Sequential Detection of Convexity from Noisy Function Evaluations

Nanjing Jian† Shane G. Henderson† Susan R. Hunter‡

† School of Operations Research and Information Engineering, Cornell University

‡ School of Industrial Engineering, Purdue University

INFORMS Annual Meeting, San Francisco, CA

November 9, 2014

Problem Statement

- Consider a function $g : S \subseteq \mathbb{R}^d \to \mathbb{R}$ that can only be evaluated with the presence of noise at r points $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_r) \in \mathbb{R}^d$. Let the true values of the function g at $\boldsymbol{x}$ be denoted $\boldsymbol{g} = (g(\boldsymbol{x}_1), g(\boldsymbol{x}_2), \ldots, g(\boldsymbol{x}_r))^T$.
- We wish to determine the convexity/non-convexity of $\boldsymbol{g}$ with some probabilistic guarantee, using only estimates of its values obtained through simulation at the points $\boldsymbol{x}$.

♠ $\boldsymbol{g}$ is convex if a convex function exists that coincides with $\boldsymbol{g}$ at those points.

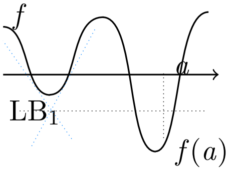
Motivation

- Learning about black-box functions

Motivation

- Learning about black-box functions

- Stopping rule for global (stochastic) optimization algorithms



Motivation

Previous research: One-shot frequentist hypothesis test, with the number of samples predetermined.

Dim	Distance	Regression parameters
1	Juditsky and Nemirovski [2002]	Baraud et al. [2005] Diack and Thomas-Agnan [1998] Meyer [2012] Wang and Meyer [2011]
> 1	Silvapulle and Sen [2001]	Lau [1978] Abrevaya and Jiang [2005]

Motivation

Previous research: One-shot frequentist hypothesis test, with the number of samples predetermined.

Dim	Distance	Regression parameters
1	Juditsky and Nemirovski [2002]	Baraud et al. [2005]
		Diack and Thomas-Agnan [1998]
		Meyer [2012]
		Wang and Meyer [2011]
> 1	Silvapulle and Sen [2001]	Lau [1978]
		Abrevaya and Jiang [2005]

Our Goal: A sequential algorithm in the Bayesian setting with indefinite number of samples and can be stopped at any time.

Assumptions

- We obtain realizations of a random vector $\boldsymbol{Y} = \boldsymbol{f} + \boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim N(\mathbf{0}, \Gamma) \in \mathbb{R}^r$, and Γ positive-definite if known.
- ♠ Γ is not necessarily diagonal for the use of Common Random Numbers.
- Conditional on $\boldsymbol{f}$, the samples $(\boldsymbol{y}_n : n = 1, 2, \ldots)$ in each iteration consists of i.i.d. random vectors.

Main Idea

In each sampling iteration n,

1. Obtain a new set of samples $\boldsymbol{y}$ of the function $\boldsymbol{f}$.

Main Idea

In each sampling iteration n,

1. Obtain a new set of samples $\boldsymbol{y}$ of the function $\boldsymbol{f}$.
2. With a conjugate prior† model, use the posterior of the last iteration as the prior and update the posterior hyper-parameters accordingly.

   † Normal-Normal with uninformative prior when Γ is known, Normal-Inverse-Wishart with Jeffery's Prior when Γ is unknown.

Main Idea

In each sampling iteration n,

1. Obtain a new set of samples $\boldsymbol{y}$ of the function $\boldsymbol{f}$.
2. With a conjugate prior† model, use the posterior of the last iteration as the prior and update the posterior hyper-parameters accordingly.
   † Normal-Normal with uninformative prior when Γ is known, Normal-Inverse-Wishart with Jeffery's Prior when Γ is unknown.
3. Estimate $P(\boldsymbol{f} \in \mathbb{C} | \mathscr{A}_n)$ with a subroutine based on the current posterior distribution.

Main Idea

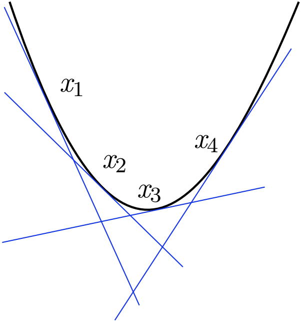
In each sampling iteration n,

1. Obtain a new set of samples $\boldsymbol{y}$ of the function $\boldsymbol{f}$.
2. With a conjugate prior[†] model, use the posterior of the last iteration as the prior and update the posterior hyper-parameters accordingly.
   † Normal-Normal with uninformative prior when Γ is known, Normal-Inverse-Wishart with Jeffery's Prior when Γ is unknown.
3. Estimate $P(\boldsymbol{f} \in \mathbb{C} | \mathscr{A}_n)$ with a subroutine based on the current posterior distribution.

Convergence

Let $p_n = P(\boldsymbol{f} \in \mathbb{C} | \mathscr{A}_n)$ be the n-iteration posterior probability that $\boldsymbol{f}$ is convex. As $n \to \infty$, $p_n - \mathbb{1}\{\boldsymbol{f} \in \mathbb{C}\} \to 0$ a.s.

Convexity



$\boldsymbol{g} \in \mathbb{C}$ if and only if each of the following LS(i),$i \in \{1, \ldots, r\}$

$$\boldsymbol{a}_i^T \boldsymbol{x}_i + b_i = g(\boldsymbol{x}_i)$$
$$\boldsymbol{a}_i^T \boldsymbol{x}_j + b_i \leq g(\boldsymbol{x}_j), \ \forall j \in \{1, \ldots, r\} \backslash \{i\}.$$

is feasible in the variables $\boldsymbol{a}_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$ (Murty [1988]).

Vanilla Monte Carlo Method

In each iteration of the main algorithm, after updating the hyper-parameters of the posterior distribution,

1. Simulate m i.i.d. samples $\boldsymbol{y}_k^n$ from the posterior distribution $\boldsymbol{f}|\mathscr{A}_n$.

Vanilla Monte Carlo Method

In each iteration of the main algorithm, after updating the hyper-parameters of the posterior distribution,

1. Simulate m i.i.d. samples $\boldsymbol{y}_k^n$ from the posterior distribution $\boldsymbol{f}|\mathscr{A}_n$.
2. For each sample $\boldsymbol{y}_k^n$, set $\boldsymbol{g} = \boldsymbol{y}_k^n$ in LS(i), $i = 1, \ldots, r$. Obtain an indicator $\mathbb{1}\left\{\boldsymbol{y}_k^n \in \mathbb{C}\right\}$ that is 0 if any of LS(i), $i = 1, \ldots, r$ is infeasible and 1 otherwise.

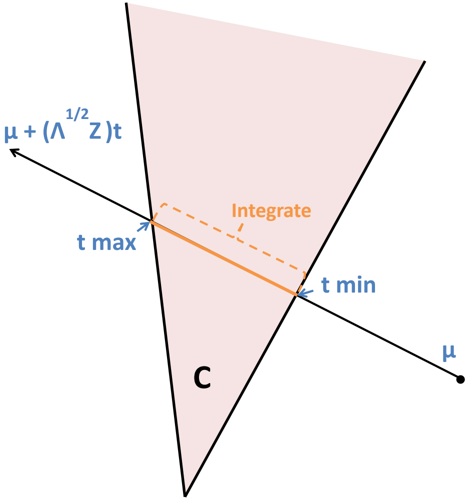
Vanilla Monte Carlo Method

In each iteration of the main algorithm, after updating the hyper-parameters of the posterior distribution,

1. Simulate m i.i.d. samples $\boldsymbol{y}_k^n$ from the posterior distribution $\boldsymbol{f}|\mathscr{A}_n$.
2. For each sample $\boldsymbol{y}_k^n$, set $\boldsymbol{g} = \boldsymbol{y}_k^n$ in $\text{LS}(i), i = 1, \ldots, r$. Obtain an indicator $\mathbb{1}\{\boldsymbol{y}_k^n \in \mathbb{C}\}$ that is 0 if any of $\text{LS}(i), i = 1, \ldots, r$ is infeasible and 1 otherwise.
3. Output the estimator $p_n = \frac{1}{m}\sum_{k=1}^{m} \mathbb{1}\{\boldsymbol{y}_k^n \in \mathbb{C}\}$ as the average of all indicators.

Conditional Monte Carlo Method

Instead of obtaining a $0 - 1$ estimator...



Conditional Monte Carlo Method

This method achieves variance reduction compared to the vanilla Monte Carlo estimator.
At each iteration n,

1. Simulate m i.i.d. samples z_k uniformly on the sphere S^{r-1}.

Conditional Monte Carlo Method

This method achieves variance reduction compared to the vanilla Monte Carlo estimator.
At each iteration n,

1. Simulate m i.i.d. samples z_k uniformly on the sphere S^{r-1}.
2. For each sample z_k, set $\boldsymbol{g} = \mu_n + (\Lambda_n^{1/2} z_k) t_i$ in LS(i), $i = 1, \ldots, r$, and use LS(i) as constraints to solve linear programs with optimal objective values $t_{\min}(i) = \min\ t_i$ and $t_{\max}(i) = \max\ t_i$.

Conditional Monte Carlo Method

This method achieves variance reduction compared to the vanilla Monte Carlo estimator.
At each iteration n,

1. Simulate m i.i.d. samples z_k uniformly on the sphere S^{r-1}.
2. For each sample z_k, set $\boldsymbol{g} = \mu_n + (\Lambda_n^{1/2} z_k) t_i$ in LS(i), $i = 1, \ldots, r$, and use LS(i) as constraints to solve linear programs with optimal objective values $t_{\min}(i) = \min\, t_i$ and $t_{\max}(i) = \max\, t_i$.
3. Set $t_{\min} = \max_{i=1,2,\ldots,r} t_{\min}(i)$ and $t_{\max} = \min_{i=1,2,\ldots,r} t_{\max}(i)$.

Conditional Monte Carlo Method

This method achieves variance reduction compared to the vanilla Monte Carlo estimator.
At each iteration n,

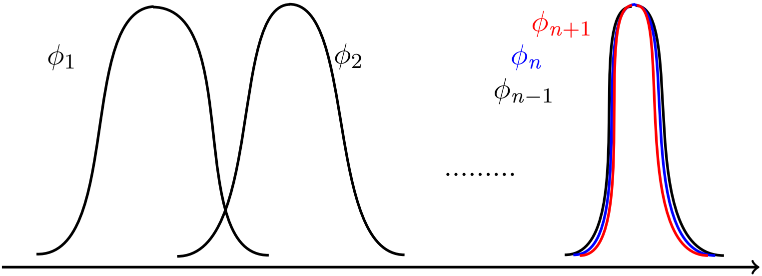
1. Simulate m i.i.d. samples z_k uniformly on the sphere S^{r-1}.
2. For each sample z_k, set $\boldsymbol{g} = \mu_n + (\Lambda_n^{1/2} z_k) t_i$ in LS$(i), i = 1, \dots, r$, and use LS(i) as constraints to solve linear programs with optimal objective values $t_{\min}(i) = \min\ t_i$ and $t_{\max}(i) = \max\ t_i$.
3. Set $t_{\min} = \max_{i=1,2,\dots,r} t_{\min}(i)$ and $t_{\max} = \min_{i=1,2,\dots,r} t_{\max}(i)$.
4. Calculate the integral $\frac{\beta(r)}{2} \int_{t_{\min}}^{t_{\max}} \phi(t z_k) |t^{r-1}| dt$.

Conditional Monte Carlo Method

This method achieves variance reduction compared to the vanilla Monte Carlo estimator.
At each iteration n,

1. Simulate m i.i.d. samples z_k uniformly on the sphere S^{r-1}.
2. For each sample z_k, set $\boldsymbol{g} = \mu_n + (\Lambda_n^{1/2} z_k) t_i$ in LS(i), $i = 1, \ldots, r$, and use LS(i) as constraints to solve linear programs with optimal objective values $t_{\min}(i) = \min\ t_i$ and $t_{\max}(i) = \max\ t_i$.
3. Set $t_{\min} = \max_{i=1,2,\ldots,r} t_{\min}(i)$ and $t_{\max} = \min_{i=1,2,\ldots,r} t_{\max}(i)$.
4. Calculate the integral $\frac{\beta(r)}{2} \int_{t_{\min}}^{t_{\max}} \phi(t z_k) |t^{r-1}| dt$.
5. Output the estimator p_n as a sample average of such integrations obtained from m samples.

Change of Measure Method



As n grows large, $\frac{\phi_{n+\ell}}{\phi_n}$ becomes close to 1, where ϕ_n is the density of $\boldsymbol{f}|\mathscr{A}_n$.

Change of Measure Method

By utilizing the samples and results in an earlier iteration, this method saves computational time.

At iteration n, calculate p_n using vanilla Monte Carlo with samples $\boldsymbol{y}_k, k = 1, \ldots, m$. In iteration $n + \ell$,

Opt 1 Reuse all the samples $\boldsymbol{y}_k^n$ in the n-th iteration: $p_{n+\ell} = \frac{1}{m} \sum_{k=1}^{m} \mathbb{1} \{\boldsymbol{y}_k^n \in \mathbb{C}\} \frac{\phi_{n+\ell}(\boldsymbol{y}_k^n)}{\phi_n(\boldsymbol{y}_k^n)}$.

Opt 2 Randomly reuse part of the samples (say, set S) in the n-th iteration and generate new samples as needed: $p_{n+\ell} = \frac{1}{|S|} \sum_{k \in S} \mathbb{1} \{\boldsymbol{y}_k^n \in \mathbb{C}\} \frac{\phi_{n+\ell}(\boldsymbol{y}_k^n)}{\phi_n(\boldsymbol{y}_k^n)} + \frac{1}{m-|S|} \sum_{k=1}^{m-|S|} \mathbb{1} \left\{\boldsymbol{y}_k^{n+\ell} \in \mathbb{C}\right\}$.

Inverted Bowl

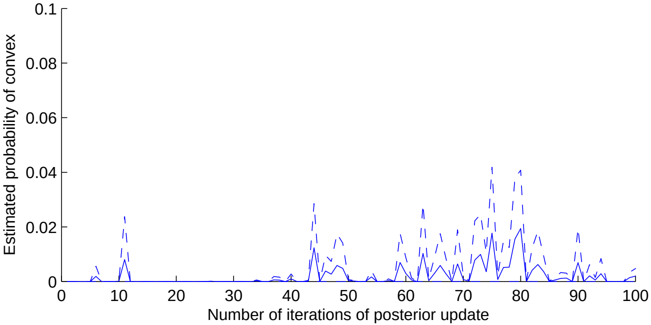


Figure : p_n for $g = -\|\boldsymbol{x}\|^2, \boldsymbol{x} \in [-1, 1]^{30}$, $r = 61$, Γ has 10^4 on the diagonal.

Plane

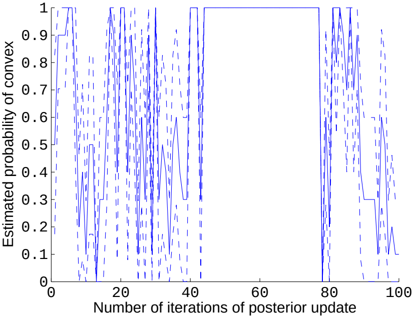


Figure : p_n for $g = 0, \boldsymbol{x} \in [-1, 1]^2$, $r = 5$, Γ has 1 on the diagonal.

Where to sample?

An interesting example for $g = \|\boldsymbol{x}\|^2$, $\boldsymbol{x} \in [-1, 1]^{30}$, $r = 60$, and Γ has 10^4 on the diagonal.

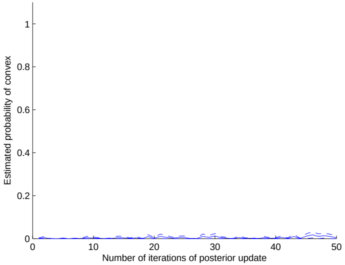


Figure : Sampling the 60 points uniformly at random in space.

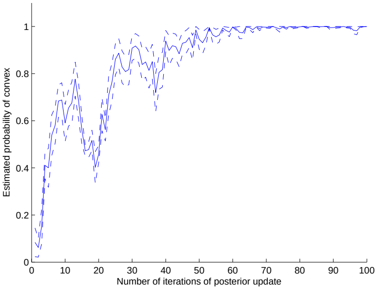


Figure : Sampling along 20 random lines with 3 points on each.

What happened?

For easiness of illustration, consider a 2-dimensional function with the following level sets:

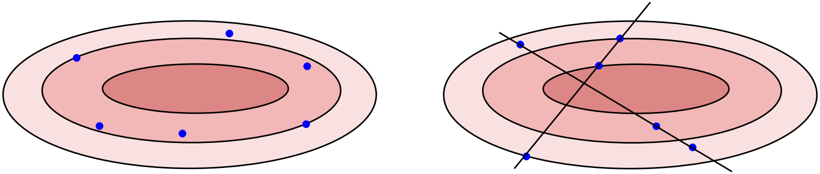


Figure : Sampling uniformly vs. Sampling along random lines.

1. Motivation
2. Algorithm
3. Subroutine Alternatives
4. Numerical Experiments
5. Conclusion

Conclusion

We suggested

- a sequential method for detecting convexity/non-convexity of noisy functions
- a Monte Carlo method for estimating probability of convex
- a conditional Monte Carlo method for variance reduction
- a change of measure method for speed improvement

Next steps:

- the number and locations of sampled points
- uneven sample size at each sampled point

Reference I

Jason Abrevaya and Wei Jiang. A nonparametric approach to measuring and testing curvature. *Journal of Business & Economic Statistics*, 23:1–19, 2005. URL `http://EconPapers.repec.org/RePEc:bes:jnlbes:v:23:y:2005:p:1-19`.

Y. Baraud, S. Huet, and B. Laurent. Testing convex hypotheses on the mean of a Gaussian vector. application to testing qualitative hypotheses on a regression function. *The Annals of Statistics*, 33(1):214–257, 2005.

C. A. T. Diack and C. Thomas-Agnan. A nonparametric test of the non-convexity of regression. *Nonparametric Statistics*, 9:335–362, 1998.

A. Juditsky and A. Nemirovski. On nonparametric tests of positivity/monotonicity/convexity. *The Annals of Statistics*, 30(2):498–527, 2002.

Lawrence J Lau. Testing and imposing monoticity, convexity, and quasi-convexity constraints. *Electronic Journal of Statistics*, 1:409–453, 1978.

Mary C. Meyer. Constrained penalized splines. *Canadian Journal of Statistics*, 40(1):190–206, 2012. ISSN 1708-945X. doi: 10.1002/cjs.10137. URL `http://dx.doi.org/10.1002/cjs.10137`.

K. G. Murty. *Linear Complementarity, Linear and Nonlinear Programming*. Heldermann Verlag, Berlin, 1988.

Mervyn J. Silvapulle and Pranab K. Sen. *Constrained Statistical Inference: Order, Inequality, and Shape Restrictions*, chapter 3, pages 59–141. John Wiley & Sons, Inc., 2001. ISBN 9781118165614. doi: 10.1002/9781118165614.ch3. URL `http://dx.doi.org/10.1002/9781118165614.ch3`.

Jianqiang C. Wang and Mary C. Meyer. Testing the monotonicity or convexity of a function using regression splines. *Canadian Journal of Statistics*, 39(1):89–107, 2011. ISSN 1708-945X. doi: 10.1002/cjs.10094. URL `http://dx.doi.org/10.1002/cjs.10094`.