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 \( \text{if } (t > 0) \text{ ...} \)
  – 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

(a) SymGibbs

(b) MH

(c) HMC (high noise)

(d) HMC (low noise)

(e) SMC (high noise)

(f) SMC (low noise)
Contribution 1:
We present an effective sampler for GMs with piecewise factors

Example:
\[
\begin{align*}
\frac{x}{y} & \quad \text{if} \quad x + y > 0 \\
\frac{x^2 + y}{y^2} & \quad \text{if} \quad x + y < 0, y > 0 \\
\vdots & \quad \vdots
\end{align*}
\]
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

```matlab
% 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: \( pr(V_1, M_2, V_2|P_{tot} = 3) ? \)

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

\[ \delta(f(x)) = \sum_{r \in \text{roots}(f(x))} \frac{\delta(x - r)}{|\frac{\partial f(x)}{\partial x}|} \]

\[ P_{tot} = M_1V_1 + M_2V_2, \text{ so:} \]
\[ pr(P_{tot}) = \delta(M_1V_1 + M_2V_2 - 3) \]
\[ = \frac{\delta \left( M_1 - \left( \frac{3 - M_2V_2}{V_1} \right) \right)}{|V_1|} \]
**Contribution 2: Algebraic Constraints**

- **An example:**

\[ pr(M_1) = U(M_1; 0.1, 2.1) \]
\[ pr(V_1) = U(V_1; -2, 2) \]
\[ pr(M_2) = U(M_2; 0.1, 2.1) \]
\[ pr(V_2|V_1) = U(V_2; -2, V_1) \]

Observation: \( P_{tot} = 3 \)

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

\[ pr(V_1, M_2, V_2|P_{tot} = 3) \propto \int_{0}^{\infty} pr(P_{tot}, V_1, M_2, V_2)pr(V_1, M_2, V_2|P_{tot} = 3) \]

\[ P_{tot} = M_1V_1 + M_2V_2 \] \( \text{so,} \)

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

\[ \delta \left( M_1 - \left( \frac{3 - M_2V_2}{V_1} \right) \right) \]

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

Divisions, absolute values! PPFs are closed under them 😊

Next: PPFs are closed under fractional substitutions! 😊
**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,  
  $M_1 V_1 + M_2 V_2 = 3$

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

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

  \[ \alpha \begin{cases} 
  \frac{1}{V_1(V_1 + 2)} & \text{if } 0 < V_1, 0.1 < \frac{3 - M_2 V_2}{V_1} < 2.1, 1 < M_2 < 3, \\
  -1 & \text{if } 0 > V_1, 0.1 < \frac{3 - M_2 V_2}{V_1} < 2.1, 1 < M_2 < 3, \\
  0 & \text{if } -2 < V_1 < 2, -2 < V_2 < V_1 \\
  & \text{otherwise} \end{cases} \]

  \[ pr(P_{tot} | \ldots) = \frac{\delta \left( M_1 - \left( \frac{3 - M_2 V_2}{V_1} \right) \right)}{|V_1|} \]
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 on 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.

\[
X_1 \sim \text{pr}(X_1|X_2 = b)
\]

\[
X_2 \sim \text{pr}(X_2|X_1 = a')
\]

\( n \) univariate CDFs

= \( n \) integrations per sample 😞
IDEA!
What if we can compute CDFs symbolically and prior to sampling?

Current sample
\( (X_1 = a, X_2 = b) \)

Intermediate sample
\( (X_1 = a', X_2 = b) \)

Next sample
\( (X_1 = a', X_2 = b') \)

Gibbs sampling
- Remember that in Gibbs, sampling from an \( N \)-dimensional function is done in \( N \) steps.

\[ X_1 \sim p(X_1 | X_2 = b) \]

\[ X_2 \sim p(X_2 | X_1 = a') \]

\( n \) univariate CDFs \( \Rightarrow n \) integrations per sample
Gibbs sampling

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

\[
X_1 \rightarrow F_{X_1}(x) := \int_{-\infty}^{x} P(X_1 = t, X_2, \ldots, X_n) dt \\
\vdots \\
X_n \rightarrow F_{X_n}(x) := \int_{-\infty}^{x} P(X_1, \ldots, X_{n-1}, X_n = t) dt
\]

n integrals rather than n \times \text{samples}

Mapping variables to symbolic CDFs

Only need to do one integral which is possible for a large class of PPFs
Returning to our example:

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

\[
\begin{cases}
\frac{1}{V_1(V_1 + 2)} & \text{if } 0 < V_1, V_1 < 30 - 10M_2V_2, \frac{3 - M_2V_2}{2.1} < V_1, 1 < M_2 < 3, \\
\frac{-1}{V_1(V_1 + 2)} & \text{if } 0 > V_1, V_1 > 30 - 10M_2V_2, \frac{3 - M_2V_2}{2.1} > V_1, 1 < M_2 < 3, \\
0 & \text{otherwise}
\end{cases}
\]
Returning to our example:

\[ V_1 \xrightarrow{\text{maps}} F_{V_1}(v_1) \]

\[
= \int_{v_1}^{v_1} \int_{v_1}^{v_1} \left( \frac{1}{V_1(V_1 + 2)} \right) \]

\[
\int_{v_1}^{v_1} \left( \frac{-1}{V_1(V_1 + 2)} \right) \]

\[
\text{if } 0 < V_1, V_1 < 30 - 10M_2V_2, \quad \frac{3 - M_2V_2}{2.1} < V_1, 1 < M_2 < 3, \\
-2 < V_1, V_1 < 2, -2 < V_2, V_2 < V_1 \\
\]

\[
\text{if } 0 > V_1, V_1 > 30 - 10M_2V_2, \quad \frac{3 - M_2V_2}{2.1} > V_1, 1 < M_2 < 3, \\
-2 < V_1 < 2, -2 < V_2 < V_1 \quad \text{otherwise} \]

Let’s just consider one statement.
Returning to our example:

\[
\int_{v_1=-\infty}^{v_1} \left( \frac{1}{V_1(V_1 + 2)} \right) \begin{cases} 
if & 0 < V_1, V_1 < 30 - 10M_2V_2, \frac{3 - M_2V_2}{2.1} < V_1, 1 < M_2 < 3, \\
& -2 < V_1, V_1 < 2, -2 < V_2, V_2 < V_1 
\end{cases} 
\right) dV_1
\]
Returning to our example:

\[ \int_{v_1=-\infty}^{v_1} \left( \frac{1}{V_1(V_1 + 2)} \right) \, dV_1 \]

\[ = \begin{cases} 
\int_{v_1=\max\{0, V_2, 3-M_2V_2 \}}^{\min\{v_1, 30 - 10M_2V_2, 2\}} \frac{1}{V_1(V_1 + 2)} \, dV_1 & \text{if } \min\{v_1, 30 - 10M_2V_2, 2\} > \max\{0, V_2, 3-M_2V_2 \}, \\
0 & \text{if } 1 < M_2 < 3, -2 < V_2, \ \\
\int_{v_1=\min\{v_1, 30 - 10M_2V_2, 2\}}^{\max\{0, V_2, 3-M_2V_2 \}} \frac{1}{V_1(V_1 + 2)} \, dV_1 & \text{otherwise} 
\end{cases} \]

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, $S \times N$ (univariate) CDFs are computed.
   • In Symbolic Gibbs, $N$ (analytical) CDFs and $S \times N$ function evaluations are required.

Much faster!
Results
Results
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 (determinism)
  – 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!