialized 1st sample y_0

Step 2: Draw x_i and y_i from the conditional distribution as follows:

$$\begin{aligned} x_i &\sim p(x|y = y_{i-1}) \\ y_i &\sim p(y|x = x_i) \end{aligned}$$

Step 3: Accept every candidate from $x_1 \ldots x_N$

Estimate probability density

Gibbs sampling can be used to estimate density by averaging the final conditional densities from multiple Gibbs sequence[1]. For bivariate,

- Generate m Gibbs sequence, and get $y_1 \dots y_m$, the final Y observations from each Gibbs sequence.
- Estimate $f(x)$ by

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

The theory behind this: Because $f(x, y) = f(y) \cdot f(x|y)$, the marginal density $f(x) = \int f(x, y)dy = \int f(x|y) \cdot f(y)dy$

Compare the density

Example: bivariate normal distribution with mean μ_1, μ_2 and variance σ_1^2, σ_2^2, and correlation ρ. And[4]

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

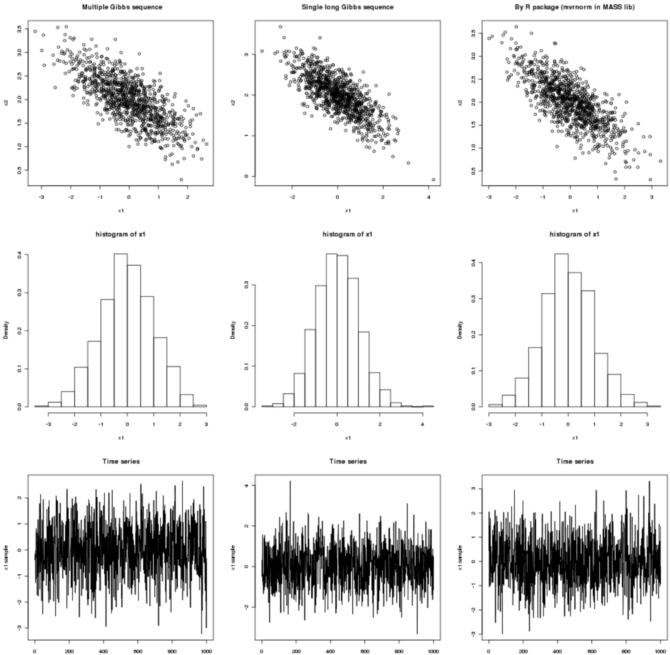
Compare the density functions recovered from samples generated by three emthods.

1: Generate $i.i.d$ samples x_2 from m Gibbs sequences, and compute $f(x_1)$ based on conditional distribution.

2: Generate samples x_1(correlated) from one long Gibbs sequences by extracting every rth observation, and compute $f(x_1)$ by kernel density estimation.

3: Generate samples by *mvrnorm* in R (MASS library), and compute $f(x_1)$ by kernel density estimation.

Test on 3 methods by nenerating 1000 samples

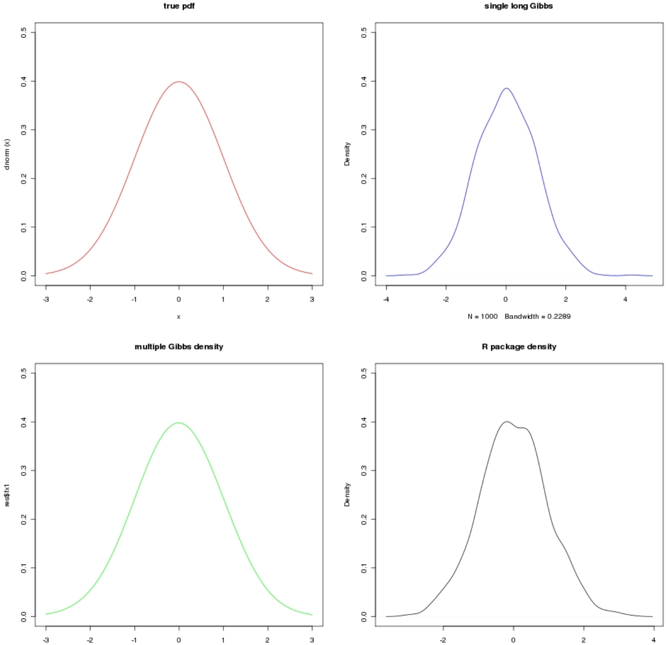


Compare the 3 methods

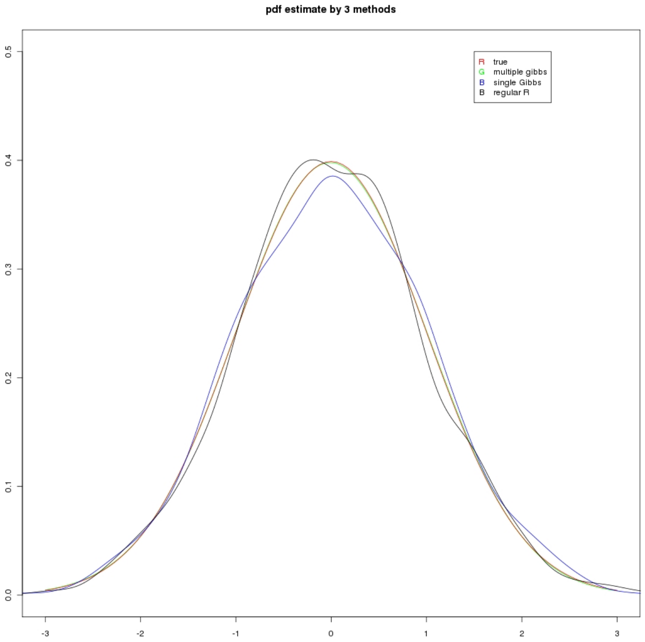
It's hard to see which is best from only sample mean and variance.

	μ	σ^2	ρ
multiple Gibbs	-0.022	1.008	-0.74
Single Gibbs	0.025	1.000	-0.76
R method	0.009	0.947	-0.75
true	0	1	-0.75

estimated pdf with 3 methods



estimated pdf with 3 methods



Discussion

The $f(x|y_i)$ have more information about $f(x)$ than $x_1 \dots x_N$ alone, and will get better estimates. The intuition behind this feature is the Rao-Blackwell theorem[2].

- George Casella and Edward I. George, *Explaining the gibbs sampler*, The American Statistician **46** (1992), no. 3, 167–174.
- 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.
- S. Geman and D. Geman, *Stochastic relaxation, gibbs distributions, and the bayesian restoration of images*, (1990), 452–472.
- Maria L. Rizzo, *Statistical computing with R*, Chapman & Hall/CRC, Boca Raton, FL, 2008, ISBN 1-584-88545-9.