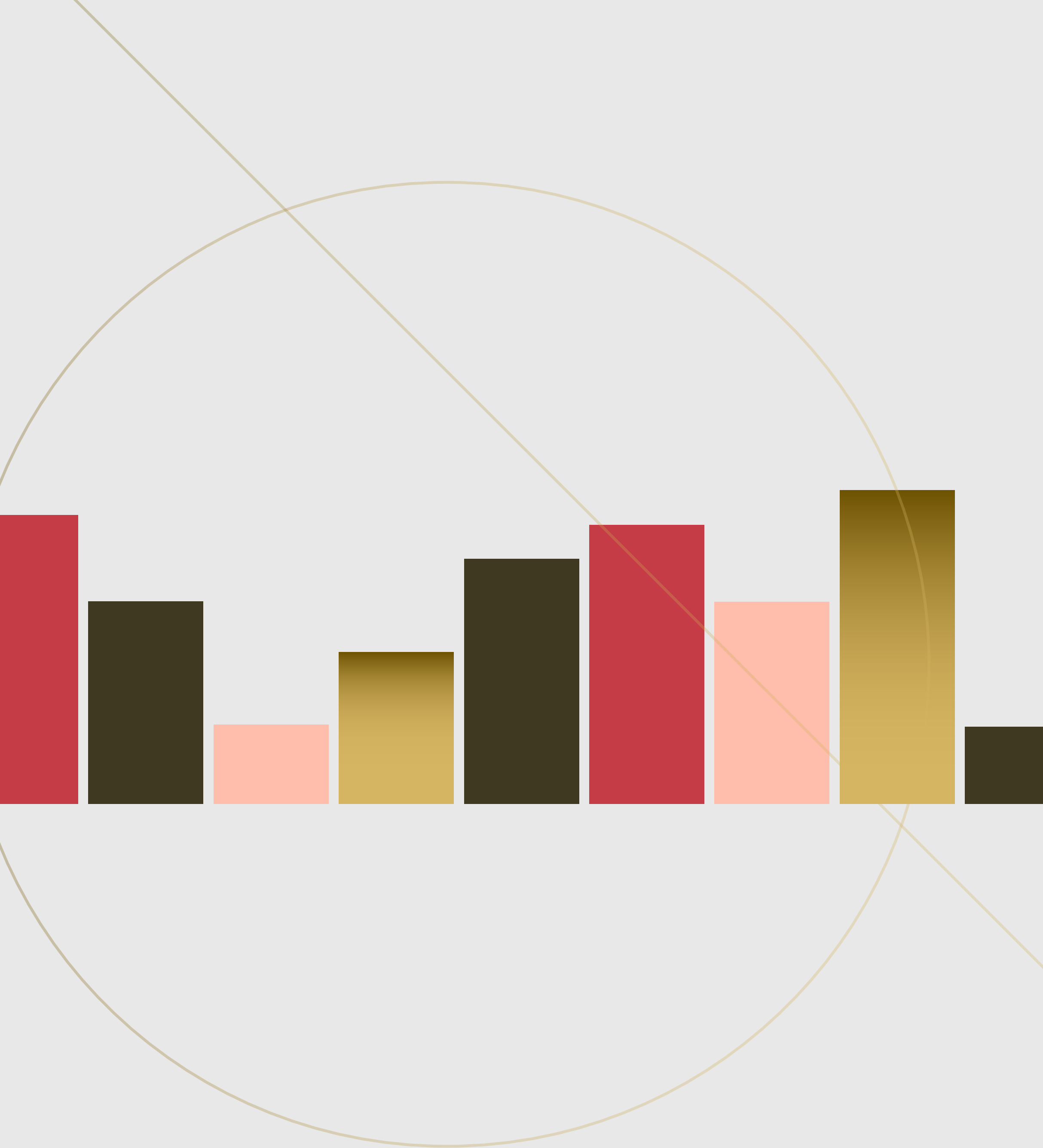

Sampling based approaches to calculating marginal densities

- Gelfand and Smith (1990)

Journal of American Statistical Association

Notations

[ ] means density

[X, Y] - Joint

[X|Y] - conditional

[X] - marginal

$$[X|Y] = \int \underbrace{[X|Y,Z,W]\,[Z|W,Y]\,[W|Y]}_{\text{integration with respect to } Z \text{ and } W}$$

(will read based on LHS)

Substitution or Data Augmentation Algorithm

$$[X] = \int [X|Y] * [Y]\, dy \quad - ①$$

$$[Y] = \int [Y|X] * [X]\, dx \quad - ②$$

Can be interpreted as expectation of conditional w.r.t marginal densities.

Substituting ② in ①

$$[X] = \int \underbrace{[X|Y] \int \underbrace{[Y|X'][X']}_{\text{integration w.r.t } X}}_{\text{integration w.r.t } Y}$$

$$[X] = \int\int [X|Y][Y|X'][X']$$

$$X = \int \overbrace{h(X, X')}^{\text{How you can go from } x' \text{ to } X \text{ using } y} [X']$$

where $$h(X, X') = \int \underbrace{[X|Y][Y|X']}_{\text{integration w.r.t } X'}$$

(though no comments been made in paper)

$X \equiv X'$ which means identical which is a stronger statement than '='.

In the RHS, if we replace $[X']$ by $[X]_i$, then the new fixed point equation will look like

$$[X]_{i+1} = \int h(X, X')[X]_i = \underbrace{I_h}_{\text{integral operator associated with } h.} [X]_i \qquad -(*)$$

Under mild conditions, the following conditions hold for iterations

1) Uniqueness : $[X]$ is unique solution to $(*)$.
2) Convergence : For any starting $[X]_0$, the sequence $[X]_1, [X]_2 \cdots$ defined by $[X]_{i+1} = I_h [X]_i$ converges to $[X]$. (We are not commenting on the distribution of $[X_i, Y_i]$, we are only talking about X_i's from each iteration).
3) Rate : $\int |[X]_i - [X]| \longrightarrow 0$

Extending for 3 variable case

$$[X] = \int [X, Z \mid Y][Y]$$
$$[Y] = \int [Y, X \mid Z][Z]$$
$$[Z] = \int [Z, Y \mid X][X]$$

After substitution of [Z] in [Y] and then [Y] in [X]

$$\left.\begin{aligned} [X] &= \int [X, Z \mid Y] \int [Y, X \mid Z] \int [Z, Y \mid X][X] \\ &= \int [X, Z \mid Y] \int [Y, X'' \mid Z] \int [Z, Y \mid X'][X'] \\ &= h(X, X''; X'][X'] \end{aligned}\right\}$$

Not sure

$$[X]_{i+1} = h(X, X'', X')[X]_i$$

Important condition emerge here
There conditional distributions $[U_{r, r \neq s} \mid U_s]$ ∀s, uniquely determines the joint density. [different from convergence condition as it converges to marginal density]

Substitution Sampling

Assumption: $[X \mid Y]$ and $[Y \mid X]$ are available.

Step 1: Draw $X^{(0)}$ from $[X]_0$ - initial distribution
(it can be degenerate distribution as well which means for a univariate, distribution will take a single value)

Step 2: Given $X^{(0)}$, draw $Y^{(1)} \sim [Y \mid X^{(0)}]$ (given).

$$[Y]_1 = \int [Y \mid X][X]_0$$

Step 3 : Draw $X^{(1)} \sim [X \mid Y^{(1)}]$. Thus $[X]_{(1)} = \int [X \mid Y][Y_1]$

$\therefore$ $X^{(1)} \sim [X]_1 = \int h(X, X')[X']_0$ [it will converge to $[X]$ using the conditions]

Repetition of this cycle will give $(Y^{(2)}, X^{(2)}) \dots (Y^{(i)}, X^{(i)})$. [Thus form a procedure]
Repitition of the procedure will give m i.i.d pairs. $(X_j^{(i)}, Y_j^{(i)})$ (we are repeating the whole procedure of let's say n iterations again as whether one procedure of n iterations, X_i's will not be independent draws)

we have independence accross j but we have dependence within j.

Using convergence condition, $X^i \xrightarrow{d} X \sim [X]$
$Y^i \xrightarrow{d} Y \sim [Y]$

If we use $Y_j^{(i)}$ which is i^{th} iteration from the j^{th} procedure for $j = 1, 2 \dots m$, we can estimate the marginal of $[X]$

$$\therefore [\hat{X}]_i = \frac{1}{m} \sum_{j=1}^{m} [X \mid Y_j^{(i)}]$$

Three variable case

Assume : $[X, Y \mid Z]$, $[Y, X \mid Z]$, $[Z, X \mid Y]$ are available

$$\begin{array}{l}{[X]} =\\ {[Y]} =\\ {[Z]} =\end{array}\begin{array}{c}\boxed{\begin{array}{l}\int [X,Y|Z][Z]\\ \int [Y,Z|X][X]\\ \int [X,Z|Y][Y]\end{array}}\\ \hookrightarrow \text{order 1}\end{array}\begin{array}{l}=\\=\\=\end{array}\begin{array}{c}\left.\boxed{\begin{array}{l}\int [X,Z|Y][Y]\\ \int [X,Y|Z][Z]\\ \int [Z,Y|X][X]\end{array}}\right\}\\ \hookrightarrow \text{order 2}\end{array}\begin{array}{l}\text{integration w.r.t}\\ \text{variables not}\\ \text{present in LHS}\end{array}$$

Step 1 : Draw $X^{(0)} \sim [X]_0$

Step 2 : Draw $(Z^{(0)'}, Y^{(0)'}) \sim [Z, Y | X^{(0)}]$

$\therefore \; [Z, Y]_0' = \int [Z, Y | X][X]_0$

Step 3 : Draw $(Y^{(1)}, X^{(0)'}) \sim [Y, X | Z^{(0)'}]$

Step 4 : Draw $(X^{(1)}, Z^{(1)}) \sim [X, Z | Y^{(1)}]$

Repeat each cycle i times to produce $(X^{(i)}, Y^{(i)}, Z^{(i)})$.

$(X^{(i)} \to X \sim [X], \; Y^{(i)} \to Y \sim [Y]$ and $Z^{(i)} \to Z \sim [Z])$

Repeat the entire process j times to obtain i.i.d $(X_j^{(i)}, Y_j^{(i)}, Z_j^{(i)})$

$(X_j^{(i)}, Y_j^{(i)}, Z_j^{(i)}) \to [X, Y, Z]$ (joint density)

$\to \underbrace{[X | Y, Z][Y | Z][Z]}$

Density estimator :

$$[\hat{X}]_i = \frac{1}{m} \sum_{j=1}^{m} [X | \underbrace{Y_j^{(i)}, Z_j^{(i)}}]$$

where $Y_j^{(i)}$ should be drawn from $[Y|Z]$ & $Z_j^{(i)}$ should be drawn from $[Z]$.

Let's say if $[X, Z | Y]$ is not available, then it can be sampled from $[X|Z, Y]$ and $[Z|Y]$.
That is, availability for full conditional and reduced conditional.
To do substitution sampling, either we need 3 joint conditional densities, i.e $[Y, X|Z]$, $[Z, X|Y]$ and $[Y, Z|X]$ or six full and partial conditional densities, i.e $[Y|X, Z]$, $[X|Z]$, $[Z|X, Y]$, $[X|Y]$, $[Y|Z, X]$ and $[Z|X]$. [We will explore this more in next section]

Gibbs Sampling

$$[X] = \int [X|Y, Z][Y|Z][Z]$$
$$[Y] = \int [Y|X, Z][Z|X][X]$$
$$[Z] = \int [Z|X, Y][X|Y][Y]$$

6 distributions are needed for substitution sampling. (rarely we have it)

Gibbs sampler do it with only full conditionals

Given arbitrary starting set of values $U_1^{(0)}, U_2^{(0)} \ldots U_k^{(0)}$, we draw

$$U_1^{(1)} \sim [U_1 \mid U_2^{(0)} \cdots U_k^{(0)}]$$
$$U_2^{(1)} \sim [U_2 \mid U_1^{(1)}, \cdots U_k^{(0)}]$$
$$\vdots \quad \vdots \quad \vdots \quad \vdots$$
$$U_k^{(1)} \sim [U_k \mid U_1^{(1)}, \cdots U_{k-1}^{(1)}]$$

Each variable is reached in a natural order.

after i such cycles, we arrive at $(U_1^{(i)}, \cdots U_k^{(i)})$

Under mild conditions, the following holds

1) convergence - $(U_1^{(i)}, \cdots U_k^{(i)}) \to [U_1, U_2 \cdots U_k]$ (joint density)
$U_s^{(i)} \xrightarrow{d} U_s \sim [U_s]$ as $i \to \infty$ $\forall s$ (marginal)

2) rate

3) Ergodic Theorem - $\lim_{i \to \infty} \frac{1}{i} \sum_{l=1}^{i} T(U_1^{(l)} \cdots U_k^{(l)}) \xrightarrow{a.s} E(T(U_1, \cdots U_k))$
$\forall$ measurable function T.

[Unlike substitution sampling, the procedure itself is not repeated]

If the procedure itself is repeated, then the marginal density can be estimated as in the previous section.

Relationship between Gibbs sampling and Substitution Sampling

3 variable case

In case we only have full conditionals, how will we run substitution sampling [instead of Gibbs]

For substitution sampling, we need 6 conditionals.

$[X] = \int [X|Y,Z][Z|Y][Y]$ - a

$[Y] = \int [Y|X,Z][X|Z][Z]$ - b

$[Z] = \int [Z|X,Y][Y|X][X]$ - c

Let's start by estimating $[Y|X]$

form a sub-substituting algorithm

$$\left.\begin{aligned} [Y|X] &= \int [Y|X,Z][Z|X] \\ [Z|X] &= \int [Z|X,Y][Y|X] \end{aligned}\right\}$$ Similar to Data augmentation

We have to draw from full conditionals only to accomodate for reduced conditional densities.

We will do same procedure for $[Z|Y]$ and $[X|Z]$

It will look like this

=> $[X|Z,Y]$ will have a sub loop of $[Z|X,Y]$ and $[X|Z,Y]$ to accomadate for $[Z|Y]$ [$Y^{(0)}$ is needed]

=> $[Y|X,Z]$ will have a subloop of $[X|Z,Y]$ and $[Y|X,Z]$ to accomodate for $[X|Z]$ [$Z^{(0)}$ is needed]

=> $[Z|Y,X]$ will have a subloop of $[Y|X,Z]$ and $[Z|X,Y]$ to accomodate for $[Y|X]$ [$X^{(0)}$ is needed]

Procedure

Fix $x^{(0)}$, $y^{(0)}$ and $z^{(0)}$

Step 1: Draw $Z^{(0)'}$ from $[Z \mid Y^{(0)}, X^{(0)}]$

Step 2: Draw $Y^{(0)'}$ from $[Y \mid Z^{(0)'}, X^{(0)}]$

Step 3: Draw $Z^{(1)}$ from $[Z \mid X^{(0)}, Y^{(0)'}]$

$Y^{(0)'}$ will be used in step 1 in next cycle

$Y^{(0)'}$ is been drawn from loop of full conditionals with $X^{(0)}$ fixed. $\therefore [Y]_0' \xrightarrow{d} Y \mid X$

Step 4: Draw $Y^{(0)''}$ from $[Y \mid X^{(0)}, Z^{(0)}]$

Step 5: Draw $X^{(0)'}$ from $[X \mid Y^{(0)''}, Z^{(0)}]$

Step 6: Draw $Y^{(1)}$ from $[Y \mid X^{(0)'}, Z^{(1)}]$

$X^{(0)'}$ will be used in step 4 in next cycle

$X^{(0)'}$ is been drawn from loop of full conditionals with $Z^{(0)}$ fixed. $\therefore [X]_0'' \xrightarrow{d} X \mid Z$

Step 7: Draw $X^{(0)''}$ from $[X \mid Z^{(0)}, Y^{(0)}]$

Step 8: Draw $Z^{(0)''}$ from $[Z \mid X^{(0)''}, Y^{(0)}]$

Step 9: Draw $X^{(1)}$ from $[X \mid Z^{(0)''}, Y^{(1)}]$

$Z^{(0)''}$ will be used in step 7 in next cycle

$Z^{(0)''}$ is been drawn from loop of full conditionals with $Y^{(0)}$ fixed. $\therefore [Z]_0'' \xrightarrow{d} Z \mid Y$

In the paper, a simplified version is given which is shorter and efficient.

Step 1: Draw $Y^{(0)'}$ from $[Y \mid X^{(0)}, Z^{(0)}]$

Step 2: Draw $Z^{(0)'}$ from $[Z \mid Y^{(0)'}, X^{(0)}]$

Step 3: Draw $X^{(0)'}$ from $[X \mid Y^{(0)'}, Z^{(0)'}]$

Step 4: Draw $Y^{(1)}$ from $[Y \mid X^{(0)'}, Z^{(0)'}]$

Step 5: Draw $Z^{(1)}$ from $[Z \mid Y^{(1)}, X^{(0)'}]$

Step 6: Draw $X^{(1)}$ from $[X \mid Y^{(1)}, Z^{(1)}]$

It looks like 2 iterations of Gibbs sampler.