Support Points

Simon Mak, V. Roshan Joseph

Georgia Institute of Technology

Wittawat Jitkrittum

Gatsby tea talk
21 Oct 2016

Overview

Support points (https://arxiv.org/abs/1609.01811, 7 Sep 2016)
Simon Mak, V. Roshan Joseph

- Generate representative points from a continuous distribution F.

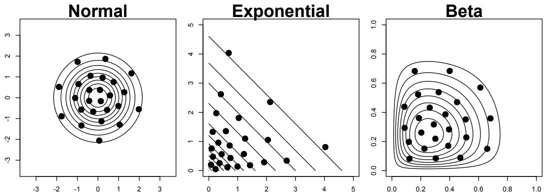

Figure 1: *$n = 25$ support points for the two-dimensional $N(0,1)$, $Exp(1)$ and $Beta(2,4)$ distributions. Lines represent density contours.*

- By minimizing the distance-based energy statistic E [Székely and Rizzo(2013)].
- Applications:
  1. Alternative to MCMC thinning.
  2. Optimally allocate runs for stochastic simulations. Want the distribution of $g(\mathbf{x})$ where $\mathbf{x} \sim F$. Expensive to simulate g. So, find support points for F first.
  3. Numerical integration. Alternative to Monte-Carlo particles. Consistent.

Overview

Support points (https://arxiv.org/abs/1609.01811, 7 Sep 2016)
Simon Mak, V. Roshan Joseph

- Generate representative points from a continuous distribution F.

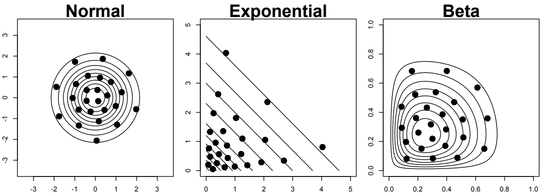

Figure 1: *$n = 25$ support points for the two-dimensional $N(0,1)$, $Exp(1)$ and $Beta(2,4)$ distributions. Lines represent density contours.*

- By minimizing the distance-based energy statistic E [Székely and Rizzo(2013)].
- Applications:
  1. Alternative to MCMC thinning.
  2. Optimally allocate runs for stochastic simulations. Want the distribution of $g(\mathbf{x})$ where $\mathbf{x} \sim F$. Expensive to simulate g. So, find support points for F first.
  3. Numerical integration. Alternative to Monte-Carlo particles. Consistent.

How to obtain the support points for F?

- Let F, G be two cont. distributions on $\mathcal{X} \subseteq \mathbb{R}^p$. Let $\mathbf{x}, \mathbf{x}' \overset{i.i.d.}{\sim} G$ and $\mathbf{y}, \mathbf{y}' \overset{i.i.d.}{\sim} F$.

$$E(F, G) := 2\mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{y}}\|\mathbf{x} - \mathbf{y}\|_2 - \mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{x}'}\|\mathbf{x} - \mathbf{x}'\|_2 - \mathbb{E}_{\mathbf{y}}\mathbb{E}_{\mathbf{y}'}\|\mathbf{y} - \mathbf{y}'\|_2.$$

- Assume $G = F_n$ (empirical distribution function, e.d.f.) for $\{\mathbf{x}_i\}_{i=1}^n$.
- In practice, $\{\mathbf{y}_j\}_{j=1}^N \sim F$. For a fixed point set size n, the support points are

$$\begin{aligned}\{\xi_i\}_{i=1}^n &= \arg\min_{\mathbf{x}_1,\ldots,\mathbf{x}_n} E\left(\{\mathbf{y}_j\}_{j=1}^N, \{\mathbf{x}_i\}_{i=1}^n\right) \\ &= \arg\min_{\mathbf{x}_1,\ldots,\mathbf{x}_n} \frac{2}{nN}\sum_{i=1}^n\sum_{j=1}^N \|\mathbf{x}_i - \mathbf{y}_j\|_2 - \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|\mathbf{x}_i - \mathbf{x}_j\|_2.\end{aligned}$$

  - Using a block coordinate descent,

$$\mathbf{x}_i = \arg\min_{\mathbf{x}} \frac{2}{N}\sum_{j=1}^N \|\mathbf{x} - \mathbf{y}_j\|_2 - \frac{1}{n}\sum_{k\neq i} \|\mathbf{x} - \mathbf{x}_k\|^2$$

How to obtain the support points for F?

- Let F, G be two cont. distributions on $\mathcal{X} \subseteq \mathbb{R}^p$. Let $\mathbf{x}, \mathbf{x}' \overset{i.i.d.}{\sim} G$ and $\mathbf{y}, \mathbf{y}' \overset{i.i.d.}{\sim} F$.

$$E(F, G) := 2\mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{y}}\|\mathbf{x} - \mathbf{y}\|_2 - \mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{x}'}\|\mathbf{x} - \mathbf{x}'\|_2 - \mathbb{E}_{\mathbf{y}}\mathbb{E}_{\mathbf{y}'}\|\mathbf{y} - \mathbf{y}'\|_2.$$

- Assume $G = F_n$ (empirical distribution function, e.d.f.) for $\{\mathbf{x}_i\}_{i=1}^n$.
- In practice, $\{\mathbf{y}_j\}_{j=1}^N \sim F$. For a fixed point set size n, the support points are

$$\begin{aligned}\{\boldsymbol{\xi}_i\}_{i=1}^n &= \arg\min_{\mathbf{x}_1, \ldots, \mathbf{x}_n} E\left(\{\mathbf{y}_j\}_{j=1}^N, \{\mathbf{x}_i\}_{i=1}^n\right) \\ &= \arg\min_{\mathbf{x}_1, \ldots, \mathbf{x}_n} \frac{2}{nN}\sum_{i=1}^n\sum_{j=1}^N \|\mathbf{x}_i - \mathbf{y}_j\|_2 - \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|\mathbf{x}_i - \mathbf{x}_j\|_2.\end{aligned}$$

- Using a block coordinate descent,

$$\mathbf{x}_i = \arg\min_{\mathbf{x}} \frac{2}{N}\sum_{j=1}^N \|\mathbf{x} - \mathbf{y}_j\|_2 - \frac{1}{n}\sum_{k \neq i} \|\mathbf{x} - \mathbf{x}_k\|^2$$

How to obtain the support points for F?

- Let F, G be two cont. distributions on $\mathcal{X} \subseteq \mathbb{R}^p$. Let $\mathbf{x}, \mathbf{x}' \overset{i.i.d.}{\sim} G$ and $\mathbf{y}, \mathbf{y}' \overset{i.i.d.}{\sim} F$.

$$E(F, G) := 2\mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{y}}\|\mathbf{x} - \mathbf{y}\|_2 - \mathbb{E}_{\mathbf{x}}\mathbb{E}_{\mathbf{x}'}\|\mathbf{x} - \mathbf{x}'\|_2 - \mathbb{E}_{\mathbf{y}}\mathbb{E}_{\mathbf{y}'}\|\mathbf{y} - \mathbf{y}'\|_2.$$

- Assume $G = F_n$ (empirical distribution function, e.d.f.) for $\{\mathbf{x}_i\}_{i=1}^n$.
- In practice, $\{\mathbf{y}_j\}_{j=1}^N \sim F$. For a fixed point set size n, the support points are

$$\begin{aligned}\{\boldsymbol{\xi}_i\}_{i=1}^n &= \arg\min_{\mathbf{x}_1,\ldots,\mathbf{x}_n} E\left(\{\mathbf{y}_j\}_{j=1}^N, \{\mathbf{x}_i\}_{i=1}^n\right)\\ &= \arg\min_{\mathbf{x}_1,\ldots,\mathbf{x}_n} \frac{2}{nN}\sum_{i=1}^n\sum_{j=1}^N \|\mathbf{x}_i - \mathbf{y}_j\|_2 - \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n \|\mathbf{x}_i - \mathbf{x}_j\|_2.\end{aligned}$$

  - Using a block coordinate descent,

$$\mathbf{x}_i = \arg\min_{\mathbf{x}} \frac{2}{N}\sum_{j=1}^N \|\mathbf{x} - \mathbf{y}_j\|_2 - \frac{1}{n}\sum_{k\neq i}\|\mathbf{x} - \mathbf{x}_k\|^2$$

Theoretical properties

- Proposition 1: $E(F,G)$ is a metric i.e., $E(F,G)=0$ if and only if $F=G$ [Székely and Rizzo(2013)].
- $E(F,F_n)$ has desirable properties for an optimization by majorization minimization.
- Theorem 2: Let $\mathbf{x} \sim F$ and $\mathbf{x}_n \sim F_n$ (e.d.f. of the support points). Then, $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$.
  - As $n \to \infty$, the histogram of $\mathbf{x}_n$ matches F.
- Corollary 1: For any continuous map $g : \mathcal{X} \to \mathbb{R}$
  1. $g(\mathbf{x}_n) \xrightarrow{d} g(\mathbf{x})$
  2. If $\mathcal{X}$ is compact, then $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{\xi}_i) = \mathbb{E}[g(\mathbf{x})]$. That is, support points are consistent for integration use asymptotically.

Theoretical properties

- Proposition 1: $E(F, G)$ is a metric i.e., $E(F, G) = 0$ if and only if $F = G$ [Székely and Rizzo(2013)].
- $E(F, F_n)$ has desirable properties for an optimization by majorization minimization.
- Theorem 2: Let $\mathbf{x} \sim F$ and $\mathbf{x}_n \sim F_n$ (e.d.f. of the support points). Then, $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$.
  - As $n \to \infty$, the histogram of $\mathbf{x}_n$ matches F.
- Corollary 1: For any continuous map $g : \mathcal{X} \to \mathbb{R}$
  1. $g(\mathbf{x}_n) \xrightarrow{d} g(\mathbf{x})$
  2. If $\mathcal{X}$ is compact, then $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{\xi}_i) = \mathbb{E}[g(\mathbf{x})]$. That is, support points are consistent for integration use asymptotically.

Theoretical properties

- Proposition 1: $E(F, G)$ is a metric i.e., $E(F, G) = 0$ if and only if $F = G$ [Székely and Rizzo(2013)].
- $E(F, F_n)$ has desirable properties for an optimization by majorization minimization.
- Theorem 2: Let $\mathbf{x} \sim F$ and $\mathbf{x}_n \sim F_n$ (e.d.f. of the support points). Then, $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$.
    - As $n \to \infty$, the histogram of $\mathbf{x}_n$ matches F.
- Corollary 1: For any continuous map $g : \mathcal{X} \to \mathbb{R}$
    1. $g(\mathbf{x}_n) \xrightarrow{d} g(\mathbf{x})$
    2. If $\mathcal{X}$ is compact, then $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{\xi}_i) = \mathbb{E}[g(\mathbf{x})]$. That is, support points are consistent for integration use asymptotically.

Theoretical properties

- Proposition 1: $E(F, G)$ is a metric i.e., $E(F, G) = 0$ if and only if $F = G$ [Székely and Rizzo(2013)].
- $E(F, F_n)$ has desirable properties for an optimization by majorization minimization.
- Theorem 2: Let $\mathbf{x} \sim F$ and $\mathbf{x}_n \sim F_n$ (e.d.f. of the support points). Then, $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$.
  - As $n \to \infty$, the histogram of $\mathbf{x}_n$ matches F.
- Corollary 1: For any continuous map $g : \mathcal{X} \to \mathbb{R}$
  1. $g(\mathbf{x}_n) \xrightarrow{d} g(\mathbf{x})$
  2. If $\mathcal{X}$ is compact, then $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g(\boldsymbol{\xi}_i) = \mathbb{E}[g(\mathbf{x})]$. That is, support points are consistent for integration use asymptotically.

More on theoretical properties

- Theorem 3: For a polynomial integrand g, $\left|\mathbb{E}_{\mathbf{x}\sim F} g(\mathbf{x}) - \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{\xi}_i)\right| \leq \sqrt{V(g)E(F,F_n)}$ where $V(g)$ is a constant depending on g.
  - $\Longrightarrow$ If $E(F,F_n)$ is small, the integration error is also small.
- Theorem 4: For any $0 < \nu < 1$, there exists $C_{\nu,p} > 0$ such that

$$E(F,F_n) \leq \frac{C_{\nu,p}}{n(\log n)^{(1-\nu)/p}}.$$

- This provides an integration error convergence rate

$$\left|\mathbb{E}_{\mathbf{x}\sim F}\left[g(\mathbf{x})\right] - \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{\xi}_i)\right| \leq \sqrt{\frac{V(g)C_{\nu,p}}{n(\log n)^{(1-\nu)/p}}}$$

More on theoretical properties

- Theorem 3: For a polynomial integrand g, $\left|\mathbb{E}_{\mathbf{x}\sim F} g(\mathbf{x}) - \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{\xi}_i)\right| \leq \sqrt{V(g)E(F,F_n)}$ where $V(g)$ is a constant depending on g.
  - $\Longrightarrow$ If $E(F,F_n)$ is small, the integration error is also small.
- Theorem 4: For any $0<\nu<1$, there exists $C_{\nu,p}>0$ such that

$$E(F,F_n) \leq \frac{C_{\nu,p}}{n(\log n)^{(1-\nu)/p}}.$$

- This provides an integration error convergence rate

$$\left|\mathbb{E}_{\mathbf{x}\sim F}\left[g(\mathbf{x})\right] - \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{\xi}_i)\right| \leq \sqrt{\frac{V(g)C_{\nu,p}}{n(\log n)^{(1-\nu)/p}}}$$

Comparison with MC, and QMC (in terms of integration errors)

- Support points: $\left|\mathbb{E}_{\mathbf{x}\sim F}\left[g(\mathbf{x})\right] - \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{\xi}_i)\right| \le \sqrt{\frac{V(g)C_{\nu,p}}{n(\log n)^{(1-\nu)/p}}} = O\left(n^{-1/2}(\log n)^{(1-\nu)/2p}\right)$
- Monte-Carlo (MC):

$$O\left(n^{-1/2}\sqrt{\log\log n}\right)$$

  (independent of p).
- For a fixed p, the error of support points drops faster than MC.
- When p is allowed to vary, MC may have an advantage i.e., constant $C_{\nu,p}$ can rapidly grow.
- Assume $F = U[0,1]^p$. Quasi MC (QMC):

$$O(n^{-1}(\log n)^p).$$

  Faster than support points.
- Claim: In simulations, support points perform better than QMC. Perhaps, the upper bound is not tight.

Comparison with MC, and QMC (in terms of integration errors)

- Support points: $\left|\mathbb{E}_{\mathbf{x}\sim F}\left[g(\mathbf{x})\right] - \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{\xi}_i)\right| \leq \sqrt{\frac{V(g)C_{\nu,p}}{n(\log n)^{(1-\nu)/p}}} = O\left(n^{-1/2}(\log n)^{(1-\nu)/2p}\right)$
- Monte-Carlo (MC):
$$O\left(n^{-1/2}\sqrt{\log\log n}\right)$$
(independent of p).
- For a fixed p, the error of support points drops faster than MC.
- When p is allowed to vary, MC may have an advantage i.e., constant $C_{\nu,p}$ can rapidly grow.
- Assume $F = U[0,1]^p$. Quasi MC (QMC):
$$O(n^{-1}(\log n)^p).$$
Faster than support points.
- Claim: In simulations, support points perform better than QMC. Perhaps, the upper bound is not tight.

Comparison with MC, and QMC (in terms of integration errors)

- Support points: $\left|\mathbb{E}_{\mathbf{x}\sim F}\left[g(\mathbf{x})\right] - \frac{1}{n}\sum_{i=1}^{n} g(\boldsymbol{\xi}_i)\right| \leq \sqrt{\frac{V(g)C_{\nu,p}}{n(\log n)^{(1-\nu)/p}}} = O\left(n^{-1/2}(\log n)^{(1-\nu)/2p}\right)$
- Monte-Carlo (MC):

$$O\left(n^{-1/2}\sqrt{\log\log n}\right)$$

  (independent of p).
- For a fixed p, the error of support points drops faster than MC.
- When p is allowed to vary, MC may have an advantage i.e., constant $C_{\nu,p}$ can rapidly grow.
- Assume $F = U[0,1]^p$. Quasi MC (QMC):

$$O(n^{-1}(\log n)^p).$$

  Faster than support points.
- Claim: In simulations, support points perform better than QMC. Perhaps, the upper bound is not tight.

Illustration

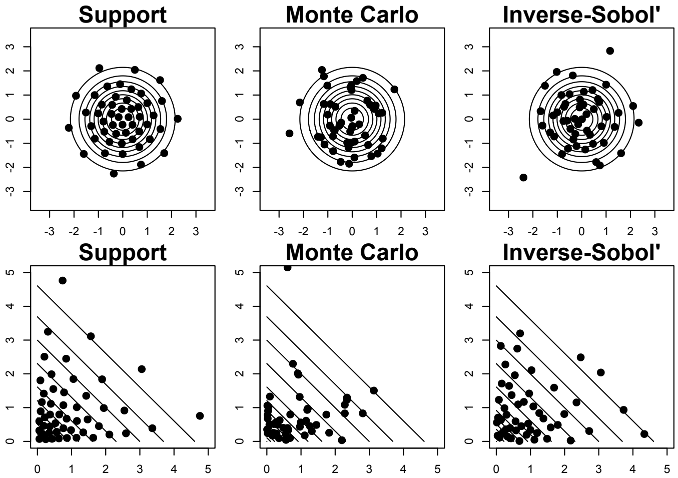

Figure 3: *$n = 50$ support points, MC samples and IT-RSS points for i.i.d. $N(0,1)$ and $Exp(1)$ in $p = 2$ dimensions. Lines represent density contours.*

Key:

- Concentrated around regions with higher density.
- Space filling.

Computation time

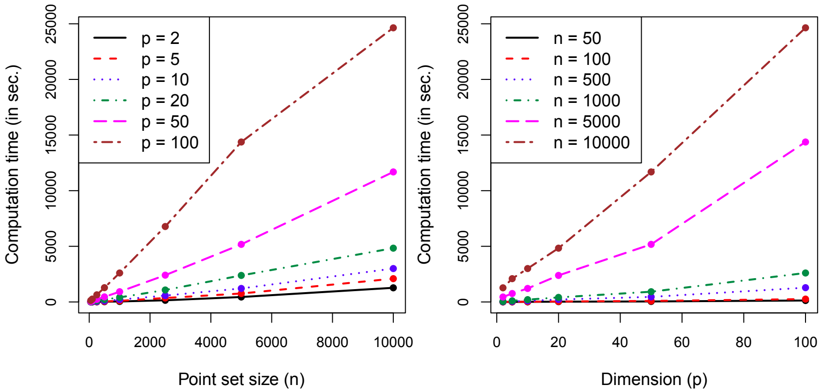

Figure 4: *Computation time (in sec.) of support points vs. point set size (n) and dimension (p) for the i.i.d.* $N(0,1)$ *distribution.*

- Linear running time in n and p.
- Still quite slow.

Simulation: numerical integration

1. Gaussian peak function (GAPK): $g(\mathbf{x}) = \exp\left(-\sum_{l=1}^{p} \alpha_l^2 (x_l - u_l)^2\right)$
2. Oscillatory function (OSC): $g(\mathbf{x}) = \cos\left(2\pi u_1 + \sum_{l=1}^{p} \beta_l x_l\right)$

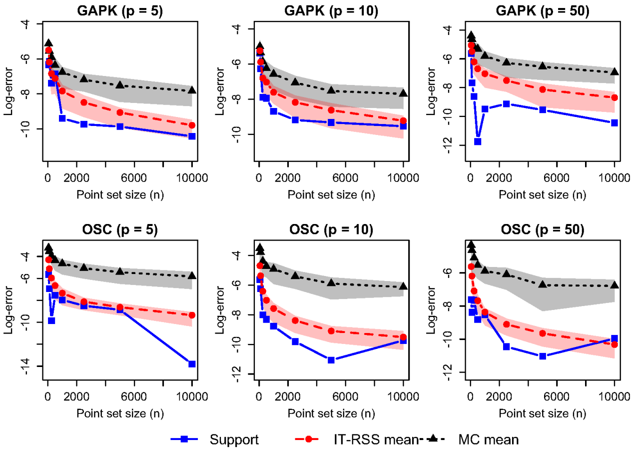

Figure 5: *Log-absolute integration errors for the i.i.d. Exp(1) distribution with $p = 5, 10$ and 50. Shaded bands mark the region between 25-th and 75-th quantiles.*

- F = multivariate $\text{Exp}(1)$.
- Better integration performance compared to QMC and MC.

Simulation: uncertainty propagation

- g: an expensive (computationally, monetarily) simulator. $\mathbf{x}$: input.
- Distribution of $g(\mathbf{x})$ = system output uncertainty.
- Want to estimate this using few simulations.
- Borehole physical model. Simulate the flow rate of water through a borehole.

Input	*Distribution*
Borehole radius	$N(0.01, 0.01618)$
Radius of influence	$\text{Lognormal}(7.71, 1.0056)$
Upper transmissivity	$U[63070, 115600]$
Upper head	$U[990, 1110]$
Lower transmissivity	$U[63.1, 116]$
Lower head	$U[700, 820]$
Borehole length	$U[1120, 1680]$
Borehole conductivity	$U[9855, 12045]$

Table 1: *Inputs and their uncertainty distributions for the borehole model.*

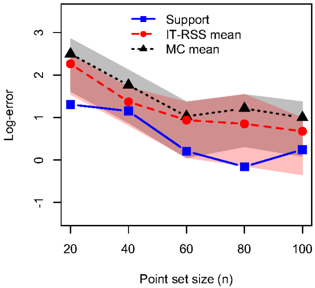

Figure 6: *Log-absolute errors for estimating $\mathbb{E}[g(\mathbf{X})]$ in the borehole model.*

Questions?

Thank you

References I

Gábor J. Székely and Maria L. Rizzo.
Energy statistics: A class of statistics based on distances.
Journal of Statistical Planning and Inference, 143(8):1249–1272, August 2013.
URL
http://www.sciencedirect.com/science/article/pii/S03783758130006