Closed-form Gibbs Sampling for Graphical Models with Algebraic constraints

Hadi Mohasel Afshar

Scott Sanner

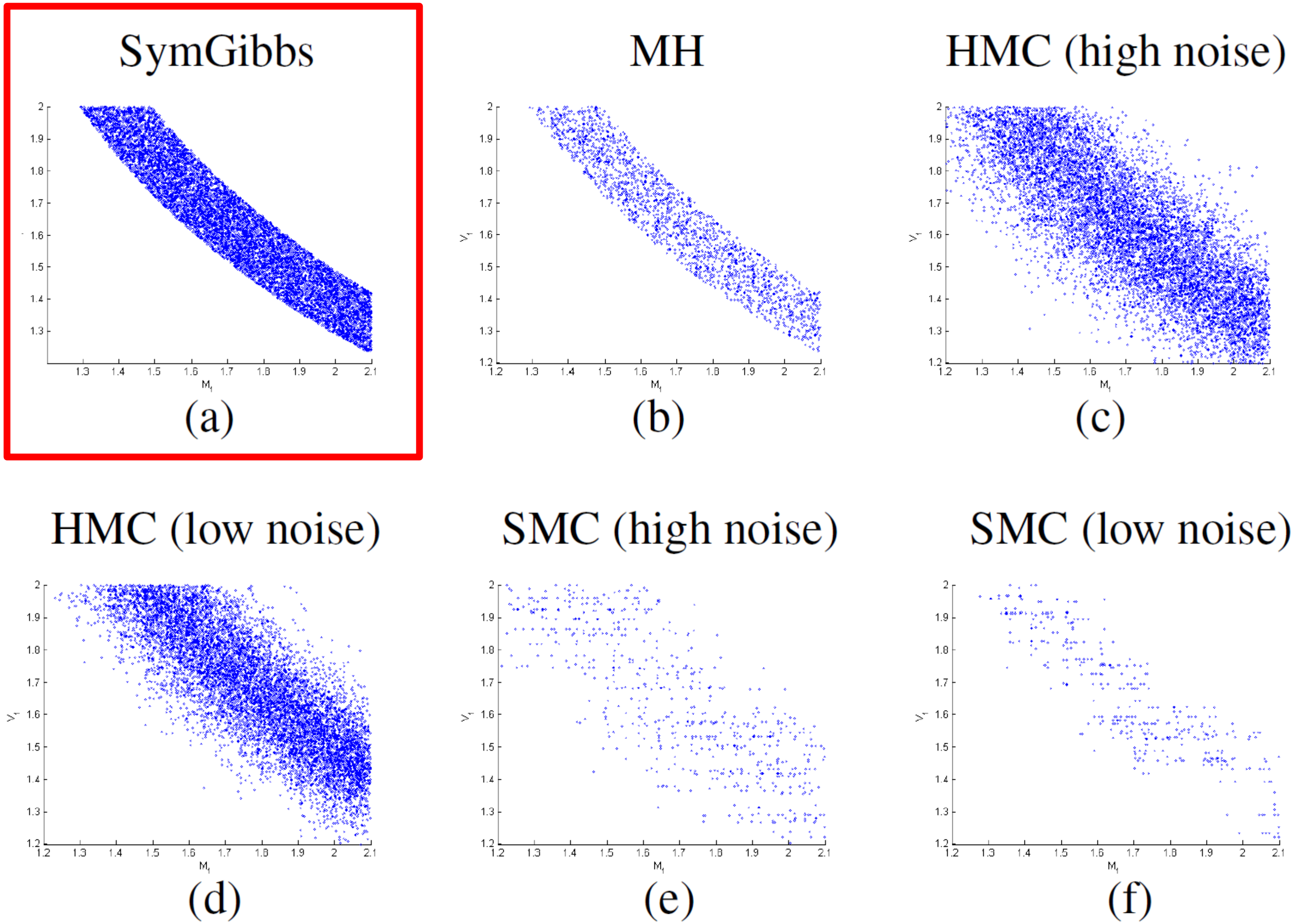
Christfried Webers

Inference in Hybrid Graphical Models / Probabilistic Programs

- Limitations of BUGS, PyMC, Anglican, and STAN
  - They don't handle piecewise functions well
    - i.e., slow convergence with conditionals ***if (t > 0)*** …
  - They don't handle simple algebraic constraints
    - i.e., you cannot assign ***x = y + 1***
      (you have to add noise)

Our solution efficiently handles piecewise functions and algebraic constraints

Sneak Preview: Sampling in Piecewise Models with Algebraic Constraints



Contribution 1:

We present an effective sampler for **GMs with piecewise factors**

Polynomial-piecewise fractional functions (PPFs)

Fractional

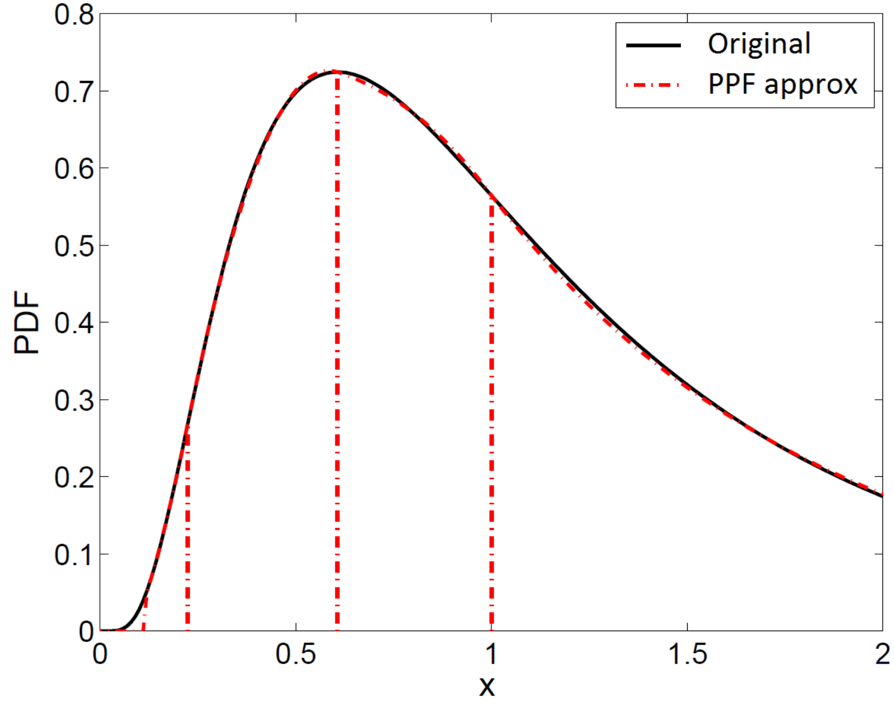
Polynomial

Example:

$$\begin{cases} \frac{x}{y} & if \quad x + y > 0 \\ \frac{x^2+y}{y^2} & if \quad x + y < 0, y > 0 \\ \vdots & \quad\quad \vdots \end{cases}$$

PPFs can be used for:

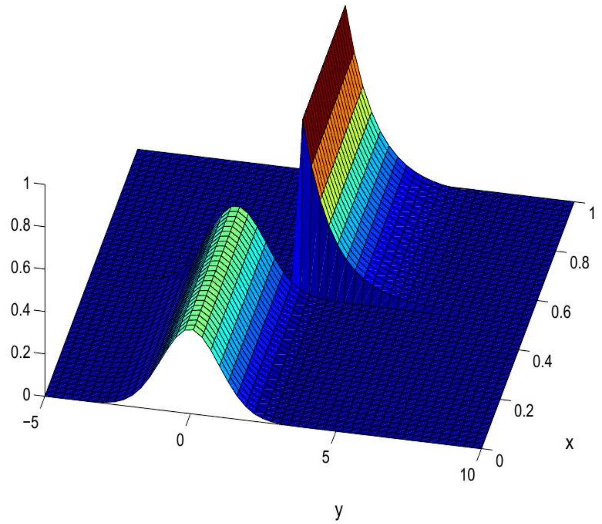
- Representing truncated/finite support models
- Approximating arbitrary models



PPFs can be used for:

- Representing truncated/finite support models
- Approximating arbitrary models
- Probabilistic programming

```
% Draw from uniform (0, 1)
x = rand;
if (x < 0.5)
    % Draw from standard Normal
    y = randn;
else
    % Draw from Gamma(1, 1)
    y = randg + 2.0;
end
```

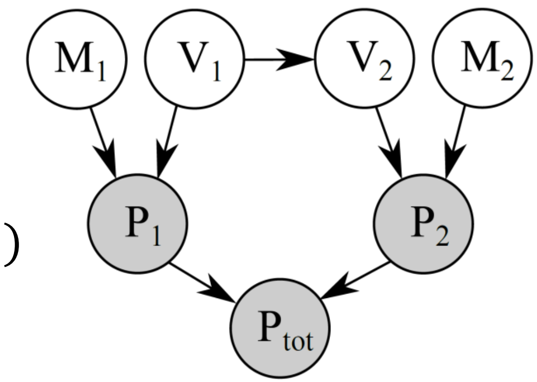


PPFs can be used for:

- Representing truncated/finite support models
- Approximating arbitrary models
- Probabilistic programming
- Bayesian inference: piecewise priors and likelihoods
- Algebraic constraints!

Contribution 2: Algebraic Constraints

- **An example:**



$pr(M_1) = U(M_1; 0.1, 2.1)$ $pr(V_1) = U(V_1; -2, 2)$

$pr(M_2) = U(M_2; 0.1, 2.1)$ $pr(V_2|V_1) = U(V_2; -2, V_1)$

Observation: $P_{tot} = 3$

Query: $pr(V_1, M_2, V_2 | P_{tot} = 3)$?

$M_1V_1 + M_2V_2 = 3$

$$pr(V_1, M_2, V_2 | P_{tot} = 3) \propto \int_{-\infty}^{\infty} pr(M_1, V_1, M_2, V_2) pr(P_{tot} | M_1, V_1, M_2, V_2) dM_1$$

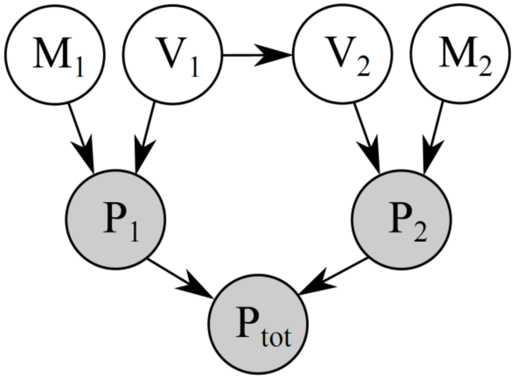
$$\delta(f(x)) = \sum_{r \in roots(f(x))} \frac{\delta(x - r)}{\left|\frac{\partial f(x)}{\partial x}\right|}$$

$P_{tot} = M_1V_1 + M_2V_2$, so:

$$pr(P_{tot}) = \delta(M_1V_1 + M_2V_2 - 3) = \frac{\delta\left(M_1 - \left(\frac{3 - M_2V_2}{V_1}\right)\right)}{|V_1|}$$

Contribution 2: Algebraic Constraints

- **An example:**



$pr(M_1) = U(M_1; 0.1, 2.1) \quad pr(V_1) = U(V_1; -2, 2)$

$pr(M_2) = U(M_2; 0.1, 2.1) \quad pr(V_2|V_1) = U(V_2; -2, V_1)$

Observation: $P_{tot} = 3$

Query: $pr(V_1, M_2, V_2|P$

Divisions, absolute values!
PPFs are closed under them ☺

$pr(V_1, M_2, V_2|P_{tot} = 3) \propto \int_{-}^{\infty}$ $, V_1, M_2, V_2) pr(P$

Next: PPFs are closed under fractional substitutions! ☺

$$\delta(f(x)) = \sum_{r \in roots(f(x))} \frac{\delta(x - r)}{\left|\frac{\partial f(x)}{\partial x}\right|}$$

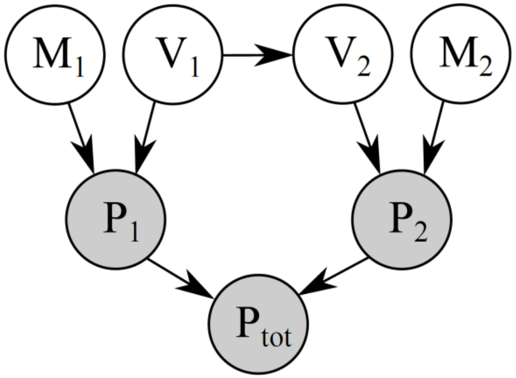
$P_{tot} = M_1V_1 + M_2V_2$, so

$$pr(P_{tot}) = \delta(M_1V_1 + M_2V_2 - 3)$$

$$= \frac{\delta\left(M_1 - \left(\frac{3 - M_2V_2}{V_1}\right)\right)}{|V_1|}$$

Collapse Algebraic Constraints

- **Collapse out M_1:**



$pr(M_1) = U(M_1; 0.1, 2.1) \quad pr(V_1) = U(V_1; -2, 2)$

$pr(M_2) = U(M_2; 0.1, 2.1) \quad pr(V_2|V_1) = U(V_2; -2, V_1)$

Observation: $P_{tot} = 3$ that is, $M_1V_1 + M_2V_2 = 3$

$$pr(P_{tot}|\ldots) = \frac{\delta\left(M_1 - \left(\frac{3 - M_2V_2}{V_1}\right)\right)}{|V_1|}$$

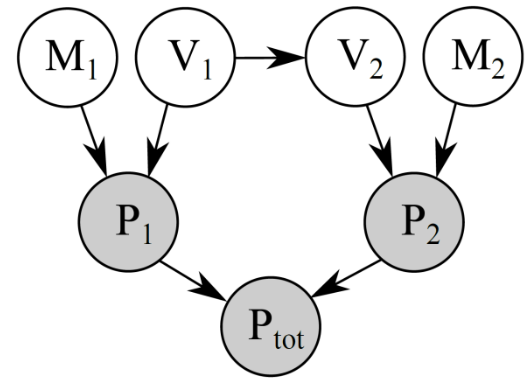
Query: $pr(V_1, M_2, V_2 | P_{tot} = 3)$?

$$pr(V_1, M_2, V_2 | P_{tot} = 3) \propto \int_{-\infty}^{\infty} pr(M_1, V_1, M_2, V_2) pr(P_{tot} | M_1, V_1, M_2, V_2) dM_1$$

$$\propto \begin{cases} \dfrac{1}{V_1(V_1+2)} & if\ 0 < V_1, 0.1 < \dfrac{3 - M_2V_2}{V_1} < 2.1, 1 < M_2 < 3, \\ & -2 < V_1 < 2, -2 < V_2 < V_1 \\ \dfrac{-1}{V_1(V_1+2)} & if\ 0 > V_1, 0.1 < \dfrac{3 - M_2V_2}{V_1} < 2.1, 1 < M_2 < 3, \\ & -2 < V_1 < 2, -2 < V_2 < V_1 \\ 0 & otherwise \end{cases}$$

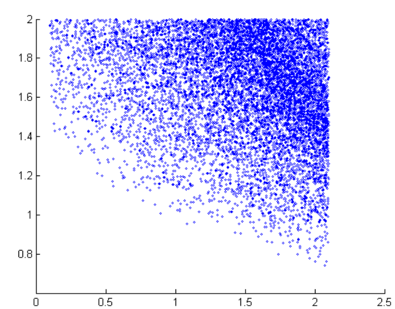
Sneak Preview: Inference Results

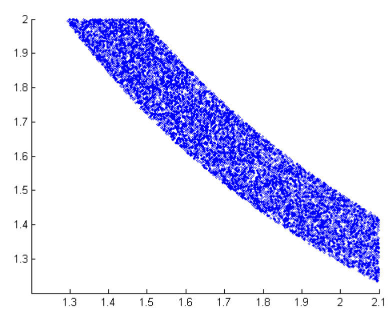
Even in this simple example, posteriors are **multimodal** and **piecewise**!

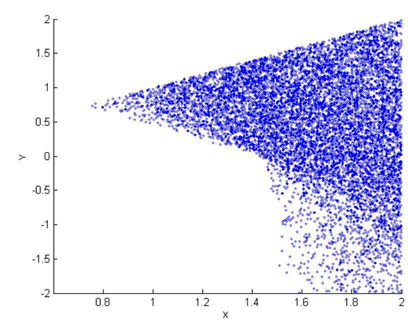


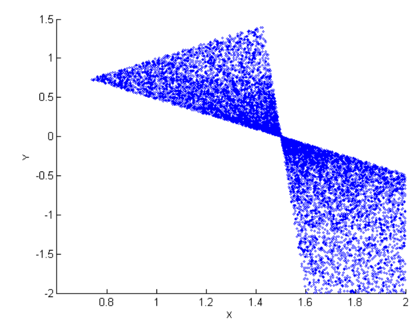
$pr(M_1, V_1 | P_{tot} = 3, V_2 = 0.2)$

$pr(V_1, V_2 | P_{tot} = 3, M_1 = 2)$









$pr(M_1, V_1 | P_{tot} = 3)$

$pr(V_1, V_2 | P_{tot} = 3)$

Where are we?

- We’ve written down expressive PPF models
- We’ve collapsed out determinism (constraints)
- We still need to do inference in the collapsed PPF model…

Inference in Piecewise Algebraic Models

- Closed-form solution? **Generally, impossible! ☹**
- Metropolis Hastings? **Low acceptance rate! ☹**

Due to high KL-divergence between the proposal and target densities

Inference in Piecewise Algebraic Models

- Closed-form solution? Generally, impossible! ☹
- Metropolis Hastings? Low acceptance rate! ☹
- Hamiltonian Monte Carlo? Low acceptance rate! ☹

Also no discrete variables!

Since HMC leap-frog mechanism relies one the assumption of smoothness.

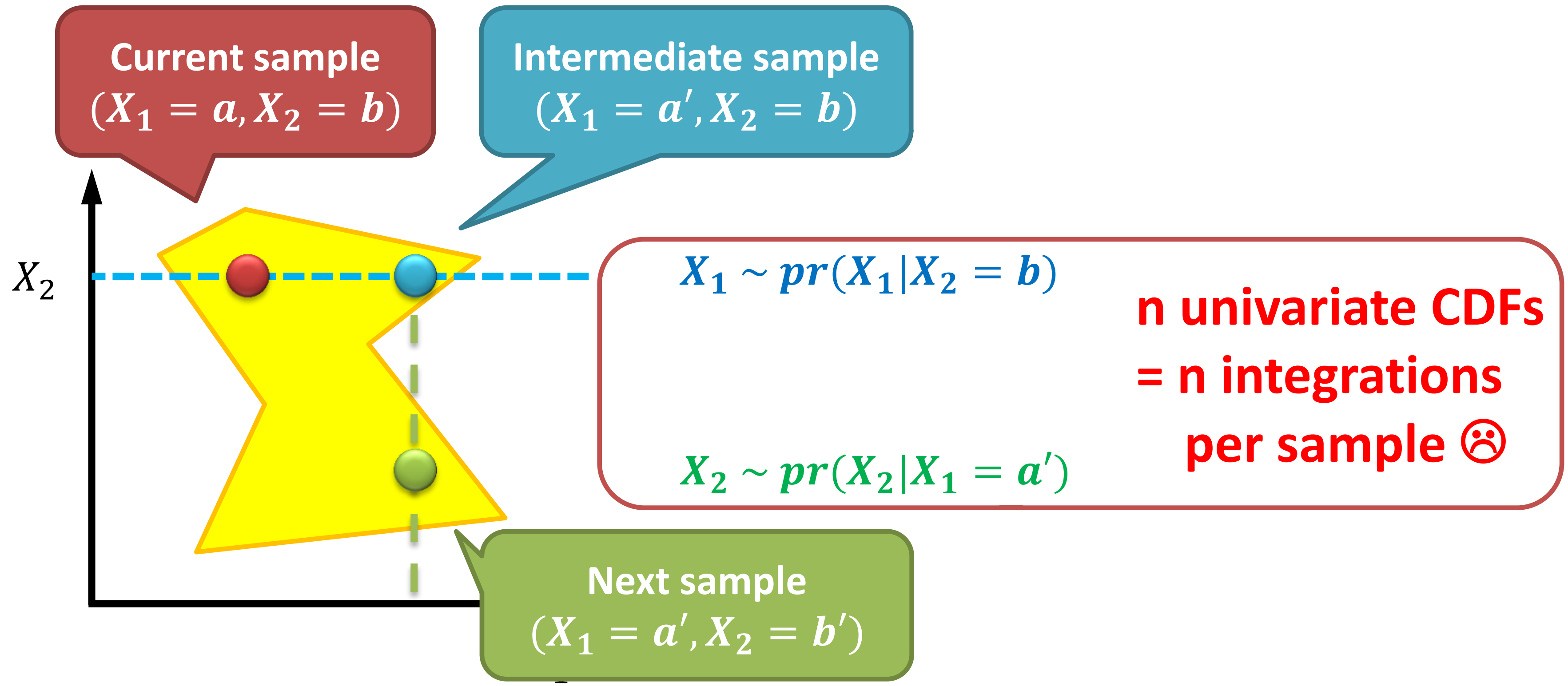
Inference in Piecewise Algebraic Models

- Closed-form solution? Generally, impossible! ☹
- Metropolis Hastings? Low acceptance rate! ☹
- Hamiltonian Monte Carlo? Low acceptance rate! ☹
- Slice Sampling? Poor performance on multimodal densities! ☹
- Gibbs sampling? Slow, due to per sample (multiple) CDF computation (integration)! ☹

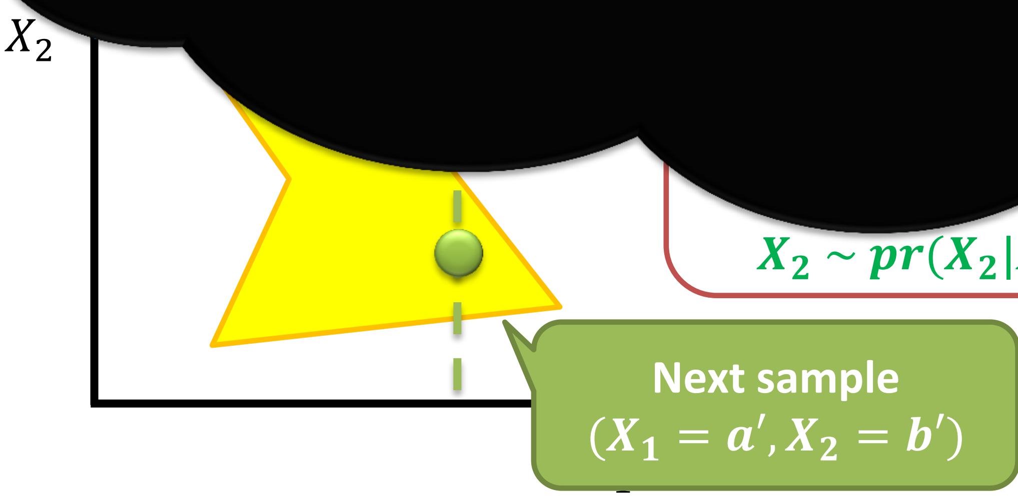
We are going to make it fast! ☺

Gibbs sampling

- Remember that in Gibbs, sampling from an **n** dimensional function is done in **n** steps.



IDEA!
What if we can compute CDFs symbolically and prior to sampling?



$X_2 \sim pr(X_2 | X_1 = a')$

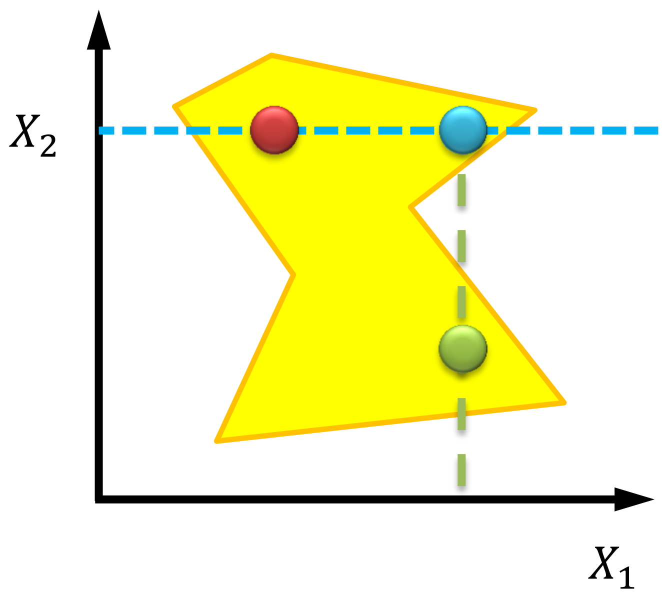
variate CDFs
tegrations
sample ☹

Gibbs sampling

- Remember that in Gibbs, sampling from an n dimensional function is done in n steps.

Only need to do one integral which is possible for a large class of PPFs

Mapping **variables** to symbolic **CDFs**



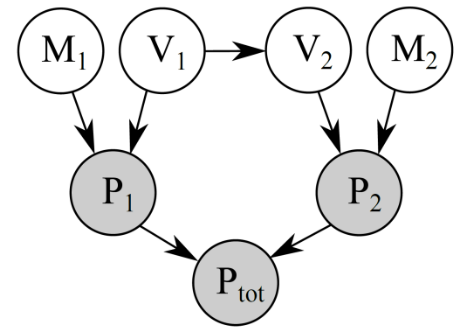
$$X_1 \longrightarrow F_{X_1}(x) := \int_{-\infty}^{x} P(X_1 = t, X_2, \ldots, X_n)dt$$

$$\vdots$$

$$X_n \longrightarrow F_{X_n}(x) := \int_{-\infty}^{x} P(X_1, \ldots, X_{n-1}, X_n = t)dt$$

n integrals rather than **n**× **#samples**

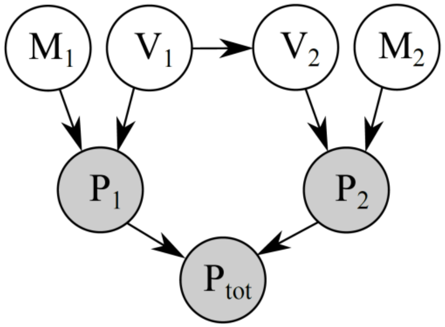
Returning to our example:



$$pr(V_1, M_2, V_2 | P_{tot} = 3) \qquad \propto$$

$$\begin{cases} \dfrac{1}{V_1(V_1+2)} & if\ 0 < V_1, V_1 < 30 - 10M_2V_2, \dfrac{3 - M_2V_2}{2.1} < V_1, 1 < M_2 < 3, \\ & -2 < V_1, V_1 < 2, -2 < V_2, V_2 < V_1 \\ \dfrac{-1}{V_1(V_1+2)} & if\ 0 > V_1, V_1 > 30 - 10M_2V_2, \dfrac{3 - M_2V_2}{2.1} > V_1, 1 < M_2 < 3, \\ & -2 < V_1 < 2, -2 < V_2 < V_1 \\ 0 & otherwise \end{cases}$$

Returning to our example:



$$V_1 \xrightarrow{maps} F_{V_1}(v_1)$$

$$= \int_{V_1=-\infty}^{v_1} pr(V_1, M_2, V_2 | P_{tot} = 3)\ dV_1 \propto$$

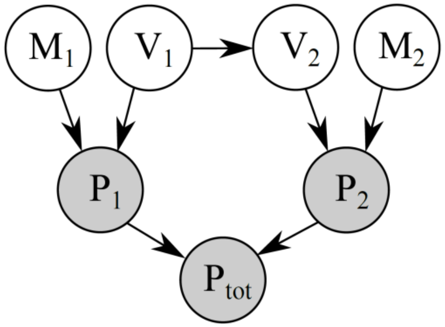
Let's just consider one statement

$$\int_{V_1=-\infty}^{v_1} \left(\frac{1}{V_1(V_1+2)} \quad \begin{array}{c} if\ 0 < V_1, V_1 < 30 - 10M_2V_2, \dfrac{3 - M_2V_2}{2.1} < V_1, 1 < M_2 < 3, \\ -2 < V_1, V_1 < 2, -2 < V_2, V_2 < V_1 \end{array} \right) dV_1$$

$$+$$

$$\int_{V_1=-\infty}^{v_1} \left(\begin{array}{cc} \dfrac{-1}{V_1(V_1+2)} & \begin{array}{c} if\ 0 > V_1, V_1 > 30 - 10M_2V_2, \dfrac{3 - M_2V_2}{2.1} > V_1, 1 < M_2 < 3, \\ -2 < V_1 < 2, -2 < V_2 < V_1 \end{array} \\ 0 & otherwise \end{array} \right) dV_1$$

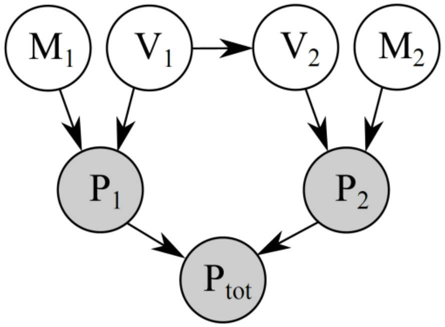
Returning to our example:



Let's just consider one statement

$$\int_{V_1=-\infty}^{v_1} \left(\frac{1}{V_1(V_1+2)} \quad if\ 0 < V_1, V_1 < 30 - 10M_2V_2, \frac{3-M_2V_2}{2.1} < V_1, 1 < M_2 < 3, \atop -2 < V_1, V_1 < 2, -2 < V_2, V_2 < V_1 \right) dV_1$$

Returning to our example:



$$\int_{V_1=-\infty}^{v_1} \left(\frac{1}{V_1(V_1+2)} \quad if\ 0 < V_1, V_1 < 30 - 10M_2V_2, \frac{3 - M_2V_2}{2.1} < V_1, 1 < M_2 < 3, \; -2 < V_1, V_1 < 2, -2 < V_2, V_2 < V_1 \right) dV_1$$

$$= \begin{cases} \displaystyle\int_{V_1=\max\{0, V_2, \frac{3-M_2V_2}{2.1}\}}^{\min\{v_1, 30-10\,M_2V_2,\, 2\}} \frac{1}{V_1(V_1+2)} dV_1 & if\ \min\{v_1, 30 - 10\,M_2V_2, 2\} > \max\left\{0, V_2, \frac{3 - M_2V_2}{2.1}\right\}, \; 1 < M_2 < 3, -2 < V_2 \\ 0 & otherwise \end{cases}$$

A large set of algebraic functions have closed-form indefinite integrals i.e. here $\int \frac{dV_1}{V_1(V_1+2)} = \frac{\log(V_1) - \log(V_1+2)}{2}$

Inference in Piecewise Algebraic GMs

1. Collapse determinism

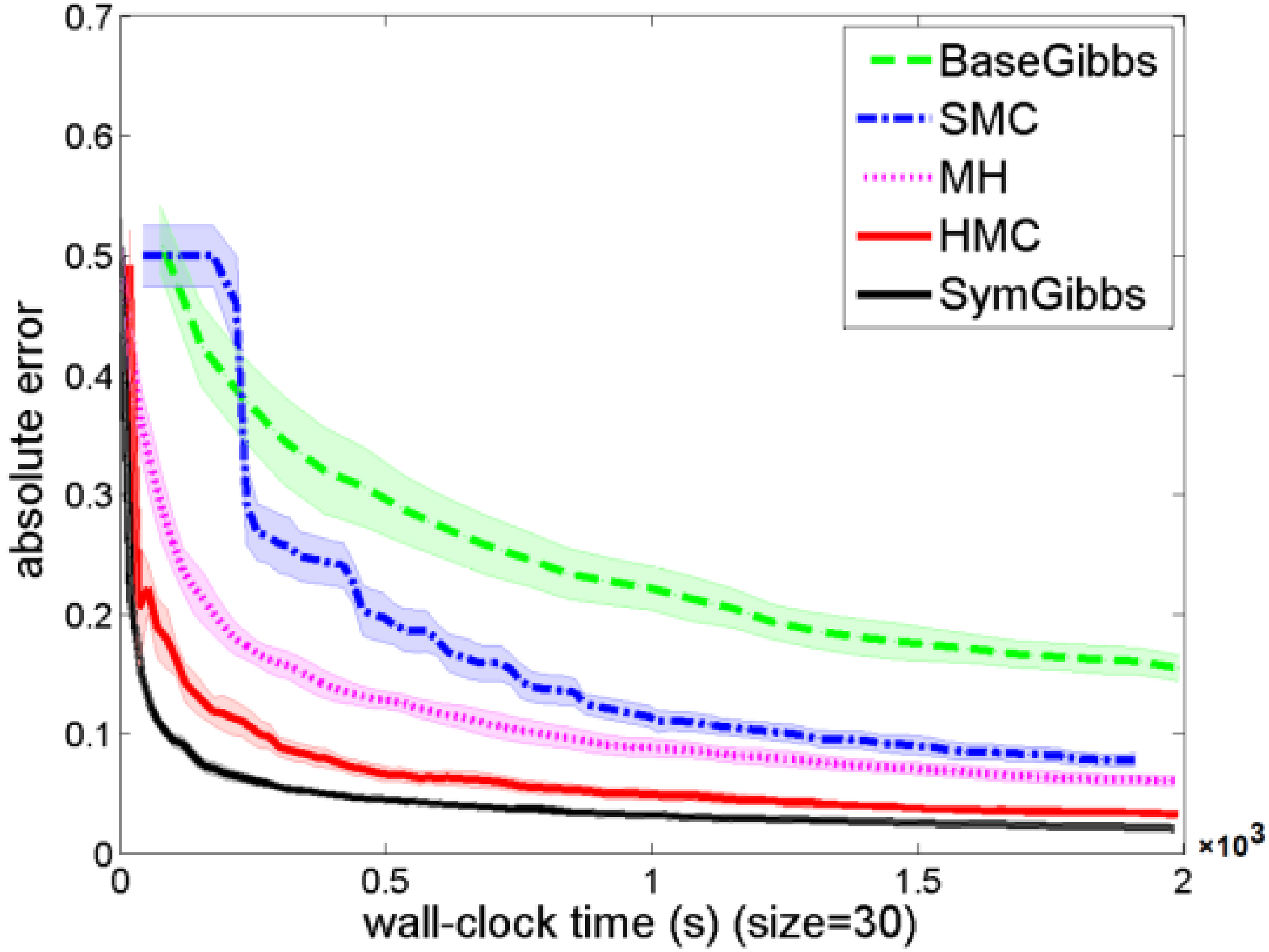
- Collapse one variable in each algebraic constraint

2. To take **S** samples from an **N**-dimensional model,

- In baseline Gibbs, **S×N** (univariate) CDFs are computed.
- In Symbolic Gibbs, **N** (analytical) CDFs and **S×N** function evaluations are required.

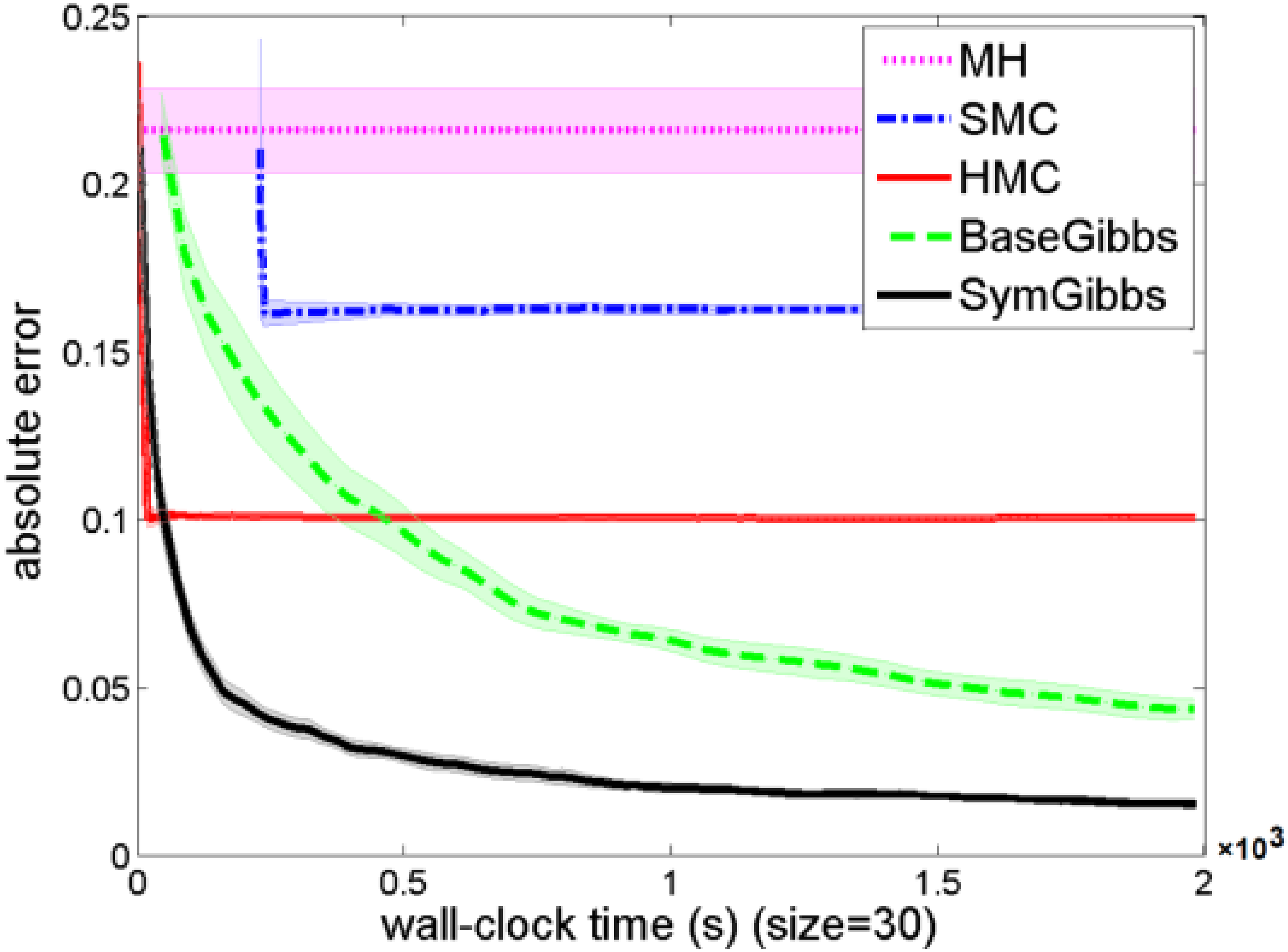
Much faster!

Results



0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
absolute error
BaseGibbs
SMC
MH
HMC
SymGibbs
0
0.5
1
1.5
2
×10^3
wall-clock time (s) (size=30)

Results



0.25
0.2
0.15
0.1
0.05
0
absolute error
0
0.5
1
1.5
2
×10^3
wall-clock time (s) (size=30)
MH
SMC
HMC
BaseGibbs
SymGibbs

Conclusions

- Expressive Graphical Models / Probabilistic Programs
  - Allow algebraic constraints
  - Represent factors as polynomial-piecewise fractions (PPFs)
  - Sufficient for rich class of probabilistic programs
    - Among **many** other applications…

- Result 1: Collapse all algebraic constraints (determism)
  - Yields symbolic substitutions into PPF form

- Result 2: PPFs are one-time integrable!
  - Symbolically pre-compute all conditions for Gibbs sampling
  - Leads to very fast Gibbs sampler!

Expressive Exact GM / PP MCMC Inference!