

# Homogenous sampling from a finite volume

Chris Hartl

July 31, 2019

## Solution description

### *Initial sampling*

The problem is to draw a set of  $N$  spatially homogenous points from an arbitrary volume enclosed by the unit hypercube and a set of arbitrary constraints. There are two concerns: efficiency and homogeneity. The objective is first to maximize homogeneity, then to minimize runtime. I'm not aware of metrics for homogeneity, but a reasonable proxy should be the median of the distance matrix

$$H(X) = \text{median}(D(X))$$

$$D = \xi \mathbf{1}^T + \mathbf{1} \xi^T - 2\Xi$$

$$\Xi = XX^T \quad \xi = \text{diag}(\Xi)$$

I thought of two possible approaches: an explicit method (given a valid point set, generate a new valid point) or an implicit method (given any point set, generate a new point which may or may not be valid). A simple interior method might be:

given: point  $x$  in the valid volume, dimensionality  $n$

- 1) Pick a direction  $v$  at random from  $U(\mathcal{S}^{n-1})$
- 2) Calculate the endpoints  $(a, b)$  of the line running through  $x$  with direction  $v$  and  $-v$ , and their distance  $d_{ab}$
- 3) Sample  $x^{(new)} = a + d_{ab}U([0, 1])(b - a)$

While a simple implicit method might be (importance sampling):

given: Set of points  $X$  in the valid volume, dimensionality  $n$ ; overdispersion  $k$ , mixture  $p$

- 1) Compute mean and covariance  $\mu, \Sigma$  of  $X$
- 2) Sample  $x^{(new)}$  from a mixture of  $N(\mu, k\Sigma)$  and  $U([0, 1]^n)$

In either case, the primary challenge appeared to be finding a good initial point. I reasoned that this could be done with basic optimization, by converting the constraints into one-sided costs:

$$g(x) > 0 \Rightarrow c(x) = \begin{cases} 0 & g(x) \geq 0 \\ -\lambda g(x) & g(x) < 0 \end{cases}$$

$-c(x)$ , when used as a potential, defines a probability density which is flat over the valid volume, with exponential tails (in units of  $g$ ) outside of it. This makes the problem amenable to advanced MCMC methods. That is, sampling from

$$\log P(x) \propto -\sum_{i=1}^k c_i(x)$$

produces a high proportion of valid points, and an approximately uniform density.

At this point, methods built on top of this baseline seemed like a path of least resistance; while implementing an explicit algorithm would require implementing an algorithm to find the first binding constraint in a given direction (root finding and for loops).

#### *Improving homogeneity*

Points drawn from a uniform distribution are by their nature heterogeneous (maximally so, as the uniform distribution has maximal entropy on  $[0, 1]^n$ ). One way to boost the heterogeneity is to identify good points that are landmarks for a neighborhood. One very simple landmark construction algorithm which I've used is to start with some representative landmark set, and take as the next landmark the point which is *furthest* from the landmark set. This is a greedy approach to choosing from among a set of points, the  $K$  points that maximize the total distance:

given: An existing landmark set  $X_l$

- 1) Compute the distances  $D(X, X_l)$
- 2) Find  $\text{argmax}_i [\min_j D(X, X_l)_{i,j}]$
- 3) Add  $X_i$  to  $X_l$

In this way, a dense *oversampling* of heterogeneous points within the volume can be converted to a sparser, more homogenous set of landmark points.

I wondered about an alternative optimization approach where, given a set of points  $X$ , the total energy (pairwise distances) was directly minimized, subject to the constraints. If every point in the set is a valid point, then using a boundary function  $\propto 1/g(x)$  presents a confinement barrier. Necessarily,, a small proportion of points will wind up as invalid after the optimization, so a small amount of oversampling is warranted.

On the example dataset, I found uniform sampling had about a 20% valid point rate, with  $H(X) \sim 0.55$ . Sampling from a potential had a 70% valid point rate, with  $H(X) \sim 0.55$ . Landmark post-processing improved  $H(X)$  to 0.59. Finally, the Landmark  $\rightarrow$  Optimization  $\rightarrow$  Landmark steps drastically improved

the median distance to  $H(X) \sim 0.706$ .

### *The Alloy Example and effective scale*

The alloy example represents a sampling space which is a tiny volume; necessitating an aggressive scale factor  $\lambda$  to generate accepting points. This large scaling factor also makes gradients unwieldy, slowing down the sampling time.

One possibility to deal with this is to re-parametrize the problem, such that  $x = f(z)$  with  $\nabla f \ll 1$ ; to do this one would potentially need different expansions of  $x$  around the feasible region of each constraint. This might be quite an interesting research project, but the right way to do this isn't immediate to me.

More generally, what has happened is that the test function  $q(X)$  being used in the MCMC ( $q \sim U([0, 1]^n)$ ) is not well adapted to the volume of the valid region. By adding parameters to  $q$ , one can identify

$$\theta^* = \operatorname{argmin}_{\theta} \operatorname{DL}[q(X; \theta), p(X)]$$

where  $p(X)$  is described by the constraint potentials. As the volume may be axis aligned, one potentially good approximation would be to use

$$q \sim N(\mu, \Sigma)$$

Fitting this to the alloy dataset achieves a reasonable acceptance rate (38%), but abysmal inter-point distances due to the Gaussianity of the approximation (0.026). Oversampling+landmarks improves this to 0.0336, and the added optimization improves  $H(X)$  to 0.036.

This won't work effectively for extreme cases of spatial nonlinearity (half moons or annuli); but point-cloud approximations (Stein variational gradient descent) could work.

## Possible future improvements

The Variational Inference can be regarded as a first step in an approximation, or as a first step in finding a better parametrization. In particular, the transformation

$$X = \mu + Z\Sigma^{1/2}$$

can be used to identify appropriate marginal bounds on  $Z$  to provide a test distribution on reasonable space. On the other hand, for completely arbitrary regions, this approach itself will fail. An alternative method which I took 5 minutes to explore was Stein variational gradient descent. Ideally, given a point cloud of size  $n$ , it iterates

$$x_i^t = x_i^{t-1} + \alpha \langle K(\xi, x_i^t) \nabla_{\xi} \log p(\xi) + \nabla_{\xi} K(\xi, x_i^t) \rangle_{\xi \in x^t}$$

which minimizes the Stein discrepancy between  $p$  and the space of functions defined by the kernel  $K$  (an approximate unnormalized likelihood) on the points  $x$ . In effect, the bandwidth of  $K$  provides a tradeoff between mode-seeking and density seeking; and this point-cloud method should perform well. I think difficulty is matching the bandwidth to the scale of the constraint set — which may even be multi-scale (e.g., funnels). SVGD is still an active area of research in Bayesian modeling; it is likely that some concepts from GPs (a deep GP version, or automatic length-scale selection) will move over in the near future.

The explicit approach for sampling the interior is likely more effective for small-volume problems; but I would have had to do much more reading and thinking to implement it. For special cases, like polytopes, it could be that there are reasonable parameterizations or dual parameterizations that make the sampling tractable and scale invariant.

Lastly, there are likely better approaches for converting point clouds to representative points (I've seen KD trees referenced quite a bit, for instance in Barnes-Hut t-SNE so I might want to look into those); as well as for the constrained optimization. Boundary functions are my go-to hack, but they're clunky and not strict bounds.