Estimating the Probability of Convexity of a Function Observed with Noise

Nanjing Jian
Shane G. Henderson

School of Operations Research and Information Engineering,
Cornell University

INFORMS Annual Meeting, Philadelphia, PA
November 3, 2015

Problem Statement

- Consider a function g: compact $S \subseteq \mathbb{R}^d \to \mathbb{R}$ 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 $\boldsymbol{x}$.

The Definition of "Convexity"

Is $g(\cdot)$ convex?

The Definition of "Convexity"

Is $g(\cdot)$ convex?

1. Infinite-dimensional
   Hilbert Space

The Definition of "Convexity"

Is $g(\cdot)$ convex?

1. Infinite-dimensional Hilbert Space

2. Usual definition

The Definition of "Convexity"

Is $g(\cdot)$ convex?

1. Infinite-dimensional Hilbert Space

2. Usual definition

3. Not computationally tractable: we are working on the theories.

The Definition of "Convexity"

Is $g(\cdot)$ convex?

1. Infinite-dimensional Hilbert Space

2. Usual definition

3. Not computationally tractable: we are working on the theories.

Not in this talk!

The Definition of "Convexity"

Is $g(\cdot)$ convex?

1. Infinite-dimensional Hilbert Space

2. Usual definition

3. Not computationally tractable: we are working on the theories.

Not in this talk!

In this talk: Is $\boldsymbol{g}$ convex?

The Definition of "Convexity"

Is $g(\cdot)$ convex?

1. Infinite-dimensional Hilbert Space
2. Usual definition
3. Not computationally tractable: we are working on the theories.

Not in this talk!

In this talk: Is $\boldsymbol{g}$ convex?

1. Finite-dimensional Euclidean Space

The Definition of "Convexity"

Is $g(\cdot)$ convex?

1. Infinite-dimensional Hilbert Space
2. Usual definition
3. Not computationally tractable: we are working on the theories.

Not in this talk!

In this talk: Is $\boldsymbol{g}$ convex?

1. Finite-dimensional Euclidean Space
2. $\boldsymbol{g}$ is convex if a convex function exists that coincides with $\boldsymbol{g}$ at those points.

The Definition of "Convexity"

Is $g(\cdot)$ convex?

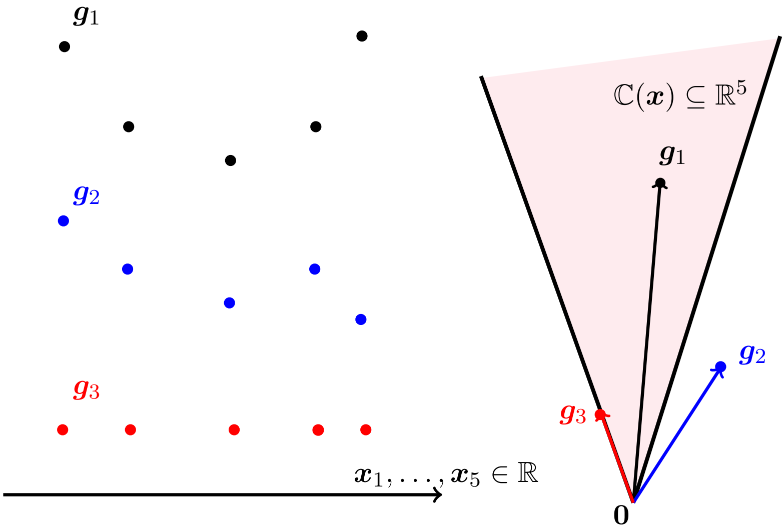
1. Infinite-dimensional Hilbert Space
2. Usual definition
3. Not computationally tractable: we are working on the theories.

Not in this talk!

In this talk: Is $\boldsymbol{g}$ convex?

1. Finite-dimensional Euclidean Space
2. $\boldsymbol{g}$ is convex if a convex function exists that coincides with $\boldsymbol{g}$ at those points.
3. More tractable

Vector Convexity and Cone of $\mathbb{C}$



Motivation

- Curiosity towards black-box functions

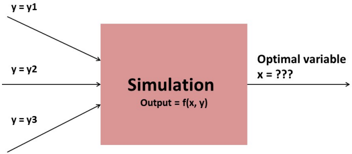
"Sorry, it's curiosity"

Motivation

- Curiosity towards black-box functions

"Sorry, it's curiosity"

- Ease of Optimization

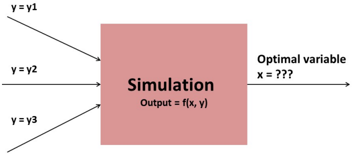


Motivation

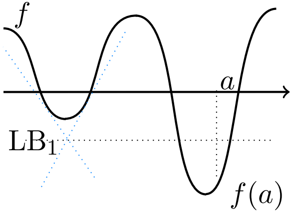
- Curiosity towards black-box functions

"Sorry, it's curiosity"

- Ease of Optimization



- Stopping rule for global (stochastic) optimization algorithms



Motivation

Previous research: One-shot frequentist hypothesis test, with the number of samples predetermined [Jian, Henderson, and Hunter, 2014].

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 [Jian, Henderson, and Hunter, 2014].

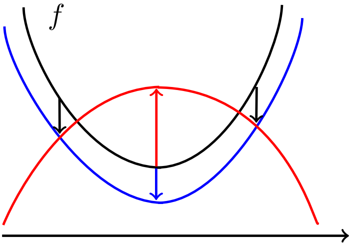
Dim	Distance	Regression parameters
1	Juditsky and Nemirovski [2002]	Baraud et al. [2005] Diack and Thomas-Agnan [1998] Meyer [2012] Wang and Meyer [2011]
> 1	Silvapulle and Sen [2001]	Lau [1978] Abrevaya and Jiang [2005]

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

1. Algorithm
2. Monte Carlo
3. Conditional MC
4. LR Based Methods
5. Implementations
6. Conclusion

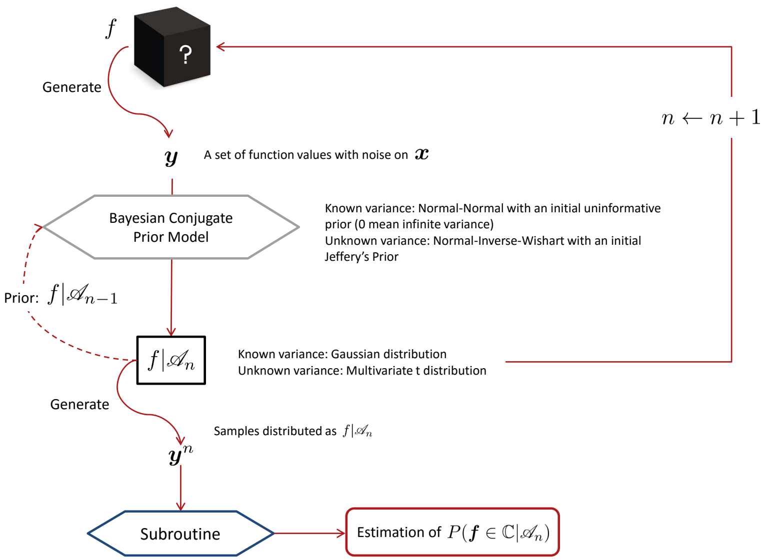
Assumptions

1. We obtain realizations of a random vector $\boldsymbol{Y} = \boldsymbol{f} + \boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim N(\mathbf{0}, \Gamma) \in \mathbb{R}^r$,with Γ positive-definite.
2. Conditional on $\boldsymbol{f}$, the samples $(\boldsymbol{y}_n : n = 1, 2, \ldots)$ in each iteration consists of i.i.d. random vectors.



Note that Γ is not necessarily diagonal because using Common Random Numbers can maintain the function structural properties (e.g. Chen et al. [2012]).

Main Idea



f
?
Generate
$n \leftarrow n+1$
$\boldsymbol{y}$
A set of function values with noise on $\boldsymbol{x}$
Bayesian Conjugate Prior Model
Known variance: Normal-Normal with an initial uninformative prior (0 mean infinite variance)
Unknown variance: Normal-Inverse-Wishart with an initial Jeffery's Prior
Prior: $f|\mathscr{A}_{n-1}$
$f|\mathscr{A}_n$
Known variance: Gaussian distribution
Unknown variance: Multivariate t distribution
Generate
Samples distributed as $f|\mathscr{A}_n$
$\boldsymbol{y}^n$
Subroutine
Estimation of $P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_n)$

Convergence

Theorem

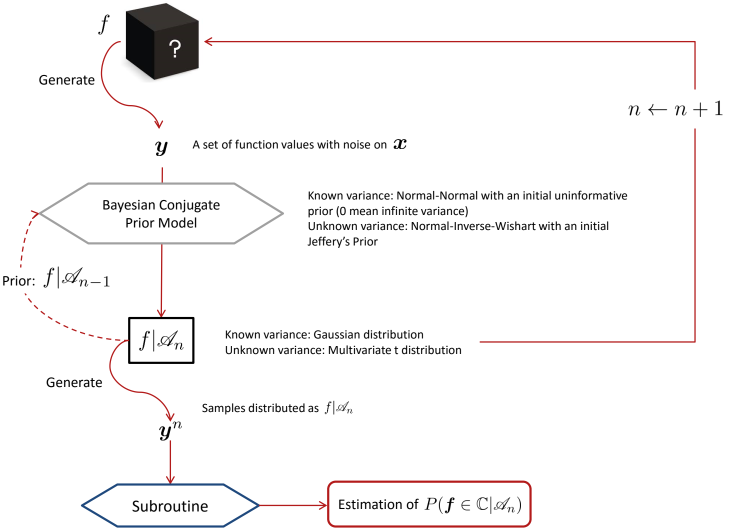
Let $p_n = P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_n)$ be the n-iteration posterior probability that $\boldsymbol{f}$ is convex. As $n \to \infty$, $p_n - \mathbb{1}\{\boldsymbol{f} \in \mathbb{C}\} \to 0$ a.s.

Proof sketch (assume Γ known):

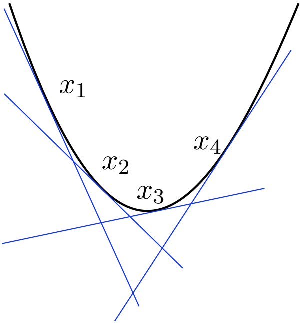
1. $P(\boldsymbol{f} \in \partial\mathbb{C}) = 0$ leaves cases $\boldsymbol{f} \notin \mathbb{C}$ and $\boldsymbol{f} \in \mathbb{C}^\circ$.
2. $\mu_n - \boldsymbol{f} \to 0$ in probability and $\Lambda_n \sim \Gamma/n \to \mathbf{0}$ as $n \to \infty$ (CLT).
3. When $\boldsymbol{f} \notin \mathbb{C}$, define $D_{\boldsymbol{f}} = \min_{\boldsymbol{h}\in\mathbb{C}} ||\boldsymbol{f} - \boldsymbol{h}||$, then $p_n - \mathbb{1}\{\boldsymbol{f} \in \mathbb{C}\} = P(\mu_n + \Lambda_n^{1/2} Z \in \mathbb{C}|\mathscr{A}_n) \leq P(\|\mu_n + \Lambda_n^{1/2} Z - \boldsymbol{f}\| \geq D_{\boldsymbol{f}}|\mathscr{A}_n) \leq P(\|\mu_n - \boldsymbol{f}\| \geq D_{\boldsymbol{f}}/2|\mathscr{A}_n) + P(\|\Lambda_n^{1/2} Z\| \geq D_{\boldsymbol{f}}/2|\mathscr{A}_n) \to_p 0$ by Markov's Inequality. A lower bound of 0 can be given when $\boldsymbol{f} \in \mathbb{C}^\circ$.
4. $(p_n : n \geq 0)$ is a uniformly integrable martingale, so the convergence is almost surely.

For the unknown Γ case, a similar proof can be constructed by conditioning on Γ.

Main Idea


f
?
Generate
$\boldsymbol{y}$
A set of function values with noise on $\boldsymbol{x}$
$n \leftarrow n+1$
Bayesian Conjugate Prior Model
Known variance: Normal-Normal with an initial uninformative prior (0 mean infinite variance)
Unknown variance: Normal-Inverse-Wishart with an initial Jeffery's Prior
Prior: $f|\mathscr{A}_{n-1}$
$f|\mathscr{A}_n$
Known variance: Gaussian distribution
Unknown variance: Multivariate t distribution
Generate
Samples distributed as $f|\mathscr{A}_n$
$\boldsymbol{y}^n$
Subroutine
Estimation of $P(\boldsymbol{f} \in \mathbb{C}|\mathscr{A}_n)$

Convexity



$\boldsymbol{g} \in \mathbb{C}$ if and only if each of the following linear system

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

Challenge: Variance Reduction

Vanilla Monte Carlo estimation converges slowly with wide confidence interval width:

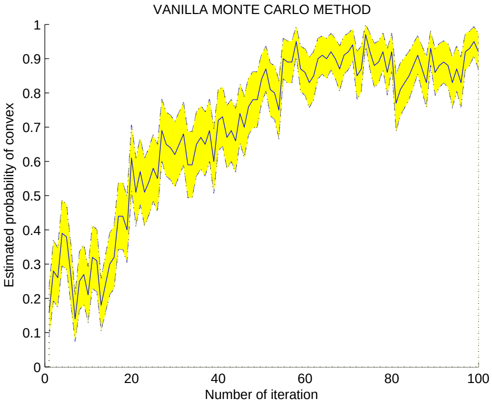
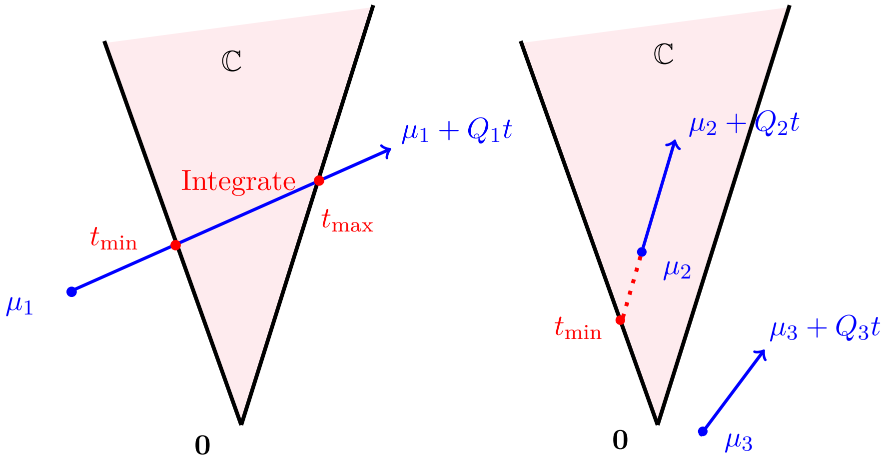


Figure: 3-d sphere function with 12 sample points.

Conditional Monte Carlo Method

Let $Q = \Lambda_n^{1/2} Z$ where $Z \sim Uniform(S^{r-1})$. Instead of obtaining a $0-1$ estimator...



Conditional Monte Carlo Method

How does this work?

$$P(\boldsymbol{f} \in \mathbb{C} | \mathscr{A}_n)$$
$$= E_n\left(\mathbb{1}\left\{\Lambda_n^{1/2} X + \mu_n \in \mathbb{C}\right\}\right), \text{ for } X \sim N(0, I) \text{ or } t_{\nu_n}(0, I)$$
$$= E_n\left(\mathbb{1}\left\{T\Lambda_n^{1/2} Z + \mu_n \in \mathbb{C}\right\}\right), \text{ for } Z \text{ uniform on } S^{r-1}$$
$$= E_n\left(E\left(\mathbb{1}\left\{T\Lambda_n^{1/2} Z + \mu_n \in \mathbb{C}\right\} | Z\right)\right)$$
$$= E_n(P(T \in [t_{\min}(Z), t_{\max}(Z)] \, | Z))$$
$$= E_n(F_{T|Z}(t_{\max}(Z)) - F_{T|Z}(t_{\min}(Z)))$$

Distribution of $T|Z$

Distribution of $T|Z$

Known variance: $F_{T|Z}(t) = \frac{1}{2} + \text{sign}(t) F_{\chi^2_r}(t^2)$.
Unknown variance: $F_{T|Z}(t) = \frac{1}{2} + \text{sign}(t) F_{F(r,\nu_n)}(t^2/r)$.

Proof sketch:

$$
\begin{aligned}
&F_{T|Z}(t) = P(T \leq t|Z) \\
=&P(T \leq 0|Z) + P(0 \leq T \leq t|Z), \text{ when } t > 0 \\
=&1/2 + 1/2 P(||X||^2 \leq t^2|Z), \text{ for } X = TZ
\end{aligned}
$$

Known variance: $X \sim N(0, I)$, so $||X||^2 \sim \chi^2_r$.

Unknown variance: $X \sim t_{\nu_n}(0, I) = \frac{N}{\sqrt{Y/\nu_n}}$ for $N \sim N(0, I)$ and $Y \sim \chi^2_{\nu_n}$. Therefore $||X||^2 = \frac{N^T N}{Y/\nu_n}$ where $N^T N \sim \chi^2_r$, so $||X||^2/r \sim F(r, \nu_n)$.

Finding $t_{\min}$ and $t_{\max}$

LP for $t_{\min}$ and $t_{\max}$

$$\begin{aligned} & t_{\min} = \min t \quad (t_{\max} = \max t) \\ \text{s.t.} \quad & \boldsymbol{a}_i^T \boldsymbol{x} + \boldsymbol{b}_i = \mu_i + (\Lambda^{1/2} Z)_i t, \ \forall i \\ & \boldsymbol{a}_i^T \boldsymbol{x}_j + \boldsymbol{b}_i \leq \mu_j + (\Lambda^{1/2} Z)_j t, \ \forall i, \forall j \neq i \\ & \boldsymbol{a}_i \in \mathbb{R}^r, \boldsymbol{b}_i \in \mathbb{R}, t \in \mathbb{R}, \ \forall i \end{aligned}$$

Similarly, this can be decomposed into r subproblems:

$$\begin{aligned} & t^i_{\min} = \min t^i \\ & (t^i_{\max} = \max t^i) \\ \text{s.t.} \quad & \boldsymbol{a}_i^T \boldsymbol{x}_i + b_i = \mu_i + (\Lambda^{1/2} Z) t^i \\ & \boldsymbol{a}_i^T \boldsymbol{x}_j + b_i \leq \mu_j + (\Lambda^{1/2} Z)_j t^i, \\ & \forall j \neq i, \end{aligned}$$

and

$$t_{\min} = \max_{i=1,2,\ldots,r} t^i_{\min},$$

$$t_{\max} = \min_{i=1,2,\ldots,r} t^i_{\max}.$$

Conditional Monte Carlo Method

At each iteration n, to estimate
$p_n = E_n(F_{T|Z}(t_{\max}(Z)) - F_{T|Z}(t_{\min}(Z)))$,

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

Conditional Monte Carlo Method

At each iteration n, to estimate
$p_n = E_n(F_{T|Z}(t_{\max}(Z)) - F_{T|Z}(t_{\min}(Z)))$,

1. Simulate m i.i.d. samples z_k uniformly on the unit shell S^{r-1}.
2. For each sample z_k, solve for $t_{\min}(z_k)$ and $t_{\max}(z_k)$.

Conditional Monte Carlo Method

At each iteration n, to estimate
$p_n = E_n(F_{T|Z}(t_{\max}(Z)) - F_{T|Z}(t_{\min}(Z)))$,

1. Simulate m i.i.d. samples z_k uniformly on the unit shell S^{r-1}.
2. For each sample z_k, solve for $t_{\min}(z_k)$ and $t_{\max}(z_k)$.
3. Calculate $F_{T|Z}(t_{\max}(z_k)) - F_{T|Z}(t_{\min}(z_k))$.

Conditional Monte Carlo Method

At each iteration n, to estimate $p_n = E_n(F_{T|Z}(t_{\max}(Z)) - F_{T|Z}(t_{\min}(Z)))$,

1. Simulate m i.i.d. samples z_k uniformly on the unit shell S^{r-1}.
2. For each sample z_k, solve for $t_{\min}(z_k)$ and $t_{\max}(z_k)$.
3. Calculate $F_{T|Z}(t_{\max}(z_k)) - F_{T|Z}(t_{\min}(z_k))$.
4. Output the estimator $\hat{p_n}$ as a sample average of such integrations obtained from m samples.

Conditional Monte Carlo Method

At each iteration n, to estimate
$p_n = E_n(F_{T|Z}(t_{\max}(Z)) - F_{T|Z}(t_{\min}(Z)))$,

1. Simulate m i.i.d. samples z_k uniformly on the unit shell S^{r-1}.
2. For each sample z_k, solve for $t_{\min}(z_k)$ and $t_{\max}(z_k)$.
3. Calculate $F_{T|Z}(t_{\max}(z_k)) - F_{T|Z}(t_{\min}(z_k))$.
4. Output the estimator $\hat{p_n}$ as a sample average of such integrations obtained from m samples.

This method achieves variance reduction compared to the vanilla Monte Carlo estimator.

Reduced Variance and New Challenge

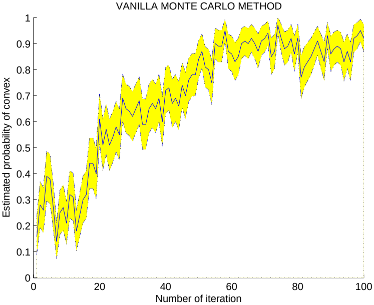


Figure: Vanilla Monte Carlo method.

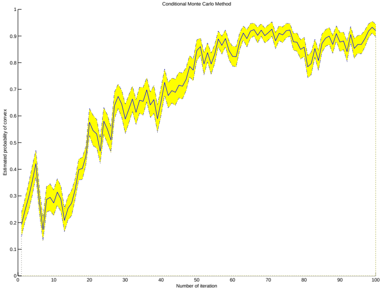
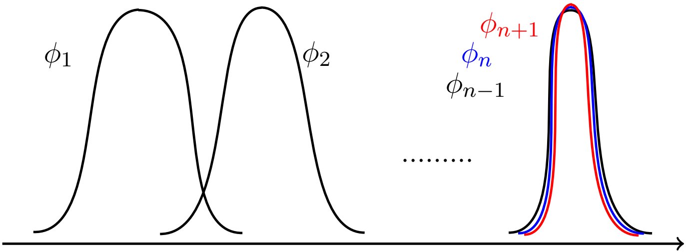


Figure: Conditional Monte Carlo method.

... but doubled the computational time to check the convexity for each sample!

Can we reuse samples?



Maybe yes: as n grows large, we might expect $\frac{\phi_{n+1}}{\phi_n}$ to become close to 1, where ϕ_n is the density of $\boldsymbol{f}|\mathscr{A}_n$. Thus p_{n+1} may be estimated by $\boldsymbol{y}_k^n, k = 1, \ldots, m$, for which $\mathbb{1}\{\boldsymbol{y}_k^n \in \mathbb{C}\}$ has been calculated in iteration n.

LR Based Methods

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

LR Based Methods

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

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

LR Based Methods

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

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

Generalized Acceptance/Rejection: Reuse part of the samples by accepting $\boldsymbol{y}_k^n$ with probability $\frac{\phi_{n+\ell}(\boldsymbol{y}_k^n)}{c\phi_n(\boldsymbol{y}_k^n)}$, then generate new samples as needed:
$\hat{p_{n+\ell}} = \frac{1}{m} \left(\sum_{k \in S} \mathbb{1}\left\{\boldsymbol{y}_k^n \in \mathbb{C}\right\} + \sum_{k=1}^{m-|S|} \mathbb{1}\left\{\boldsymbol{y}_k^{n+\ell} \in \mathbb{C}\right\} \right)$

LR Based Methods

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

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

Generalized Acceptance/Rejection: Reuse part of the samples by accepting $\boldsymbol{y}_k^n$ with probability $\frac{\phi_{n+\ell}(\boldsymbol{y}_k^n)}{c\phi_n(\boldsymbol{y}_k^n)}$, then generate new samples as needed:

$$\hat{p_{n+\ell}} = \frac{1}{m}\left(\sum_{k \in S} \mathbb{1}\left\{\boldsymbol{y}_k^n \in \mathbb{C}\right\} + \sum_{k=1}^{m-|S|} \mathbb{1}\left\{\boldsymbol{y}_k^{n+\ell} \in \mathbb{C}\right\}\right)$$

By utilizing the samples and results in an earlier iteration, these two methods saves computational time.

Theory vs. Reality

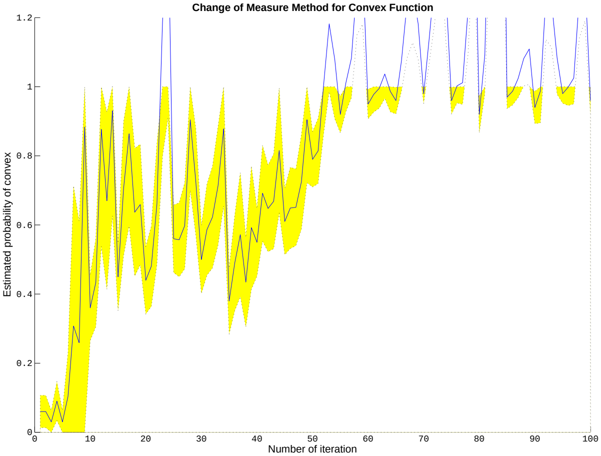


Figure: The probability of convex is estimated to be over 1.

Theory vs. Reality

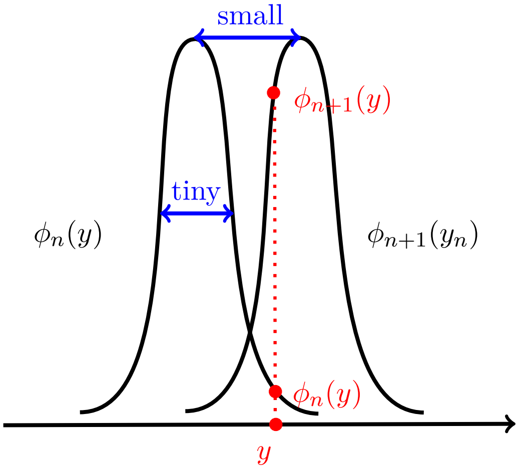


Figure: When $\frac{\phi_{n+1}(\boldsymbol{y})}{\phi_n(\boldsymbol{y})}$ is huge.

How likely does this happen?

It can be shown that $\hat{p}_{n+\ell}$ is unbiased and has finite variance [Jian et al., 2014] . However, when n becomes large, $\max_{y^n} \frac{\phi_{n+1}(y^n)}{\phi_n(y^n)} = c$ has infinite expectation.

Proof sketch: For example, assume $r = 1$, and $\Gamma = \lambda^{-2}$ is known, so the n-iteration posterior precision is $\sigma_n^{-2} = \sigma_0^{-2} + n\lambda^2$. Let $y \sim N(0, \lambda^{-2})$, and $y^n \sim N(\mu_n, \sigma_n^2)$.

$\max_{y^n} 2 \ln \frac{\phi_{n+1}(y^n)}{\phi_n(y^n)} = (\mu_n - \mu_{n+1})^2(\sigma_{n+1}^{-2} + \frac{1}{\lambda^2}(\sigma_{n+1}^{-4})) + \ln \frac{\sigma_n}{\sigma_{n+1}}$,

$(\mu_n - \mu_{n+1})^2 = ((1 - \frac{\sigma_{n+1}^2}{\sigma_n^2})\mu_n - \sigma_{n+1}^2 \lambda^2 y)^2 \to (\sigma_{n+1}^2 \lambda^2 y)^2$ and $\sigma_n/\sigma_{n+1} \to 1$. So $\max_{y^n} 2 \ln \frac{\phi_{n+1}(y^n)}{\phi_n(y^n)} \to \lambda^2 y^2 \sim \chi_1^2$.

Thus $c \geq \exp \frac{\lambda^2 y^2}{2} \sim \exp \frac{\chi_1^2}{2}$ with expectation ∞.

The Combination of Methods

Combining the Conditional Monte Carlo method with the Change of Measure method, we get an hybrid algorithm:

1. At every iteration n, simulate m i.i.d. samples z_k uniformly on the unit shell S^{r-1}. Solve for $t_{\min}(z_k)$ and $t_{\max}(z_k)$ and output the Conditional Monte Carlo estimator $\hat{p_n}$.

The Combination of Methods

Combining the Conditional Monte Carlo method with the Change of Measure method, we get an hybrid algorithm:

1. At every iteration n, simulate m i.i.d. samples z_k uniformly on the unit shell S^{r-1}. Solve for $t_{\min}(z_k)$ and $t_{\max}(z_k)$ and output the Conditional Monte Carlo estimator $\hat{p_n}$.
2. Let $\boldsymbol{y}_k = \mu_n + (\Lambda_n^{1/2} z_k) t_k$, where t_k is sampled from $T|Z = z_k$. If $t_k \in [t_{\min}(z_k), t_{\max}(z_k)]$, then $\mathbb{1}\{\boldsymbol{y}_k^n \in \mathbb{C}\} = 1$; otherwise, $\mathbb{1}\{\boldsymbol{y}_k^n \in \mathbb{C}\} = 0$.

The Combination of Methods

Combining the Conditional Monte Carlo method with the Change of Measure method, we get an hybrid algorithm:

1. At every iteration n, simulate m i.i.d. samples z_k uniformly on the unit shell S^{r-1}. Solve for $t_{\min}(z_k)$ and $t_{\max}(z_k)$ and output the Conditional Monte Carlo estimator $\hat{p_n}$.
2. Let $\boldsymbol{y}_k = \mu_n + (\Lambda_n^{1/2} z_k) t_k$, where t_k is sampled from $T|Z = z_k$. If $t_k \in [t_{\min}(z_k), t_{\max}(z_k)]$, then $\mathbb{1}\{\boldsymbol{y}_k^n \in \mathbb{C}\} = 1$; otherwise, $\mathbb{1}\{\boldsymbol{y}_k^n \in \mathbb{C}\} = 0$.
3. In the next ℓ iterations, output $\hat{p}_{n+\ell} = \frac{1}{m} \sum_{k=1}^{m} \mathbb{1}\{\boldsymbol{y}_k^n \in \mathbb{C}\} \frac{\phi_{n+\ell}(\boldsymbol{y}_k^n)}{\phi_n(\boldsymbol{y}_k^n)}$.

The Combination of Methods

Combining the Conditional Monte Carlo method with the Change of Measure method, we get an hybrid algorithm:

1. At every iteration n, simulate m i.i.d. samples z_k uniformly on the unit shell S^{r-1}. Solve for $t_{\min}(z_k)$ and $t_{\max}(z_k)$ and output the Conditional Monte Carlo estimator $\hat{p_n}$.
2. Let $\boldsymbol{y}_k = \mu_n + (\Lambda_n^{1/2} z_k) t_k$, where t_k is sampled from $T|Z = z_k$. If $t_k \in [t_{\min}(z_k), t_{\max}(z_k)]$, then $\mathbb{1}\{\boldsymbol{y}_k^n \in \mathbb{C}\} = 1$; otherwise, $\mathbb{1}\{\boldsymbol{y}_k^n \in \mathbb{C}\} = 0$.
3. In the next ℓ iterations, output $\hat{p}_{n+\ell} = \frac{1}{m} \sum_{k=1}^{m} \mathbb{1}\{\boldsymbol{y}_k^n \in \mathbb{C}\} \frac{\phi_{n+\ell}(\boldsymbol{y}_k^n)}{\phi_n(\boldsymbol{y}_k^n)}$.

(Careful! $\frac{\phi_{n+\ell}(\boldsymbol{y}_k^n)}{\phi_n(\boldsymbol{y}_k^n)} \gg 1$ may happen.)

Inverted Bowl

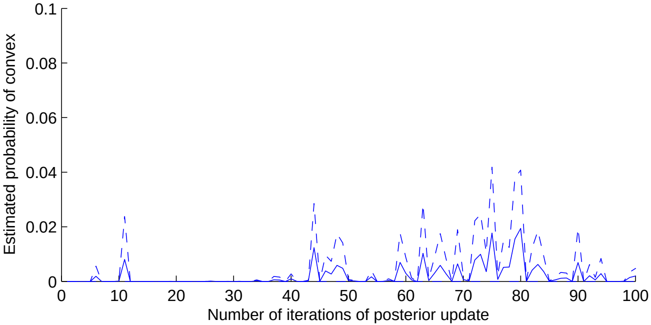


Figure: $\hat{p_n}$ for $g = -\|\boldsymbol{x}\|^2, \boldsymbol{x} \in [-1,1]^{30}$, $r = 61$, Γ has 10^4 on the diagonal.

Plane

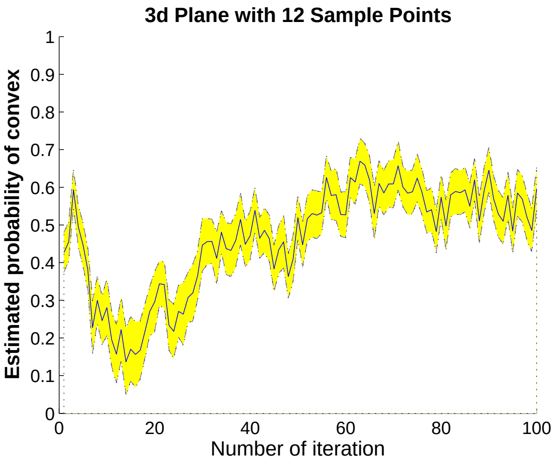


Figure: $\hat{p_n}$ for $g = 0, \boldsymbol{x} \in [-1, 1]^2$, $r = 6$, Γ has $1e-18$ on the diagonal.

Ambulances in a Square

A "real" example from SimOpt.org: What does the long run average response time behave like as a function of the ambulance base location?

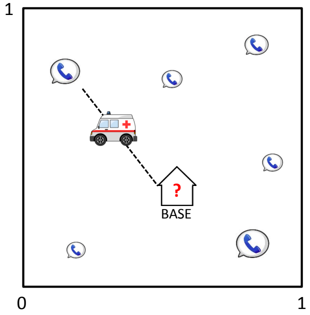


Figure: Problem illustration.

One Base, Two Bases

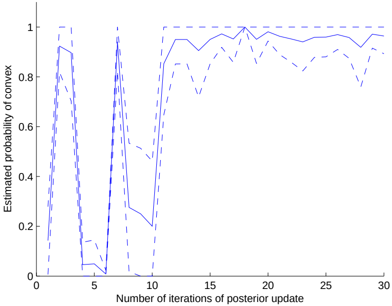


Figure: $\hat{p_n}$ for one ambulance base $(d = 2)$.

One Base, Two Bases

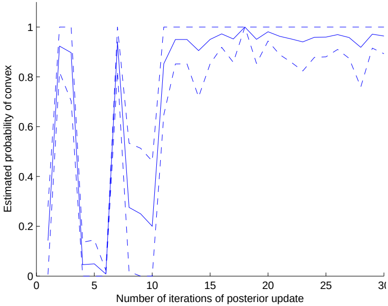


Figure: $\hat{p_n}$ for one ambulance base ($d = 2$).

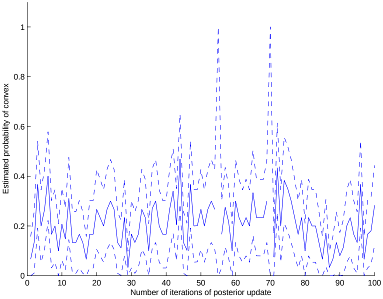


Figure: $\hat{p_n}$ for two ambulance bases ($d = 4$).

Conclusion

In this talk:

- a sequential Bayesian method for estimating the probability of convex utilizing Monte Carlo simulation
- a Conditional Monte Carlo variance reduction method
- two likelihood-ratio-based variance reduction methods with limitation

Conclusion

In this talk:

- a sequential Bayesian method for estimating the probability of convex utilizing Monte Carlo simulation
- a Conditional Monte Carlo variance reduction method
- two likelihood-ratio-based variance reduction methods with limitation

On-going research:

- The asymptotic behavior of the probability that a function is convex using a kriging model as the number of samples goes to infinity.

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.

Xi Chen, Bruce E. Ankenman, and Barry L. Nelson. The effects of common random numbers on stochastic kriging metamodels. *ACM Trans. Model. Comput. Simul.*, 22(2):7:1–7:20, March 2012. ISSN 1049-3301. doi: 10.1145/2133390.2133391. URL http://doi.acm.org/10.1145/2133390.2133391.

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

Nanjing Jian, Shane G. Henderson, and Susan R. Hunter. Sequential detection of convexity from noisy function evaluations. In *Proceedings of the 2014 Winter Simulation Conference*, WSC '14, pages 3892–3903, 2014. URL http://dl.acm.org/citation.cfm?id=2693848.2694332.

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.

Reference II

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`.

Constant for the Generalized A/R

How to find c in the generalized acceptance/rejection method, so that $c \geq \frac{\phi_{n+\ell}(\boldsymbol{y})}{\phi_n(\boldsymbol{y})}$ for all y?

Known variance: $c \geq$
$$\left\{ \frac{|\Lambda_{n+\ell}|}{|\Lambda_n|} \exp \frac{1}{\ell} (\Lambda_{n+\ell}^{-1}\mu_{n+\ell} - \Lambda_n^{-1}\mu_n)^T \Gamma (\Lambda_{n+\ell}^{-1}\mu_{n+\ell} - \Lambda_n^{-1}\mu_n) + \mu_n^T \Lambda_n^{-1}\mu_n - \mu_{n+\ell}^T \Lambda_{n+\ell}^{-1}\mu_{n+\ell} \right\}^{1/2}.$$

Unknown variance: Intractable and needs numerical estimation.

The Dual

$Q = \Lambda^{1/2} Z.$

Primal: $t_{\min} = \min t$

s.t. $\boldsymbol{a}_i^T \boldsymbol{x} + \boldsymbol{b}_i = \mu_i + Q_i t,\ \forall i$

$\boldsymbol{a}_i^T \boldsymbol{x}_j + \boldsymbol{b}_i \le \mu_j + Q_j t,$

$\forall i, \forall j \neq i$

$\boldsymbol{a}_i \in \mathbb{R}^r, \boldsymbol{b}_i \in \mathbb{R}, t \in \mathbb{R},\ \forall i$

Dual: $t_{\min} = -\min \sum_i \sum_j v_{ij} \mu_j$

s.t.

$$\sum_i \sum_j v_{ij} Q_j = 1, \quad (1)$$

$$\sum_j \boldsymbol{x}_j v_{ij} = \mathbf{0}^d,\ \forall i, \quad (2)$$

$$\sum_j v_{ij} = \mathbf{0}^r,\ \forall i, \quad (3)$$

$$v_{ij} \ge 0,\ \forall i, \forall j \neq i, \quad (4)$$

Meaning

min distance from μ to $\mathbb{C}$

$(vQ)_i \neq 0$ for the nearest face of $\mathbb{C}$

$\sum_j \boldsymbol{x}_j \frac{v_{ij}}{-v_{ii}} = \mathbf{0}^d, \forall i$

$\frac{v_{ij}}{-v_{ii}} \ge 0,\ \sum_j \frac{v_{ij}}{-v_{ii}} = 1$

$v_{i,\cdot}$ characterizes faces of $\mathbb{C}$

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.

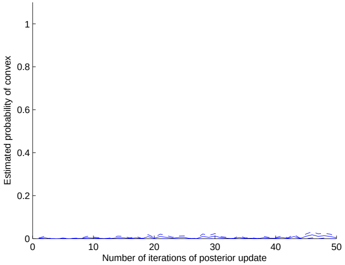


Figure: Sampling the 60 points uniformly at random in space.

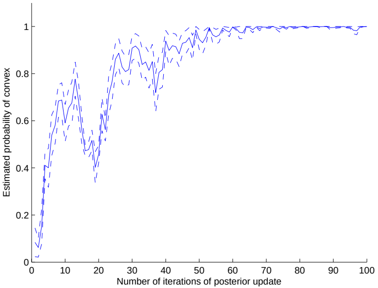


Figure: Sampling along 20 random lines with 3 points on each.

Where to sample?

Consider a 2-dimensional concave function with the following level sets (higher values in darker shades):

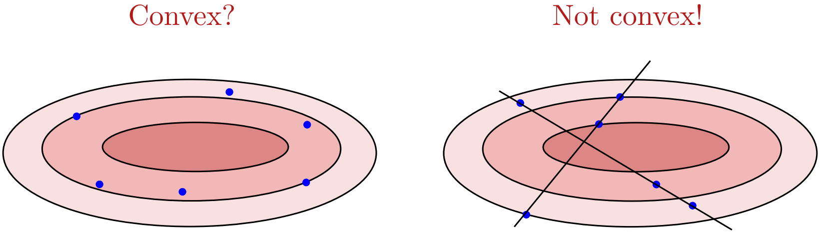


Figure: Sampling uniformly vs. Sampling along random lines.