

An Introduction to Gibbs Sampling

Scott W. Burk, Scott & White Health Care, Temple, Texas

Abstract

"Computer-intensive algorithms, such as the Gibbs sampler, have become increasingly popular statistical tools, both in applied and theoretical work." (Casella and George) (1992). The presentation was based upon two papers on Gibbs sampling. The presentation included a basic example of what the Gibbs sampler does and why, with a computer illustration done in SAS/IML. The Gibbs sampler is an iterative technique to compute a marginal based on the conditionals. A sample point from the marginal can be obtained without knowing the marginal itself by sequencing through the conditionals of the distribution. This cycling is a Markov process and requires convergence. This cycling was shown to converge (for the discrete case) using Markov chains.

What is Gibbs sampling?

Gibbs Sampling is an iterative technique that computes sample points from an unknown marginal distribution. The algorithm is very computer intensive and thus with the advent of high speed computers is gaining a large following. It frees statisticians from dealing with calculations, allowing them to focus on statistical aspects of the problem.

Where is Gibbs sampling used?

There are two primary areas where Gibbs sampling is used: 1) Bayesian analysis, primarily to generate posterior distributions and 2) classical analysis, primarily to calculate the likelihood function. The early 90's were boom years for research and application of this technique. I equate the surge of its popularity with the popularity of the bootstrap when it was first introduced by Efron. The Current Index to Statistics lists many references to Gibbs sampling. Some specific theory areas: a) Generalized linear models – Dellaportas and Smith (1990) b) Mixture Models – Diebolt and Robert (1990) c) Evaluating computing algorithms – Eddy and Schervish (1990) d) Normal data models – Gelfand, Hill, and Lee (1990) g) Constrained parameter estimation – Gelfand et al. (1992). Some specific application areas: a) DNA sequence modeling – Churchill and Casella (1991) b) HIV modeling – Lange, Carlin, and Gelfand (1990) c) Supermarket scanner data modeling - Blattberg and George (1991) d) Capture – Recapture modeling - George and Robert (1991). These references are available in Casella and George (1992). At the time, 8 papers were technical reports and 3 were papers in JASA. It is my feeling that the novelty of the subject is the reason not more of these were published papers at the time and by now many have been submitted for publication.

A Brief History of Gibbs Sampler

German S. and German D. (1984) sparked the initial surge of popularity with their work in image-processing models. However; basic roots of the technique can be traced back to at least Metropolis N., Rosenbluth A.W., Rosenbluth N.N., Teller A.H., and Teller E. (1953), with their work in physical chemistry. Further development was made by Hastings W.K. (1970) with some work on Monte Carlo methods using Markov chains. Resurgence of interest was generated by Gelfand and Smith (1990) who showed its potential in a wide variety of conventional statistical problems.

Suppose we are given a joint density, but want characteristics of the marginal distribution. A straightforward approach would be to calculate the marginal directly by integrating

$$f(x) = \int \dots \int f(x, y_1, \dots, y_p) dy_1 \dots dy_p.$$

Then we could solve for the characteristics of interest (mean), etc.). However; the Gibbs approach generates a sample $X_1, \dots, X_M \sim f(x)$ without requiring $f(x)$. Then the population characteristic can be obtained because as the number of sample points from the marginal density goes to infinity the characteristics of the generated density converge to their population analogs with probability one.

An illustration of the Gibbs Algorithm (the two variable case).

Suppose the conditionals of the joint $f(x, y)$ are known. We generate an arbitrary y . We then substitute that y in as the conditioned value in $f(x|Y=y)$ and generate a random x . We then take that x value and substitute that value as the conditioned value in $f(y|X=x)$ and generate a random y . We repeat that process which provided us a Gibbs sequence. In general,

$Y'_0 = y'_0$ is Specified

$X'_j \sim f(x|Y' = y'_j)$

$Y'_{j+1} \sim f(y|X'_j = x'_j)$

The Gibbs Sequence

$Y'_0, X'_0, Y'_1, X'_1, Y'_2, X'_2, Y'_3, X'_3, \dots, Y'_k, X'_k.$

An illustration of the Gibbs Algorithm (the two variable case).

Suppose we have a joint density of the form:

$$f(x, y) \text{ proportional to } \left(\frac{n}{x}\right) y^{x+\alpha-1} (1-y)^{n-x+\beta-1}, x = 0, 1, \dots, n; 0 \leq y \leq 1$$

We want to get the marginal of x and know the conditionals are:

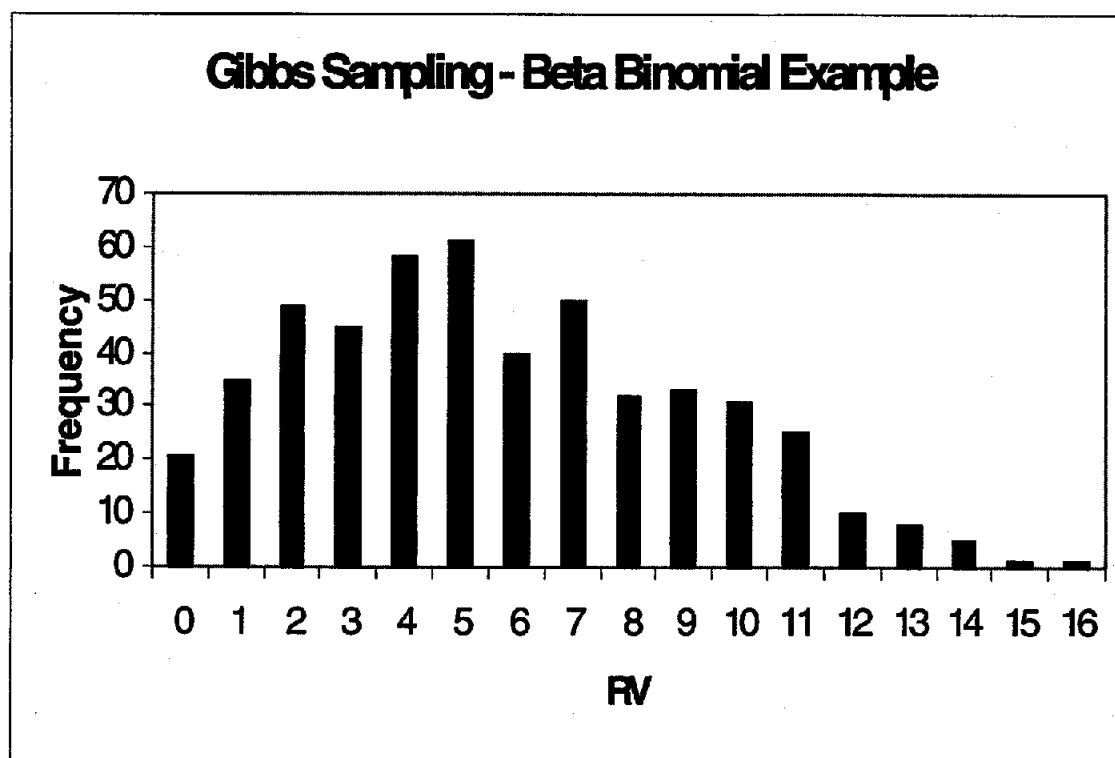
$f(x|y)$ is Binomial (n, y)

$f(y|x)$ is Beta ($x+\alpha, n-x+\beta$)

Of course, one could directly compute the marginal for this example and analytically (the beta-binomial). In fact, it is possible to get an analytical solution for any bivariate when the joint can be obtained. However the following gives a nice, easy to follow illustration. With the following SAS/IML code we generate 500 observations ($m=500$) from the conditional distributions. The parameters are: $n=16$, $\alpha=2$, and $\beta=4$. The Gibbs sequence is length $k=10$.

```
PROC IML;
ALPHA=2;; BETA=4;
M=500; K=10; N=16;
DATAVEC = J(M,1,0);
DO I=1 TO M;
  Y = UNIFORM (0);
  DO J=1 TO K;
    X = RANBIN (0,N,Y);
    ALPHA1 = X + ALPHA;
    BETA1 = N - X + BETA;
    SEEDIT = UNIFORM (0);
    Y = BETAINV (SEEDIT, ALPHA1, BETA1);
  END;
  DATAVEC [I,1]=X;
END;
PRINT DATAVEC;
RUN;
QUIT;
```

The result of this program is:



Which is very close to the Beta-Binomial (the appropriate marginal) for this parameterization. Note that even in the bivariate case, sometimes it is impossible to calculate the joint or the marginals. A case is give in Casella and George (1992) where Gibbs sampling is indispensable.

Convergence of the Gibbs Sequence.

Suppose X and Y are each Bernoulli random variables with the following joint distribution:

		X	
		0	1
Y	0	p_1	p_2
	1	p_3	p_4

$0 \leq p_i, \quad p_1 + p_2 + p_3 + p_4 = 1$

Of course, the joint is a multinomial and may also be represented by:

$$\begin{bmatrix} f_{x,y}(0,0) & f_{x,y}(1,0) \\ f_{x,y}(0,1) & f_{x,y}(1,1) \end{bmatrix} = \begin{bmatrix} p_1 & p_2 \\ p_3 & p_4 \end{bmatrix}$$

Thus the marginal for y is:

$$=[f_Y(0) \ f_Y(1)] = [p_1 + p_2 \ p_3 + p_4]$$

And the conditionals are:

$$A_{Y|X} = \begin{bmatrix} \frac{p_1}{p_1 + p_3} & \frac{p_3}{p_1 + p_3} \\ \frac{p_2}{p_2 + p_4} & \frac{p_4}{p_2 + p_4} \end{bmatrix}$$

$$A_{X|Y} = \begin{bmatrix} \frac{p_1}{p_1 + p_2} & \frac{p_2}{p_1 + p_2} \\ \frac{p_3}{p_3 + p_4} & \frac{p_4}{p_3 + p_4} \end{bmatrix}$$

These conditionals can be viewed as transition matrices (i.e. the probability of going from an x state to a y state or visa versa). To go from $X_0' \rightarrow X_1'$ we must go through Y_1' so the iteration sequence $(X_0' \rightarrow Y_1' \rightarrow X_1', X_0' \rightarrow X_1')$ forms a Markov Chain with transition probability

$$P(X_1' = \chi_1 | X_0' = \chi_0) = \sum_y P(X_1' = \chi_1 | Y_1' = y) P(Y_1' = y | X_0' = \chi_0)$$

The transition probability matrix of the X' sequence is

$$A_{X|X} = A_{Y|X} A_{X|Y}$$

The transition matrix that gives $P(X_k' = \chi_k | X_0 = \chi_0)$ is $(A_{X|X})^k$. And if we write $f_k = [f_k(0) f_k(1)]$ to denote the marginal then for any k

$$f_k = f_0 A_{k|x}^k = (f_0 A_{k|x}^{k-1}) A_{k|x} = f_{k-1} A_{k|x}. \quad (1)$$

Since all the entries of $A_{k|x}$ are positive then (1) implies for any initial probability f_0 , as $k \rightarrow \infty$, f_k converges to a unique distribution f ,

$$f A_{k|x} = f.$$

As $k \rightarrow \infty$, the distribution of X_k gets closer to f_x . And we get

$$f_x A_{k|x} = f_x A_{k|y} A_{k|y|x} = f_x.$$

Conditionals Determine Marginals (The Bivariate Case)

Consider the bivariate case of determining the marginals from the joint:

$$\begin{aligned} f_x(x) &= \int f_{XY}(x, y) dy \\ &= \int [f_{XY}(x|y) f_Y(y)] dy \\ &= \int \left\{ f_{XY}(x|y) \left[\int f_{Y|X}(y|t) f_X(t) dt \right] \right\} dy \\ &= \left[\int f_{XY}(x|y) f_{Y|X}(y|t) dy \right] f_X(t) dt \\ &= \int h(x, t) f_X(t) dt \end{aligned}$$

Where (2) is a fixed-point integral equation and is a unique solution. This is shown in Gelfand and Smith (1990). And since (2) is the limiting form of the Gibbs iteration scheme so as $k \rightarrow \infty$

$$f_{X_k|X_0}(x|x_0) \rightarrow f_X(x)$$

and

$$f_{X_t|X_{t-1}}(x|t) \rightarrow h(x,t)$$

Therefore the Gibbs sequence converges to the point integral equation and its unique solution is a sample point from the marginal of x . We have demonstrated this for the 2X2 case, but the algebra works for any NXM case. Furthermore, Casella and George (1992) point out that with suitable assumptions all the theory goes through for the continuous case and the Gibbs sampler still produces a sample from the marginal distribution of x .

An Illustration of the Gibbs Algorithm (the three variable case).

Suppose we now want the marginal of a trivariate distribution. Of course

$$f_x(x) = \int \int \int f_{XYZ}(x|y,z) f_{YZ|X}(y,z|t) dy dz \Big| f_x(t) dt$$

But if we do not know the joint and we have the conditionals we would sample iteratively from $f_{X|YZ}, f_{Y|XZ}, f_{Z|XY}$. The j th iteration would then be

$$X'_j \sim f(x|Y'_j = y'_j, Z'_j = z'_j)$$

$$Y'_{j+1} \sim f(y|X'_j = x'_j, Z'_j = z'_j)$$

$$Z'_{j+1} \sim f(z|X'_j = x'_j, Y'_{j+1} = y'_{j+1})$$

This would produce the Gibbs Sequence,

$$Y'_0, Z'_0, X'_0, Y'_1, Z'_1, X'_1, Y'_2, Z'_2, X'_2, \dots$$

And for sufficiently large k , this would represent random variables being generated for the marginal distribution of x .

An example of the Gibbs Algorithm (the three variable case).

Suppose we have a joint density of the form:

$$f(x, y, n) \propto \binom{n}{x} y^{x+\alpha-1} (1-y)^{n-x+\beta-1} (e^{-\lambda})^n \left(\frac{\lambda^n}{n!} \right), x = 0, 1, \dots, n; 0 \leq y \leq 1, n = 1, 2, \dots$$

Furthermore, suppose the conditionals are:

$$f(x | y, n) \text{ is Binomial}(n, y)$$

$$f(y | x, n) \text{ is Beta}(x + \alpha, n - x + \beta)$$

$$f(n | x, y) \propto e^{-(1-y)\lambda} \frac{[(1-y)\lambda]^{n-x}}{(n-x)!}, n = x, x+1, \dots$$

Then it is possible to cycle through the conditionals as illustrated above. Note there is no closed form solution for this example and therefore Gibbs or another numerical technique must be used. The result is a distribution similar to the example given before, but is more skewed to the right by the introduction of the Poisson variability. See Casella and George (1992) for the results and nice interpretation.

Gibbs sampling is an increasingly popular technique for generating marginals when 1) the marginal has no closed form 2) when the joint is not known, but the conditionals are 3) when we have high dimensional distributions and analytical solutions are difficult or impossible. The advent of high-speed computers provides the practical application of this technique. We can expect to see application and research of Gibbs sampling exploding in the future.

References

Casella, George; George, Edward, I. (1992), "Explaining the Gibbs Sampler, "The American Statistician August 1992, Vol 46, No. 3, pp 167 – 174.

Gelfand, E.E., Smith, A.F.M., (1990) "Sampling-Based Approaches to Calculating Marginal Densities " *Journal of the American Statistical Association*, 85, 398-409

Gelfand, A.E., Smith, A.F.M., and Lee T.M. (1992), "Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampling, " *Journal of the American Statistical Association*, 87, 523-532

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

Hastings W.K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications, " *Biometrika*, 57, 97-109.

Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., and Teller E. (1953), "Equations of State Calculations by Fast Computing Machines, " *Journal of Chemical Physics*, 21, 1087-1091.

Tanner, M.A. (1991), *Tools for Statistical Inference*, New York; Springer-Verlag

For references of theory and application areas of Gibbs sampling see Casella and George (1992).