Gaussians under Linear Domain Constraints

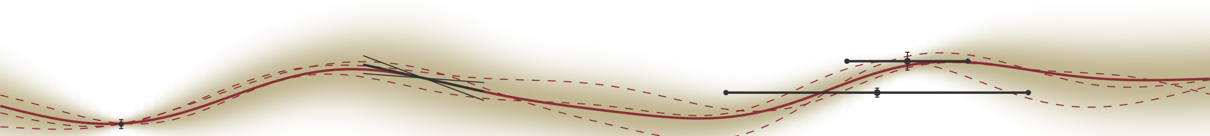
Alexandra Gessner
Trinity College Dublin
March 10, 2021

EBERHARD KARLS UNIVERSITÄT TÜBINGEN

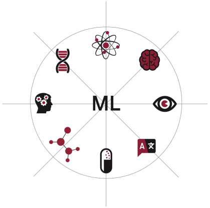
MAX PLANCK INSTITUTE FOR INTELLIGENT SYSTEMS

imprs-is International Max Planck Research School on Intelligent Systems

some of the presented work is supported by the European Research Council.



Machine Learning in Tübingen



ellis

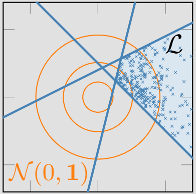
MAX PLANCK INSTITUTE
FOR INTELLIGENT SYSTEMS

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

CyberValley

Sampling and integration of a linearly constrained Gaussian

PROBLEM



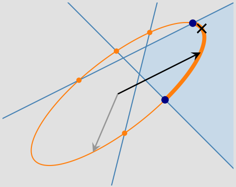
$P(\mathbf{x} \in \mathcal{L}) = \ ?$

$\mathbf{x} \sim \mathcal{N}(0,1)\ \mathbb{1}_{\mathcal{L}}\ ?$

$\rightarrow$

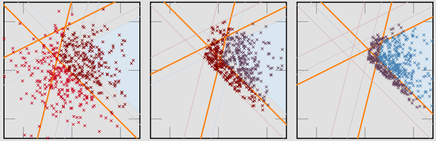
SOLUTION

Elliptical slice sampling



$+$

Multilevel splitting

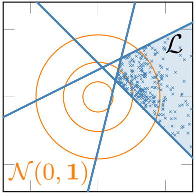


$=$ Rejection-free sampling / Estimation of the normalization constant

Problem setting

Consider M linear functions $\mathbf{A}^\mathsf{T}\mathbf{x} + \mathbf{b}$; $\mathbf{A} \in \mathbb{R}^{D\times M}, \mathbf{b} \in \mathbb{R}^M$ and define the domain

$$\mathcal{L} = \{\mathbf{x} : \mathbf{a}_m^\mathsf{T}\mathbf{x} + b_m > 0; \quad \forall m = 1, \dots, M\} \subset \mathbb{R}^D$$



GOAL

Integrate: $$Z = P(\mathbf{x} \in \mathcal{L}) = \int_{\mathbb{R}^D} \mathbb{1}_\mathcal{L} \; \mathrm{d}\,\mathcal{N}(\mathbf{x}; 0, \mathbf{1}) \qquad \mathbb{1}_\mathcal{L} = \begin{cases} 1 \text{ if } \mathbf{x} \in \mathcal{L} \\ 0 \text{ if } \mathbf{x} \notin \mathcal{L} \end{cases}$$

Sample: $$\mathbf{x} \sim \frac{1}{Z} \mathcal{N}(\mathbf{x}; 0, \mathbf{1})\, \mathbb{1}_\mathcal{L}$$

Challenges:

- Z might be tiny
- High dimensions D

Assumptions:

- Standard normal w.l.o.g.
- Typically $M \geq D$

Part I: Sampling

Basics: Slice sampling

[Neal, 1997; Neal, 2003]

MCMC algorithm with similarities to **Gibbs sampling** and **rejection sampling**.

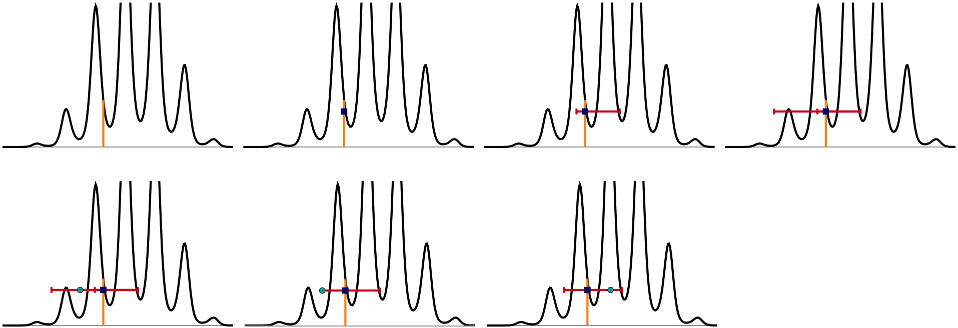


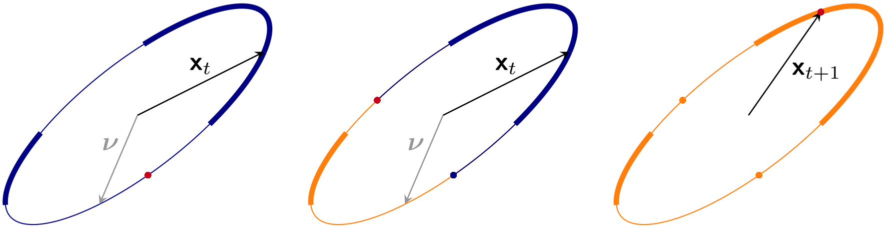
Image adapted from MacKay: ITILA, 2006

MCMC algorithm for the special case that $p_*(\mathbf{x}) = \ell(\mathbf{x})\,\mathcal{N}(\mathbf{x}, \mathbf{0}, \mathbf{\Sigma})$

Construct **1D** ellipse from state $\mathbf{x}_t$ and auxiliary vector $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{x}, \mathbf{0}, \mathbf{\Sigma})$ as

$$\mathbf{x}(\theta) = \mathbf{x}_t \cos\theta + \boldsymbol{\nu} \sin\theta$$

and perform **slice sampling** on ellipse.



Note: this algorithm is parameter-free!

Sampling from a linearly constrained Gaussian domain

An adaptation of elliptical slice sampling

Elliptical slice sampling where...

1. "Likelihood" $\ell(\mathbf{x}) = \mathbb{1}_{\mathcal{L}}$ has binary outcome, 0 or 1

   $\rightarrow$ **no likelihood threshold**

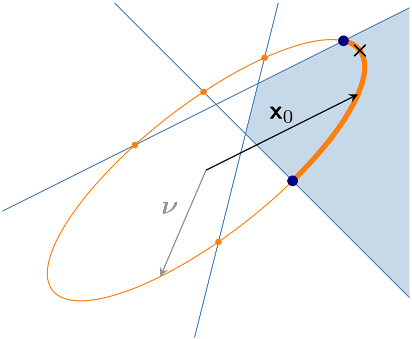
2. Intersections of ellipse and domain boundaries have closed-form solutions

$$\mathbf{A}^{\intercal}(\mathbf{x}_0 \cos\theta + \boldsymbol{\nu} \sin\theta) + \mathbf{b} = \mathbf{0}$$

$$\theta_{m,1/2} = \pm \arccos\left(-\frac{b_m}{r}\right) + \arctan\left(\frac{\mathbf{a}_m^{\intercal}\boldsymbol{\nu}}{r + \mathbf{a}_m^{\intercal}\mathbf{x}_0}\right)$$

with $r = \sqrt{(\mathbf{a}_m^{\intercal}\mathbf{x}_0)^2 + (\mathbf{a}_m^{\intercal}\boldsymbol{\nu})^2}$

$\rightarrow$ **rejection-free sampling**



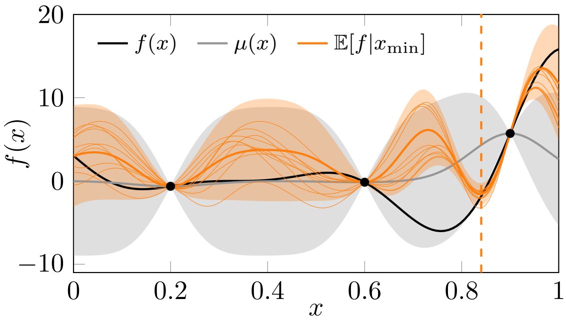
LIN-ESS

Sampling: Applications

Drawing samples from a linearly constrained multivariate normal distribution

Bayesian optimization:

Predictive entropy search requires $p(f \mid X, x_{\min})$



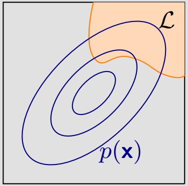
Gaussian processes

A Gaussian process $f \sim \mathcal{GP}(\mu, k)$is a random process with mean function $\mu : \mathbb{R} \to \mathbb{R}$ and covariance function $k : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that f evaluated at a finite set of inputs $\mathbf{X}$ follows a multivariate normal distribution.

Part II: Integration

Motivation: Reliability analysis

A problem in reliability analysis



$$P(\mathbf{x} \in \mathcal{L}) = \ ?$$

probability of failure

Problem

- Failure probability is extremely small
  $\rightarrow$ ~~rejection sampling~~
- Experiments (= evaluations) can be expensive

Idea

Decompose integral into easier-to-solve integrals

Multilevel splitting methods: Holmes-Diaconis-Ross, subset simulation

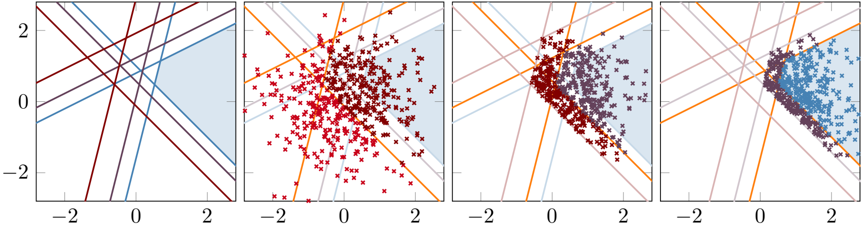
Integration via the Holmes-Diaconis-Ross algorithm

[Diaconis and Holmes 1995, Ross 2012]

Key idea: Decompose integral into product of conditional probabilities:

$$Z = P(\mathcal{L}) = P(\mathcal{L}_0) \prod_{t=1}^{T} P(\mathcal{L}_t | \mathcal{L}_{t-1}).$$

of **nested domains** $\mathbb{R}^D = \mathcal{L}_0 \supset \mathcal{L}_1 \supset \mathcal{L}_2 \supset \ldots \supset \mathcal{L}_T = \mathcal{L}$, s.t. $\mathcal{L}_t = \bigcap_{i=1}^{t} \mathcal{L}_i$.



Define nested domain by a set of **scalar shift values** $\infty = \gamma_0, \ldots, \gamma_T = 0$ with

$$\mathcal{L}_t = \{\mathbf{x} : \mathbf{a}_m^\intercal \mathbf{x} + b_m + \gamma_t > 0 \quad \forall m = 1, \ldots, M\}$$

The Holmes-Diaconis-Ross algorithm

[Diaconis and Holmes 1995, Ross 2012]

procedure HDR($\mathbf{A}, \mathbf{b}, \{\gamma_1, \dots, \gamma_T\}, N$)

  $\mathbf{X} \sim \mathcal{N}(0, \mathbf{1})$ // N samples

  $\log Z = 0$ // initialize log integral value

  for $t = 1 \dots T$ **do**

    $\mathcal{L}_t = \{\mathbf{x} : \min_m(\mathbf{a}_m^\top \mathbf{x}_n + b_m) + \gamma_t > 0\}_{n=1}^N$ // find samples inside current nesting

    $\log Z \leftarrow \log Z + \log(\#(\mathbf{X} \in \mathcal{L}_t)) - \log N$

    choose $\mathbf{x}_0 \in \mathcal{L}_t$

    $\mathbf{X} \leftarrow$ LinESS($\mathbf{A}, \mathbf{b} + \gamma_t, N, \mathbf{x}_0$) // draw new samples from constrained domain

  end for

  return $\log Z$

end procedure

HDR allows to compute the **logarithm** of the integral, $\log \hat{Z} = \sum_{t=1}^{T} \log \rho_t$, where $\rho_t = P(\mathcal{L}_t | \mathcal{L}_{t-1})$.

Integration: Subset simulation

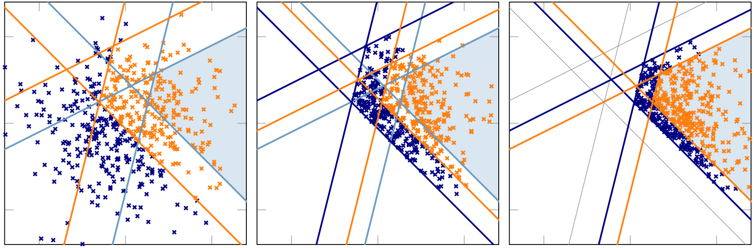
[Au and Beck 2001]

i) LIN-ESS to sample from $\mathcal{L}_t,\ t = 1 \dots T$? ✓

ii) How to choose the nested domains $\gamma_0, \dots, \gamma_T$?

Subset simulation: similar to HDR, but:

1. Fix conditional probabilities $\rho_t = P(\mathcal{L}_t | \mathcal{L}_{t-1})$ to $\rho = 1/2$
2. Sample from current nesting and choose γ_t s.t. $\lfloor N\rho \rfloor$ samples fall in $\mathcal{L}_t$.

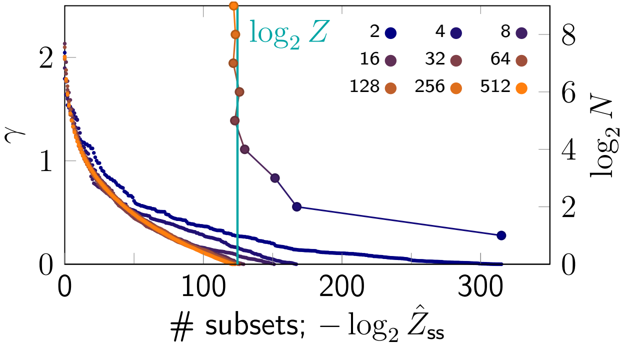


Integration: Pre-processing with subset simulation

...and some insight about the role of the nestings

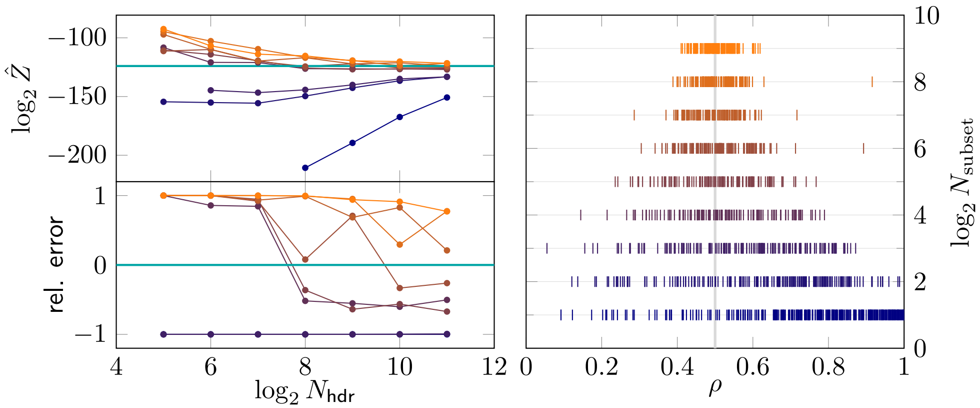
Subset simulation:

- How many samples?
- Example: 500-d shifted orthant
  $Z = 3.07 \cdot 10^{-38} = 2^{-124.6}$
- Subset simulation is biased



If $\rho = P(\mathcal{L}_t \,|\, \mathcal{L}_{t-1})\ \forall t = 0 \ldots T$ and T nestings, the subset estimator is roughly $\hat{Z}_{\text{ss}} \approx \rho^T$

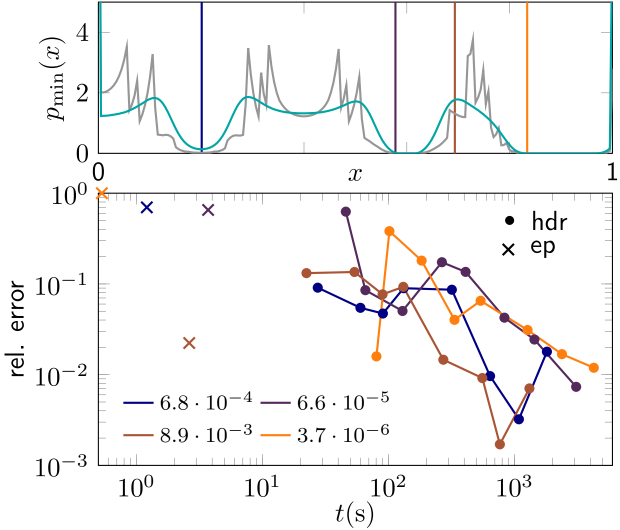
Integration: Taking over with HDR



HDR performance depends on nesting sequences: **a good nesting matters!**

Experiments: Entropy Search

Probability of x_i to be the minimum [Hennig & Schuler 2012]



Probability of minimum:

$$p_{\min}(\mathbf{x}_i) = \int d\mathbf{f}\ \mathcal{N}(\mathbf{f}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_{j \neq i} \mathbb{1}_{f(\mathbf{x}_i) < f(\mathbf{x}_j)}$$

Probabilistic tree search

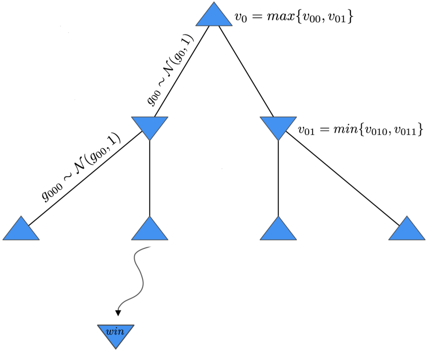
work & slides by Julia Grosse

Scores under random play:

$$p(g_0) = \mathcal{N}(g_0; 0, 1)$$
$$p(g_i|g_j) = \mathcal{N}(g_i; \mu_j, 1)$$

Scores under optimal play:

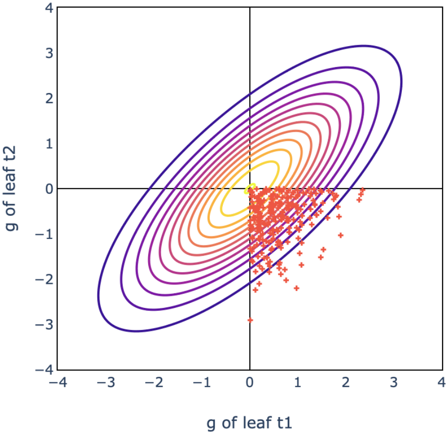
$$v_i = \begin{cases} g_i & \textit{if } i \text{ is leaf} \\ \max_{j \in \text{children}(i)}\{v_j\} & \text{if } i \text{ is MAX node} \\ \min_{j \in \text{children}(i)}\{v_j\} & \text{if } i \text{ is MIN node} \end{cases}$$



Observations are linear constraints at leaf nodes:

$$p(\text{win}|g_t) = \mathbb{I}[g_t > 0]$$
$$p(\text{loss}|g_t) = \mathbb{I}[g_t < 0]$$

Assume we observe a win at leaf t_1 and a loss at leaf t_2.
The posterior over the scores is a constraint Gaussian:



Summary

Gaussians under linear domain constraints with Oindrila Kanjilal and Philipp Hennig

(i) rejection-free sampling scheme based on elliptical slice sampling

(ii) integration scheme that works for arbitrarily small probability masses



Paper `arxiv:1910.09328`

Code `https://github.com/alpiges/LinConGauss`

Thank you!