GIBBS SAMPLER

Tomáš Karel – Dagmar Blatná

Abstract

The Gibbs sampler represents a modern statistical computing method used in theoretical and applied work. It is one of the most frequently used Markov chains. Most applications of the Gibbs sampler occur in Bayesian statistics, but it is extremely useful in frequentist (classical) models as well. Using the Gibbs sampler we can simulate the posterior probability distribution without having to express its density in the analytical form. If we can express the posterior distribution of individual parameters estimated using the conditional probability distributions, then we can gradually generate values of the individual parameters from the conditional distributions. It is the indirect way to simulate the posterior distribution of cluster, without analytical calculation and explicit expression of the probability density. This process was a great contribution to the development of the present Bayesian statistical methods. The paper aims at explaining the usage of the Gibbs sampler and at the same time introducing various methods of obtaining the required information from simulations. The problem is illustrated on the beta-binomial distribution.

Key words: Gibbs sampling, posterior probability distribution, bayesian computing method

JEL Code: C11, C15

Introduction

The Gibbs sampler is the computing statistical method used to simulate the posterior probability distribution in Bayesian statistics without having to express its density in analytical form. It is one of the most frequently used Markov chains (Cappe, 2000). The basic idea is quite elementary. We can express the posterior distribution of the individual variables (in Bayesian statistics parameters) estimated by the conditional probability distributions sampling random numbers from the known conditional distributions. It represents the indirect way how to simulate the posterior distribution without its analytical and explicit expression. The development of this computing method presented a great step in the development of Bayesian statistical methods, which has started in early seventies of the last century. This period was connected with the rapid development of computer technology. It was a breakpoint in overcoming one of the major criticism of Bayesian methods, especially of its computational cost. Practical applications of the Gibbs sampling do not mean the progress only in the field of Bayesian statistics (Greenberg, 2008). It means a great benefit for classical (frequentist) statistics as well (Casella & George, 1992). Thanks to this method we are able to get characteristics from the unknown marginal distribution using conditional probabilities. Large number of examples and applications of using the Gibbs sampler are described in Tools for Statistical Inference (Taner, 1991).

1. Illustration of Gibbs sampling

This chapter illustrates and describes the Gibbs sampler in a few illustrative examples. The Gibbs sampler is based on the procedure of Markov chains. Consider the joint probability density. We are interested in numerical characteristics (such as mean, variance) of marginal probability distributions of the individual variables

$$f(x) = \int_{y_1} \dots \int_{y_p} f(x, y_1, y_2, \dots, y_p)\; dy_1 \dots\, dy_p. \tag{1}$$

The most regular way is numerical calculation of an integral and subsequent analytical expression of the searched density of probability. If the expression of this integral is too complicated, we can use the Gibbs sampling method. We are able to make a random selection $X_1, X_2, \dots, X_n$ from the unknown marginal density $f(x)$ without the need of its analytical expression. If the quantity of generated numbers from the known conditional distributions is large enough, we are able to get characteristics (average, mode, median) and variability (standard deviation, variance). To get the sufficient characteristics the sample of random generated values n has to be large enough. To estimate the E(X) as the average value of the generated values the

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i, \tag{2}$$

by the same way we are able to estimate the variability D(X)

$$D(X) = \int_{-\infty}^{\infty} \left[x - E(X)\right]^2 f(x) dx = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{\infty} \left[x_i - \bar{x}\right]^2. \tag{3}$$

In economic practice, hundreds, or thousands of variables could come to the model. With such a large number of variables it is almost impossible (or at least very difficult) to analytically express searched marginal density $f(x)$. For a graphic illustration and understanding the principle of the Gibbs sampler we simplify this multivariate problem to only two dimensions – two variables (*X*, *Y*). Using the Gibbs sampler we are able to sample from the unknown marginal density without its explicitly analytical expression. All we need to know is only the conditional densities $f(x|y)$ and $f(y|x)$. These conditional densities are well known in statistical models of real-world examples (Casella & George, 1992). The process is done by sampling a "Gibbs sequence" of random numbers

$$(Y_0', X_0'; Y_1', X_1'; Y_2', X_2'; ...; Y_k', X_k'). \tag{4}$$

The first value of the Gibbs sequence (4) is specified a priori $Y_0' = y_0'$ another values of this sequence (4) are obtained iteratively alternating generating random numbers from conditional densities

$$\begin{aligned} &Y_0' = y_0' \\ &X_0' \sim f(x|Y_0' = y_0') \\ &Y_1' \sim f(y|X_0' = x_0') \\ &X_1' \sim f(x|Y_1' = y_1') \\ &\ldots \\ &Y_k' \sim f(y|X_{k-1}' = x_{k-1}') \\ &X_k' \sim f(x|Y_k' = y_k'). \end{aligned} \tag{5}$$

The distribution of X_k' converges to $f(x)$ as $k \to \infty$. For large enough *k* the final observation in (5) x_k' is effectively a sample point from searched marginal distribution $f(x)$

$$\begin{aligned} &X_k' \sim f(x) \qquad \text{pro } k \to \infty, \\ &X_k' \xrightarrow[\text{distribution}]{} X_k \quad \text{pro } X_k \sim f(x). \end{aligned} \tag{6}$$

To complete the understanding of the Gibbs sampler we illustrate this method on a few elementary examples.

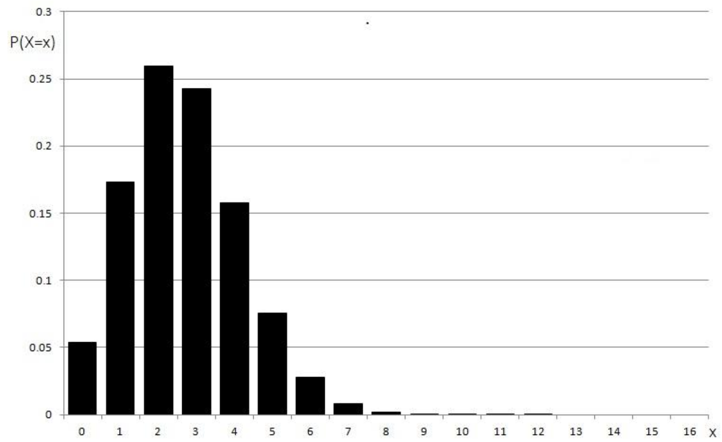
2. The simulation of marginal density

Consider a fair dice. The probability of getting 6 exactly *x* times of *n* independent throws can be obtained from the binomial distribution with parameters $n = 16$ a $\pi = y = 1/6$. Then variable *X* has the binomial probability distribution

$$P(X = x) = \binom{n}{x} y^x (1-y)^{n-x}. \tag{7}$$

Imagine a situation where we throw total 16-times with this fair dice. We are interested in mean (expected) value of getting 6 in the case of 16 independent throws. We can calculate this mean: $\mathrm{E}(X) = ny = 2,667$. Fig. 1 depicts graphical presentation of this discrete probability function.

Fig. 1: Binomial distribution Bi(16; 1/6)



Source: Author´s computations

In the more general situation, the dice could not be fair. In purely theoretical situation the probability of getting 6 could obtain any value from 0 to 1, then assume that

$$f(x;y) \propto \binom{n}{x} y^{x+\alpha-1}(1-y)^{n-x+\beta-1}, \qquad \text{pro } x = 0,1,2,...,n; \quad y \in \langle 0;1 \rangle. \tag{8}$$

We are interested in the marginal distribution $f(x)$. The formula (8) is performed with the following conditional distributions (Cassela, George, 1992)

$$\begin{aligned} f(x|y) &\sim Bi(n;y), \\ f(y|x) &\sim Beta(x+\alpha; n-x+\beta). \end{aligned} \tag{9}$$

To obtain the values from the searched marginal distribution $f(x)$ we can use the Gibbs sampler (11) and iteratively alternating generating random numbers from conditional densities $f(x|y)$ a $f(y|x)$. We can repeat the Gibbs sampling process p times and after that we get p values

$$(X_{k1}', X_{k2}', X_{k3}', ..., X_{kp}') \quad , \tag{10}$$

where

$$X_{ki} \sim f(x) \qquad \text{pro } i = 1,2,3,..., p. \tag{11}$$

In this case the usage of the Gibbs sampler is not necessary. We are able to express the searched marginal density in the analytical form (8) as the beta-binomial distribution

$$f(x)=\binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(\alpha+\beta+n)}\quad \text{pro } x=0,1,2,...,n. \tag{12}$$

Employing (12) we are able to calculate the value of searched characteristics directly. In this case it is not necessary to use an indirect method of calculation. The Gibbs sampling may be indispensable in situations where f(*x,y*), f(*x*), or f(*y*) cannot be calculated. Therefore, in our case where we are able to express the marginal density $f(x)$ analytically as

$$f(x)=\frac{f(x,y)}{f(y|x)}, \tag{13}$$

it is better to obtain the searched characteristics directly from (13) respectively (12). To illustrate the work of the Gibbs sampler we show the calculation of the characteristics mean E(*X*) and variability D(*X*) first directly from the analytical formula (12) and then by using the Gibbs sampler. Suppose 16 independent rolls of the unfair dice again. The probability of getting six on this unfair dice in 16 independent rolls has a beta-distribution with parameters $\alpha=3$ and $\beta=7$. Suppose that these parameters are chosen to represent our prior information about the dice.

The mean E(*X*) of the marginal distribution $f(x)$ can be calculated as

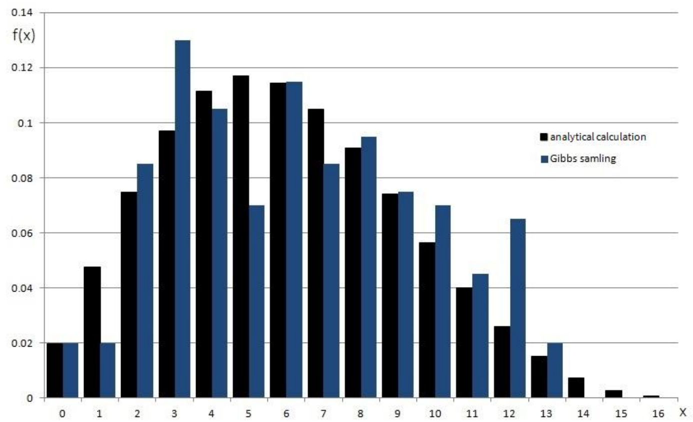
$$\begin{aligned}E(X)&=\sum_{i=1}^{n}x_iP(x_i)=\sum_{i=1}^{n}x_i\binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(\alpha+\beta+n)}=\\&=\sum_{i=0}^{16}x_i\binom{16}{x}\frac{\Gamma(3+7)}{\Gamma(3)\Gamma(7)}\frac{\Gamma(x_i+3)\Gamma(16-x_i+7)}{\Gamma(3+7+16)}=6,00.\end{aligned} \tag{14}$$

The variability D(*X*) can be calculated as

$$\begin{aligned}D(X)&=\sum_{i=0}^{n}\left[x_i-E(X)\right]^2P(x_i)=\sum_{i=0}^{n}\left[x_i-E(X)\right]^2\binom{n}{x_i}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\frac{\Gamma(x+\alpha)\Gamma(n-x+\beta)}{\Gamma(\alpha+\beta+n)}=\\&=\sum_{i=0}^{16}\left[x_i-6,00\right]^2\binom{16}{x_i}\frac{\Gamma(3+7)}{\Gamma(3)\Gamma(7)}\frac{\Gamma(x_i+3)\Gamma(16-x_i+7)}{\Gamma(3+7+16)}=0,03131.\end{aligned} \tag{15}$$

Suppose the situation where the direct calculation of the characteristics of the marginal distribution is not possible. Now the use of the Gibbs sampling may be indispensable. The first graph (Fig.2) shows comparison of marginal density computed directly from the analytical expression (black color) and the marginal density obtained using te Gibbs sampling (blue color). We used the procedure recommended by Gelfand and Smith (1990), which is based on p independent realizations of the Gibbs sampling (in our example we chose $p = 200$ and the step of iterations $k = 10$).

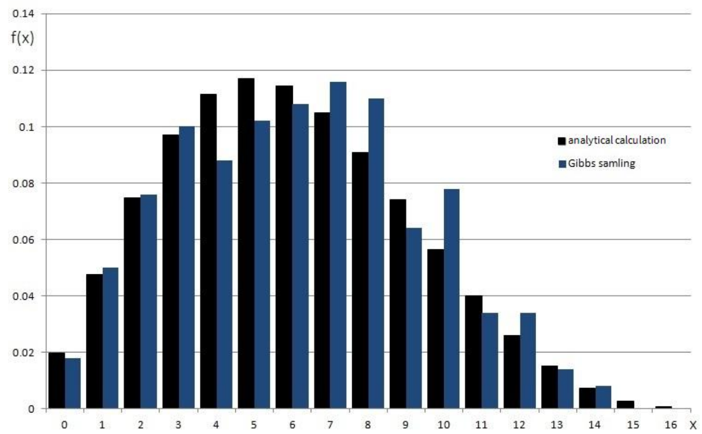
Fig. 2: Gibbs selection $p = 200$; $k = 10$



Source: Author´s computations

Figure 3 shows the simulation the Gibbs sampler with larger number of realizations $p = 500$.

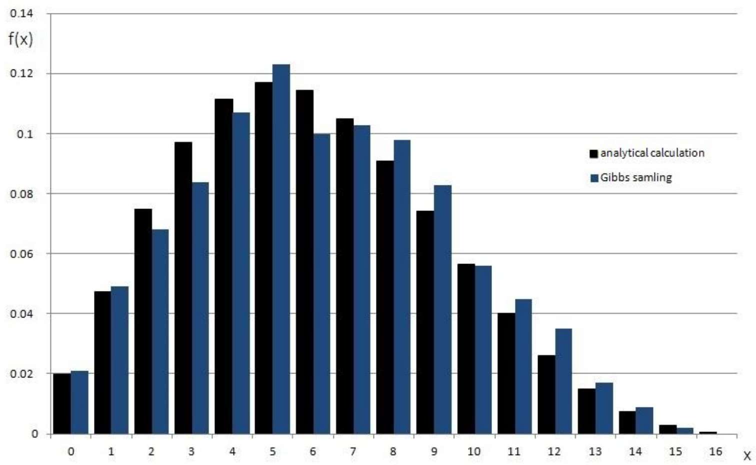
Fig. 3: Gibbs selection $p = 500$; $k = 10$



Source: Author´s computations

In the Figure 4 we can see the increasing number of the Gibbs sampling implementation where $p = 1{,}000$ and $k = 10$.

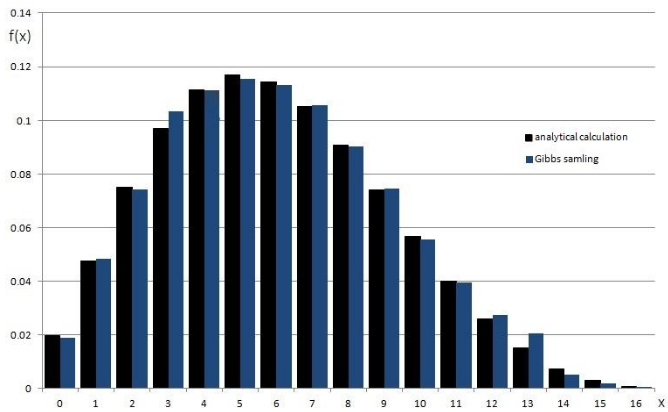
Fig. 4: Gibbs selection $p = 1{,}000$; $k = 10$



Source: Author´s computations

Another method how to get information from the Gibbs sampling design was described by Gelman and Rubin (1991). The point of this method is based on the only one long realization of the Gibbs sampling. In our example we chose k = 30,000. The first 3,000 values are taken out (as a burning process) and from the remaining 27,000 values every 10-th value is selected. These selected values are considered as a random sample from the marginal distribution $f(x)$.

Fig. 5: Gibbs selection *p* = 1; *k* = 30,000



Source: Author´s computations

The characteristics obtained from the simulation described in the previous paragraphs are shown in Tab. 1. Notice how the choice of the step of iteration k and the number of realizations p involves the searched characteristics.

Tab. 1: Characteristic from the Gibbs sampling design

	theoretical distribution	p = 200 k = 10	p = 500 k = 10	p = 1,000 k = 10	p = 1,000 k = 100	k = 30,000 p = 1; m = 10
E(*X*)	6.000	6.305	6.136	6.084	6.049	6.050
D(*X*)	0.03131	0.04415	0.03418	0.03272	0.03172	0.03099

Source: Author´s computations

In two-dimensional cases the calculation of the marginal distribution (except for special cases) is not complicated. The problem comes in the multivariate tasks where the analytical expression of the marginal density is difficult or impossible (due the large number of variables). We can illustrate the relationship between marginal and conditional distributions on the two-dimensional case (Casella, 2001). Consider two random variables X and Y and a known conditional distributions $f(x|y)$ and $f(y|x)$. We search marginal density $f(x)$ which we get as

$$f(x) = \int_Y f(x,y)\, dy. \tag{15}$$

Because the joint density $f(x,y)$ is not known, therefore is possible to use the relationship between the conditional densities $f(x|y)$, $f(y|x)$ and the marginal density $f(x)$

$$f(x) = \int_Y f(x|y) f(y)\, dy, \tag{16}$$

we can also express the unknown marginal density $f(y)$ as in (16)

$$f(y) = \int_X f(y|x) f(x)\, dx \tag{17}$$

and substitute the unknown marginal density in formula (16) with the form (17), then we get

$$\begin{aligned} f(x) &= \int_Y f(x|y) \int_T f(y|t)\, f(t)\, dt\, dy = \\ &= \int_T \left[\int_Y f(x|y) f(y|t)\, dy \right] f(t)\, dt = \\ &= \int_T h(x,t) f(t)\, dt, \end{aligned} \tag{18}$$

where $h(x,t) = \int_Y f(x|y) f(y|t)\, dy$. The marginal density $f(x)$ given in equation (18) is the limiting form of the Gibbs sampling that uses random numbers generated from the conditional distribution of simulated marginal distributions $f(x)$, if $k \to \infty$

$$f_{X_k|X_0'}(x|x_0) \to f_X(x), \tag{19}$$

then

$$f_{X_k|X_{k-1}'}(x|t) \to h(x;t). \tag{20}$$

It is evident that the expression (18) is the limiting form of

$$f_{X_k|X_0'}(x|x_0) \to \int_Y f_{X_k'|X_{k-1}'}(x|t) f_{X_{k-1}'|X_0}(t|x_0) dt. \tag{21}$$

General conditions for convergence of the Gibbs sampler are more described in the work of Roberts and Polson (1990).

3. The methods of obtaining information from the Gibbs selection

The ways how to find the sufficient iteration step *k*, which would guarantee sufficient quality simulation of the searched marginal distribution using the Gibbs sampler (11) are currently discussed (Casella & George, 1992). One of them is used in (1.11) and more described by Gelfand and Smith (Gelfand & Smith, 1990). In this approach we need a set of *p* independent realizations of the Gibbs sequences where every *k*-th value is considered to be a value from the searched marginal distributions $f(x)$. Another approach assumes the long Gibbs sequence (with a lot of iteration steps), where is chosen sufficiently high value of *k*. This value is considered to be the starting point of the simulation. From this point we continue with iterations and select every *m*-th observation. It is necessary because of autocorrelation between neighboring values generated from one the Gibbs sampler process. A potential disadvantage of this approach could be in the movement of random generated values for a long time in a small sample area (Gelman, Rubin, 1991). It is evident that the convergence rate of the Gibbs sampler to the searched marginal distribution depends on the movement of values $X_j^{'}$ throw the sample area. Other methods how to obtain information using the Gibbs sampler describe Muller (1991), Gelman and Rubin (1991), and Tierney (1991) and Tanner (1991).

Conclusion

The Gibbs sampler represents the computing method which makes it possible to simulate the values of unknown marginal distribution which is not possible to express in the analytical form. The Gibbs sampler found an application in the field of Bayesian statistics and also in the works of classical statisticians. For the Bayesian, the Gibbs sampler is mainly used to simulate posterior distributions, whereas for the classical statistician a major use represents calculation of the likelihood function and characteristics of likelihood estimators. Due to the high computational cost, this method has found application in the last few decades with the rapid development of modern computers (Muller, 1991). The illustration examples in this paper show that the convergence to the searched marginal distribution depends on the choose iteration steps *p* and number of the Gibbs sampler simulations *k*. We can see the convergence characteristics got from the simulations to analytically computed ones in the Figures 1 to 5 and then clearly arranged in the Table 1. The ultimate value of the Gibbs sampler lies in its practical potential. We can see the implementation of the Gibbs sampler to general normal data models, to DNA sequence modeling, to applications in HIV modeling problem and e.g. to outlier problems and to logistic regression (Strawderman, 2000). Now that the groundwork has been laid in the pioneering papers of Geman and Geman (1984) and Gelfand and Smith (1990). The research in the area of the Gibbs sampler is still flourishing.

Acknowledgment

The paper has been prepared under the support of the University of Economics, Prague – Internal Grant Agency; project No. IGA 128/2014. "Consequences of assumption violations of classical statistical methods and the possible use of alternative statistical techniques in economic applications".

References

Cappe, O. R. (2000). Markov Chain Monte Carlo: 10 Years and Still Running!. *Journal of the American Statistical Association*.

Casella, G. (2001). Empirical Bayes Gibbs sampling. *Biostatistics*, 485-500.

Casella, G. & George, I, E. (1992). Explaining the Gibbs Sampler. *The American Statistician*, *46*, 167-174.

Gelfan, A., E., & Smith, A. F., (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 398-409.

Gelman, A, & Rubin, D, B.,(1991). An Overview and Approach to Inference From Iterative Simulation. *Technical Report*, University of California-Berkeley, Dept. of Statistics.

Geman, S., & Geman, D. (1984), Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

Greenberg, E. (2008). *Introduction to Bayesian econometrics*. Cambridge: Cambridge University Press.

Müller, P. (1991). *A generic approach to posterior integration and Gibbs sampling*. : Purdue University, Department of Statistics.

Muller, P. *A.* (1991). Generic Approach to Posterior Integration and Gibbs Sampling, *Technical Report* 91-09, Purdue University, Dept. of Statistics.

Roberts, G., & Polson, N. (1990). A Note on the Geometric. Convergence of the Gibbs Sampler, *Technical Report*, University of Nottingham, Dept. of Mathematics.

Strawderman, R. L. (2000). Markov Chain Monte. Carlo: Stochastic Simulation for Bayesian inference. (Review). *Journal of the American Statistical Association*.

Tanner, M. A. (1991). Tools for Statistical Inference, *New York: Springer-Verlag*.
Tierney, L. (1991). Markov Chains for Exploring Posterior Distributions. *Technical Report 560*, University of Minnesota, School of Statistics.

Contact

Tomáš Karel
Department of Statistics and Probability, ,
University of Economic in Prague,
Nám. W. Churchilla 4, Praha 3
tomas.karel@vse.cz

Dagmar Blatná
Department of Statistics and Probability
University of Economic in Prague,
Nám. W. Churchilla 4, Praha 3
blatna@vse.cz