Quantile estimation under monotonicity constraint

MASCOTNUM

Saint Etienne - 08/04/2015

Vincent Moutoussamy - EDF R&D - MRI

Nicolas Bousquet, Bertrand Iooss - EDF R&D - MRI

Fabrice Gamboa, Thierry Klein - Université Paul Sabatier

Context

- Consider $g : \boldsymbol{R}^d \to \boldsymbol{R}$

Context

- Consider $g : \boldsymbol{R}^d \to \boldsymbol{R}$
- Consider an input $\mathbf{X}$ as a random vector

Context

- Consider $g : \boldsymbol{R}^d \rightarrow \boldsymbol{R}$
- Consider an input $\mathbf{X}$ as a random vector $\Rightarrow$ $g(\mathbf{X})$ is a random variable

Context

- Consider $g : \boldsymbol{R}^d \rightarrow \boldsymbol{R}$
- Consider an input $\mathbf{X}$ as a random vector $\Rightarrow$ $g(\mathbf{X})$ is a random variable
- An undesirable event is described as $\{g(\mathbf{X}) \leq q\}$

Context

- Consider $g : \boldsymbol{R}^d \rightarrow \boldsymbol{R}$
- Consider an input $\mathbf{X}$ as a random vector $\Rightarrow\ g(\mathbf{X})$ is a random variable
- An undesirable event is described as $\{g(\mathbf{X}) \leq q\}$
- Denote $F(t) = \mathbb{P}(g(\mathbf{x}) \leq t)$, the p-quantile is defined by

$$q = \inf\{t \in \boldsymbol{R} :\ F(t) > p\}.$$

Context

- Consider $g : \boldsymbol{R}^d \to \boldsymbol{R}$

- Consider an input $\mathbf{X}$ as a random vector $\Rightarrow$ $g(\mathbf{X})$ is a random variable

- An undesirable event is described as $\{g(\mathbf{X}) \leq q\}$

- Denote $F(t) = \mathbb{P}(g(\mathbf{x}) \leq t)$, the p-quantile is defined by

$$q = \inf\{t \in \boldsymbol{R} : \; F(t) > p\}.$$

- Knowing the behaviour of the component, a **monotonic** hypothesis is studied

- The monotonicity gives **sure bounds** of q.

Context

- Without lost of generality, assume
  - $g : [0,1]^d \to \boldsymbol{R}$ is a globally increasing function: $\mathbf{u}, \mathbf{v} \in [0,1]^d$ such that $\mathbf{u} \preceq \mathbf{v}$ then $g(\mathbf{u}) \leq g(\mathbf{v})$
  - $\mathbf{X}$ is uniformly distributed on $[0,1]^d$

- An adaptive method is provided to estimate a quantile: estimate q by

$$\hat{q} = \inf\{t \in \boldsymbol{R} : \hat{F}(t) > p\}.$$

with $\hat{F}$ an estimator of F.

Materials

Definition

Let $A \subset [0,1]^d$. Define

$$\mathbb{U}^-(A) = \bigcup_{\mathbf{x} \in A} \{\mathbf{u} \in [0,1]^d : \ \mathbf{u} \preceq \mathbf{x}\},$$
$$\mathbb{U}^+(A) = \bigcup_{\mathbf{x} \in A} \{\mathbf{u} \in [0,1]^d : \ \mathbf{u} \succeq \mathbf{x}\}.$$

Definition

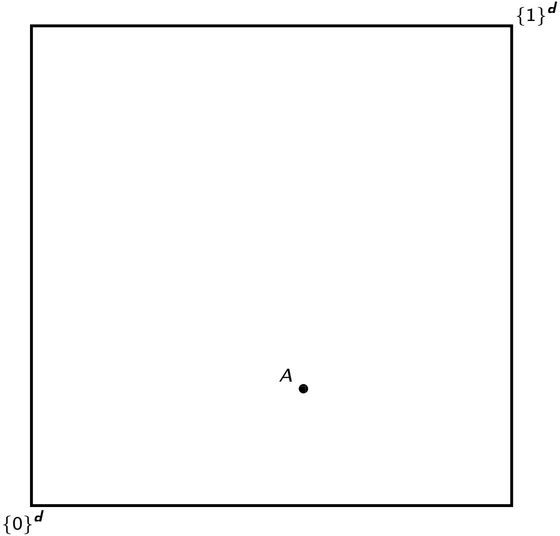
Let $\alpha \in]0,1[$ assume that S is a $(d-1)$-dimensional surface such that

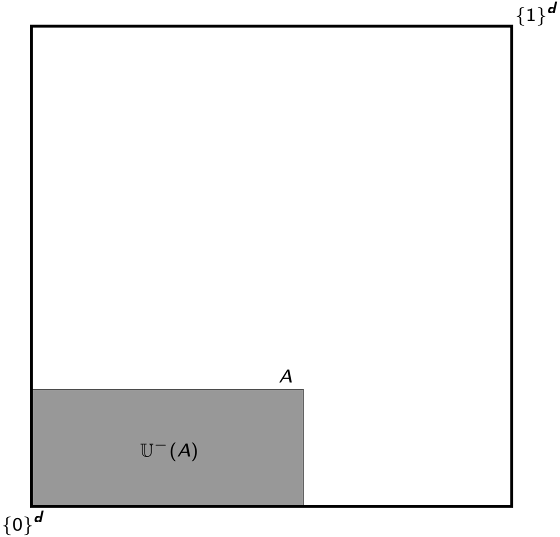
(i) for all $\mathbf{u}, \mathbf{v} \in S$, $\mathbf{u}$ is not strictly dominated by $\mathbf{v}$

(ii) $\mu(\mathbb{U}^-(S)) = \alpha$,

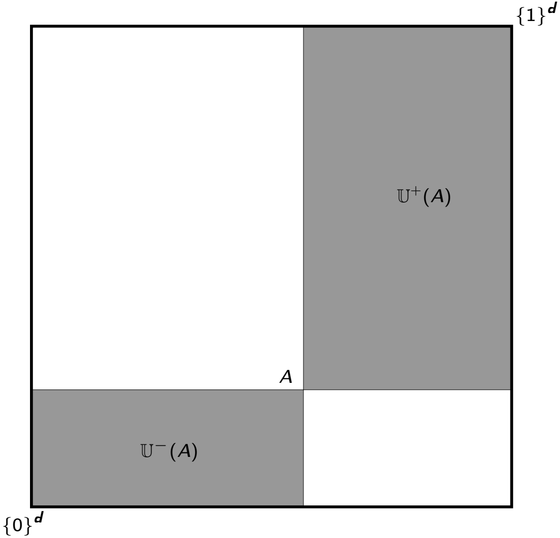
then S is said to be α-monotonic.

Remark: the set $\{\mathbf{x} \in [0,1]^d, g(x) = q\}$ is $F(q)$-monotonic.


$\{1\}^d$
A
$\{0\}^d$



$\{1\}^d$
A
$\mathbb{U}^-(A)$
$\{0\}^d$



$\{1\}^d$
$\mathbb{U}^+(A)$
A
$\mathbb{U}^-(A)$
$\{0\}^d$

Main result

Proposition

Let S_p be a p-monotonic surface. Then

$$\min_{\mathbf{x}\in S_p} g(\mathbf{x}) \leq q \leq \max_{\mathbf{x}\in S_p} g(\mathbf{x}).$$

- A p-monotonic surface is difficult to build in practice. The condition can be relaxed:

Main result

Proposition

Let S_p be a p-monotonic surface. Then

$$\min_{\mathbf{x} \in S_p} g(\mathbf{x}) \leq q \leq \max_{\mathbf{x} \in S_p} g(\mathbf{x}).$$

- A p-monotonic surface is difficult to build in practice. The condition can be relaxed:

Proposition

Let $p^- \in [0, p]$ and $p^+ \in [p, 1]$, and let S_{p^-}, S_{p^+} be respectively a (p^-)-monotonic surface and a (p^+)-monotonic surface. Then

$$\min_{\mathbf{x} \in S_{p^-}} g(\mathbf{x}) \leq q \leq \max_{\mathbf{x} \in S_{p^+}} g(\mathbf{x}).$$

Initialisation

- Let $\mathbf{x} = (x_1, \ldots, x_d) \in [0,1]^d$. Since the frontier of $\mathbb{U}^-(\mathbf{x})$ is monotonic

$$\mu(\mathbb{U}^-(\mathbf{x})) = x_1 \cdots x_d \geq p \Rightarrow q \leq g(\mathbf{x})$$

Proposition

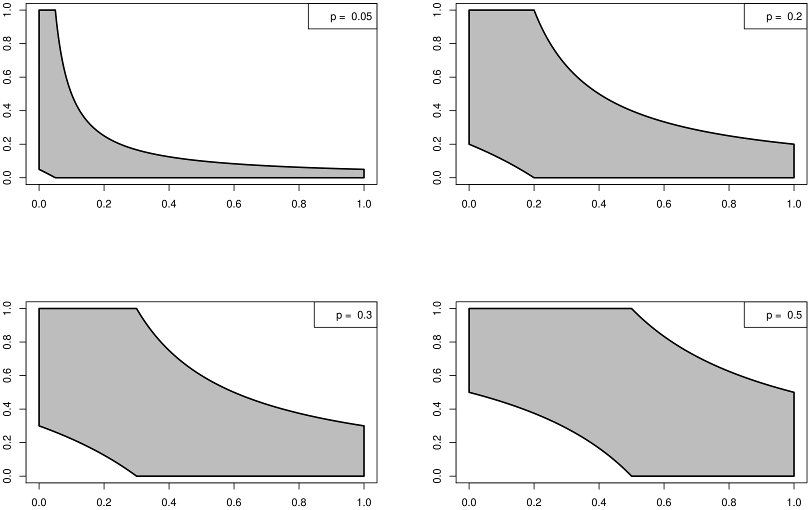
Let

$$\mathbb{W}^-(p) = \left\{ \mathbf{x} = (x_1, \ldots, x_d) \in [0,1]^d : \prod_{i=1}^{d} (1 - x_i) \geq 1 - p \right\}$$

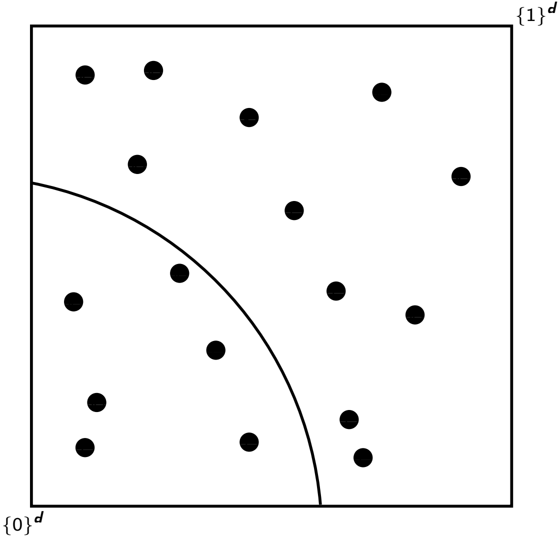
$$\mathbb{W}^+(p) = \left\{ \mathbf{x} = (x_1, \ldots, x_d) \in [0,1]^d : \prod_{i=1}^{d} x_i \geq p \right\}.$$

Then for all $(\mathbf{u}, \mathbf{v}) \in \mathbb{W}^-(p) \times \mathbb{W}^+(p),\ g(\mathbf{u}) \leq q \leq g(\mathbf{v})$.

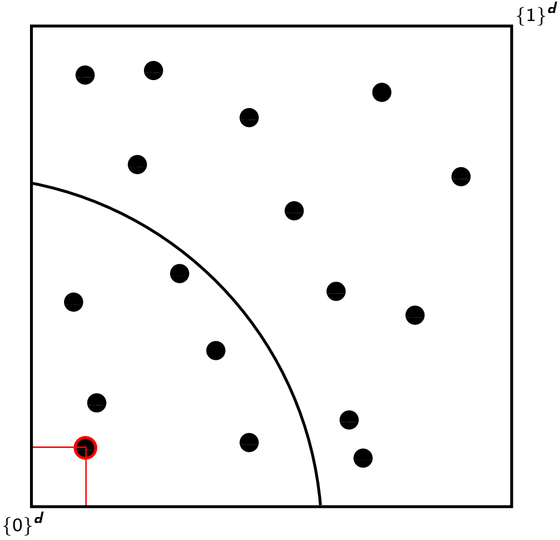
- In gray: $\mathbb{W}(p) = [0,1]^d \backslash (\mathbb{W}^-(p) \cup \mathbb{W}^+(p))$



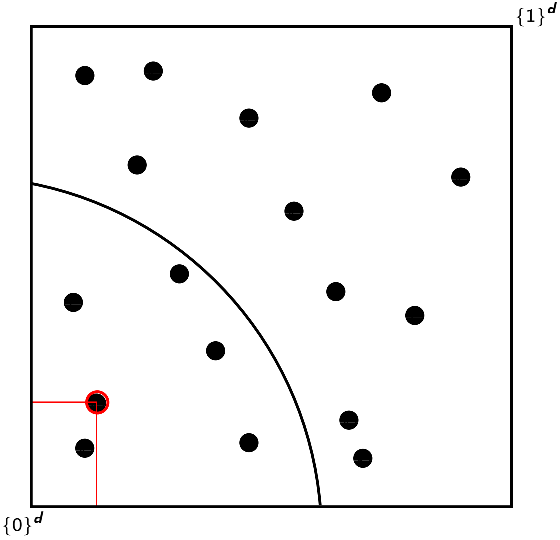
Bounding a quantile from a sample



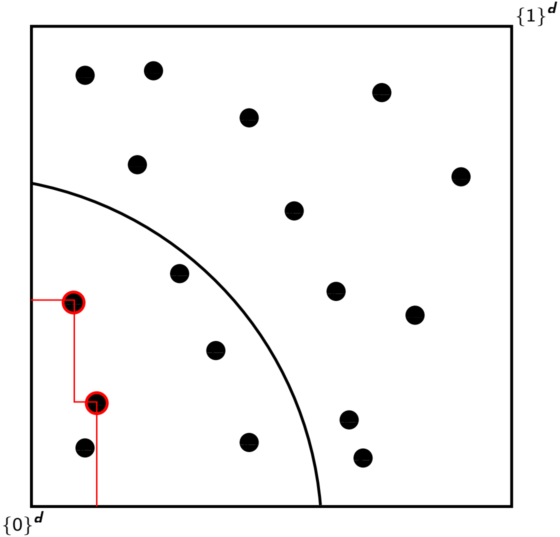
Bounding a quantile from a sample



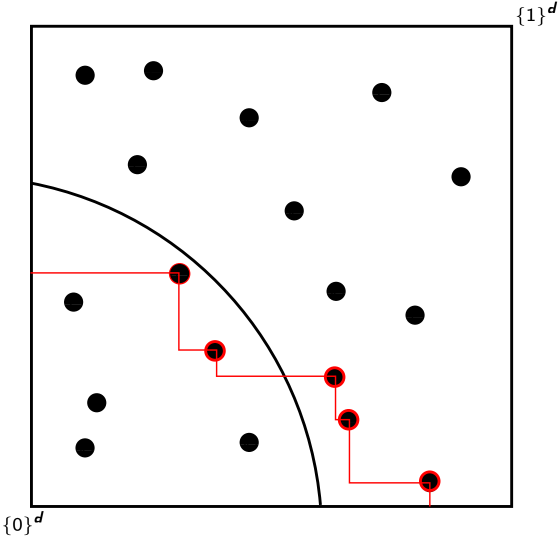
Bounding a quantile from a sample



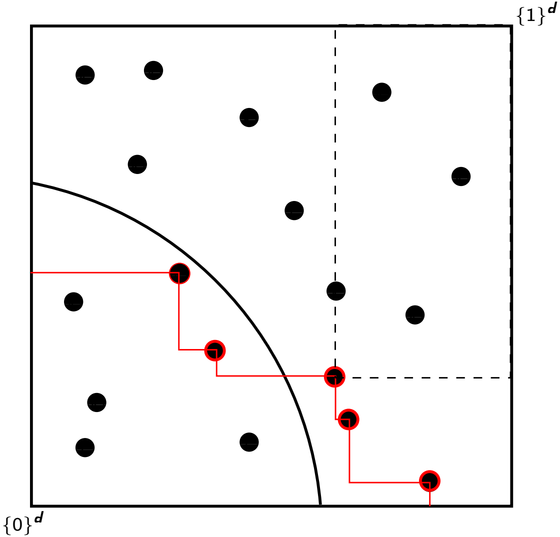
Bounding a quantile from a sample



Bounding a quantile from a sample



Bounding a quantile from a sample



Probability estimation

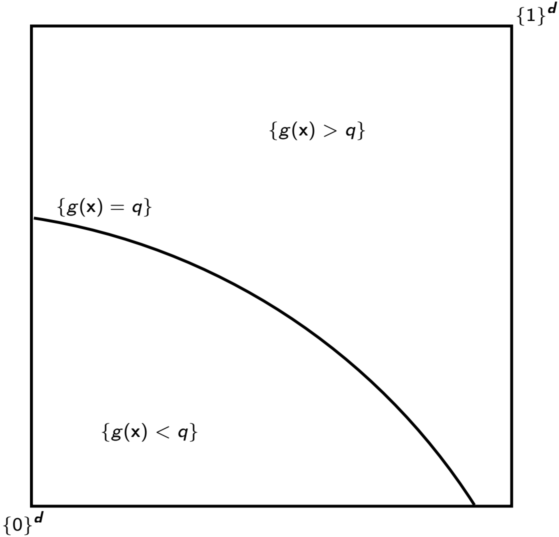
- Aim: estimate q by

$$\hat{q} = \inf\{t \in \boldsymbol{R} : \ \hat{F}(t) > p\}$$

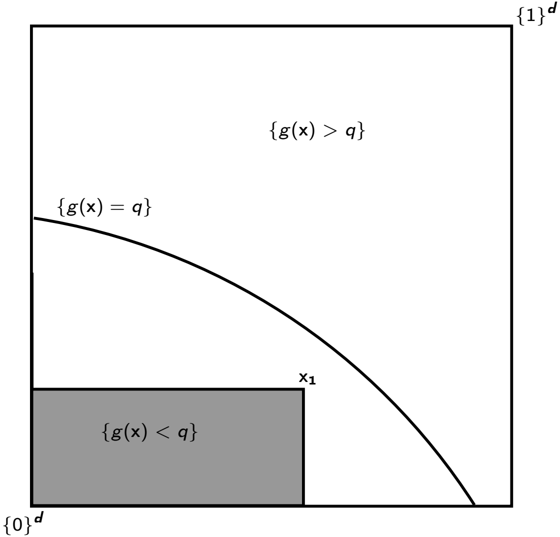
with $\hat{F}$ an estimator of F

- Taking account of the deterministic bounds, an unbiased estimator $\hat{F}$ is inspired by Bousquet (2012)

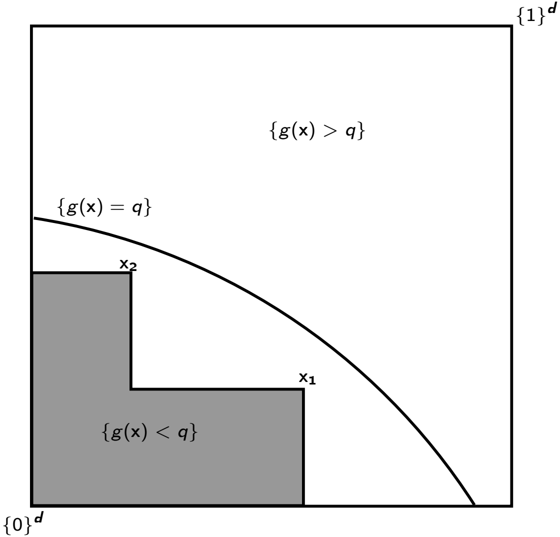
Probability estimation



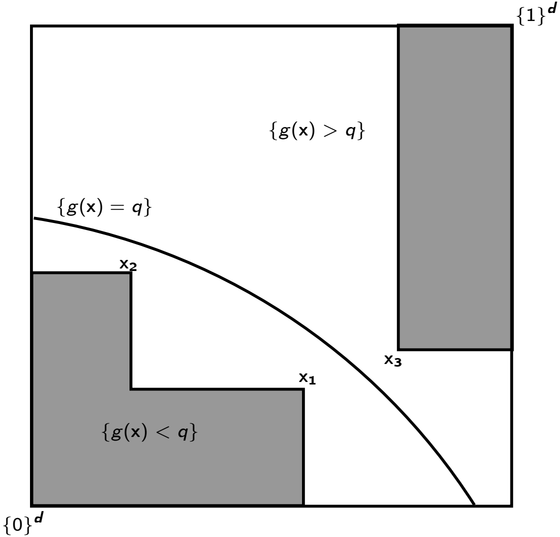
Probability estimation



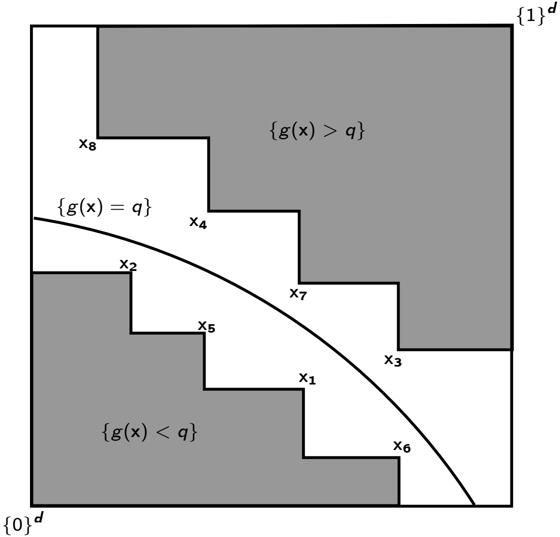
Probability estimation



Probability estimation



Probability estimation



Probability estimation

- Denote $\mathbb{U}_{k-1}$ the set where the sign of $g(.) - q$ is unknown
- Let $\mathbf{X}_k$ be uniformly distributed on $\mathbb{U}_{k-1}$, then

$$p_{k-1}^{-} + (p_{k-1}^{+} - p_{k-1}^{-})\mathbb{1}_{\{g(\mathbf{X}_k)\leq q\}}$$

is an unbiased estimator of $F(q)$.

Indeed, $\mathbb{1}_{\{g(\mathbf{X}_k)\leq q\}} \sim \mathcal{B}ernoulli\left(\frac{F(q)-p_{k-1}^{-}}{p_{k-1}^{+}-p_{k-1}^{-}}\right)$

Probability estimation

- Denote $\mathbb{U}_{k-1}$ the set where the sign of $g(.) - q$ is unknown
- Let $\mathbf{X}_k$ be uniformly distributed on $\mathbb{U}_{k-1}$, then

$$p_{k-1}^{-} + (p_{k-1}^{+} - p_{k-1}^{-})\mathbb{1}_{\{g(\mathbf{X}_k)\leq q\}}$$

is an unbiased estimator of $F(q)$.

Indeed, $\mathbb{1}_{\{g(\mathbf{X}_k)\leq q\}} \sim \mathcal{B}ernoulli\left(\frac{F(q)-p_{k-1}^{-}}{p_{k-1}^{+}-p_{k-1}^{-}}\right)$

- Let $(\mathbf{X}_k)_{k\geq 1}$ be a sequence of random vector such that for all $k \geq 1, \mathbf{X}_k$ is uniformly distributed on $\mathbb{U}_{k-1}$, then

$$F_n(q) = \frac{1}{n}\sum_{k=1}^{n} p_{k-1}^{-} + (p_{k-1}^{+} - p_{k-1}^{-})\mathbb{1}_{\{g(\mathbf{X}_k)\leq q\}}$$

is also an unbiased estimator of $F(q)$.

Application to quantile estimation

- At step 1, denote :
  - $\mathbb{U}_0 = \mathbb{W}(p)$
  - $g(\mathbf{x}^-) = \mathbf{q_0^-} \leq \mathbf{q} \leq \mathbf{q_0^+} = g(\mathbf{x}^+)$, with

$$\mathbf{x}^- = (1-(1-p)^{1/d}, \ldots, 1-(1-p)^{1/d}) \in \mathbb{W}^-(p) \subset [0,1]^d,$$
$$\mathbf{x}^+ = (p^{1/d}, \ldots, p^{1/d}) \in \mathbb{W}^+(p) \subset [0,1]^d,$$

Remark: without more information on g, $\mathbf{x}^-, \mathbf{x}^+$ are chosen arbitrary.

  - $p_0^- = \mu(\mathbb{W}^-(p))$, $p_0^+ = 1 - \mu(\mathbb{W}^+(p))$.

Application to quantile estimation

- At step 1, denote :
  - $\mathbb{U}_0 = \mathbb{W}(p)$
  - $g(\mathbf{x}^-) = \mathbf{q_0^-} \leq \mathbf{q} \leq \mathbf{q_0^+} = g(\mathbf{x}^+)$, with

$$\mathbf{x}^- = (1-(1-p)^{1/d}, \ldots, 1-(1-p)^{1/d}) \in \mathbb{W}^-(p) \subset [0,1]^d,$$
$$\mathbf{x}^+ = (p^{1/d}, \ldots, p^{1/d}) \in \mathbb{W}^+(p) \subset [0,1]^d,$$

Remark: without more information on g, $\mathbf{x}^-, \mathbf{x}^+$ are chosen arbitrary.

  - $p_0^- = \mu(\mathbb{W}^-(p))$, $p_0^+ = 1 - \mu(\mathbb{W}^+(p))$.
- Let $\mathbf{X}_1$ be uniformly distributed on $\mathbb{U}_0$, then

$$\hat{F}_1(q) = p_0^- + (p_0^+ - p_0^-)\mathbb{1}_{\{g(\mathbf{X}_0) \leq q\}}$$

is an unbiased estimator of $F(q)$.

- q is estimated by

$$\hat{q}_1 = \inf\{t \in [q_0^-, q_0^+] : \ \hat{F}_1(t) > p\}$$

Application to quantile estimation

- At step n, from $\mathbf{X}_1, \dots, \mathbf{X}_{n-1}$:
  - two bounds $q^-_{n-1} \leq q \leq q^+_{n-1}$ has been obtained
  - the non-dominated set $\mathbb{U}_{n-1}$ has been updated
- Let $\mathbf{X}_n$ be uniformly distributed on $\mathbb{U}_{n-1}$, then

$$\hat{F}_n(q) = \frac{1}{n} \sum_{k=1}^{n} p^-_{k-1} + (p^+_{k-1} - p^-_{k-1}) \mathbb{1}_{\{g(\mathbf{X}_k) \leq q\}}$$

is an unbiased estimator of $F(q)$.

- q is estimated by

$$\hat{q}_n = \inf\{t \in [q^-_{n-1}, q^+_{n-1}] : \; \hat{F}_n(t) > p\}$$

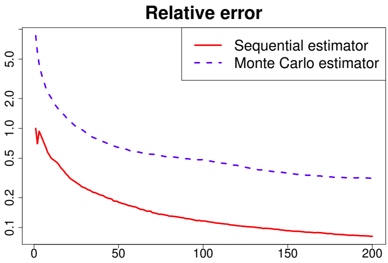
Numerical applications

- Consider an analytical example.
- Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector where $X_i \sim \Gamma(i+1, 1)$. Denoting

$$g(\mathbf{X}) = X_1 / \sum_{i=1}^{d} X_i \sim \mathcal{B}eta(2, (d+1)(d+2)/2 - 3),$$

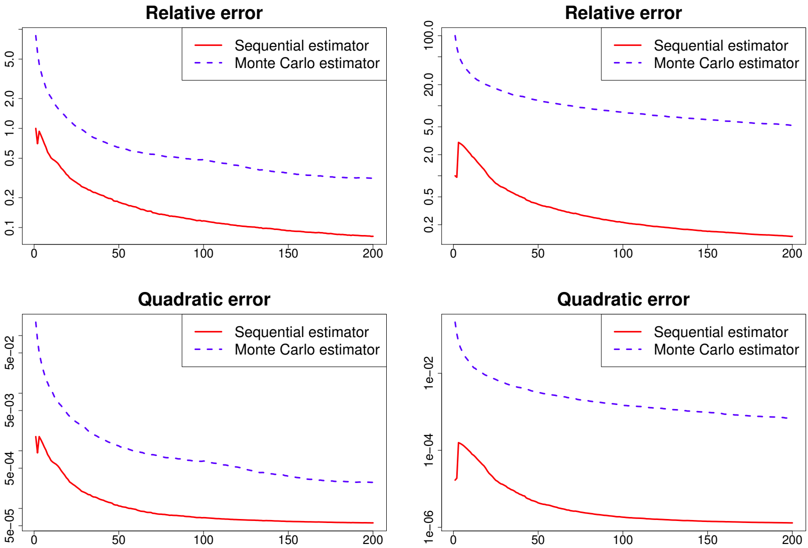
and $q_{d,p}$ the p-quantile of $g(\mathbf{X})$.
- The method is compared with a standard Monte Carlo Method for different couple (d, p) with $n = 200$ evaluations available of g
- The comparison is conducted on four different criterion:
  - the quadratic error $\mathbb{E}[(q_n - q_{d,p})^2]$
  - the relative error $\mathbb{E}[|q_n - q_{d,p}]/q_{d,p}$
  - the bias $\mathbb{E}[q_n - q_{d,p}]$
  - the length of the confidence interval at 95%

$d = 2$



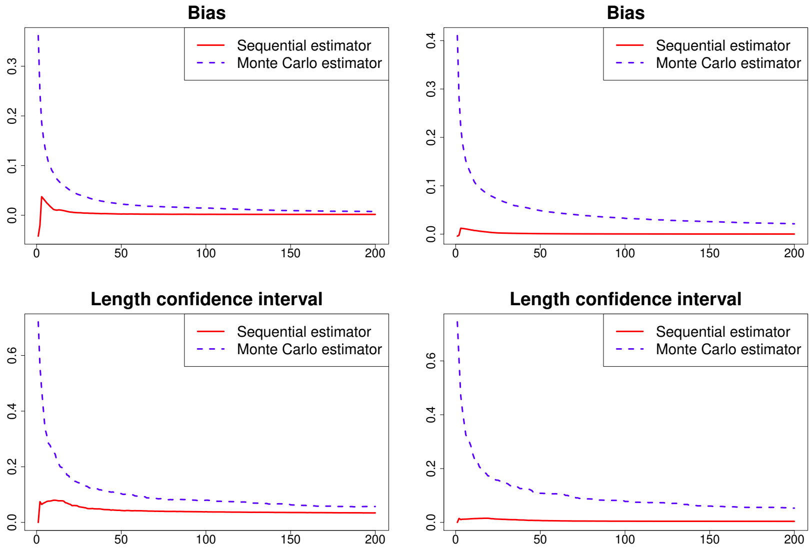
Relative error
Sequential estimator
Monte Carlo estimator
5.0
2.0
1.0
0.5
0.2
0.1
0
50
100
150
200
Relative error
Sequential estimator
Monte Carlo estimator
100.0
20.0
5.0
2.0
0.5
0.2
0
50
100
150
200

$d = 2$



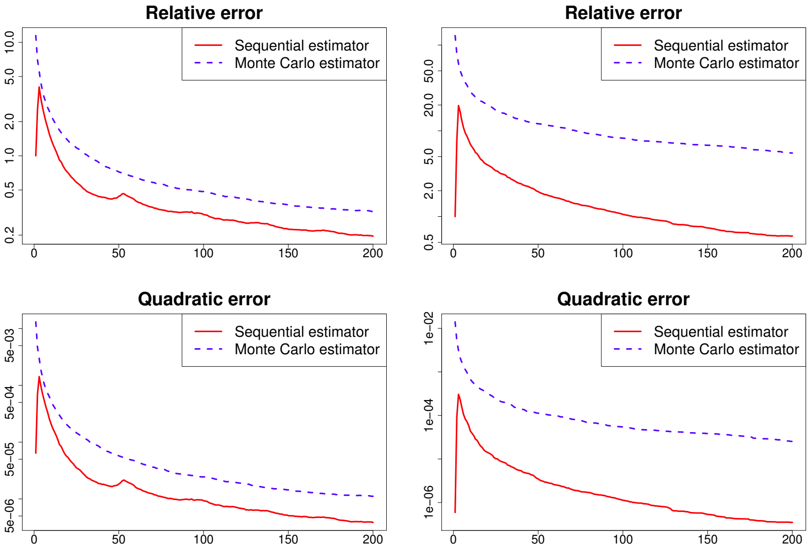
Left: $p = 10^{-2}$. **Right:** $p = 10^{-4}$

$d = 2$



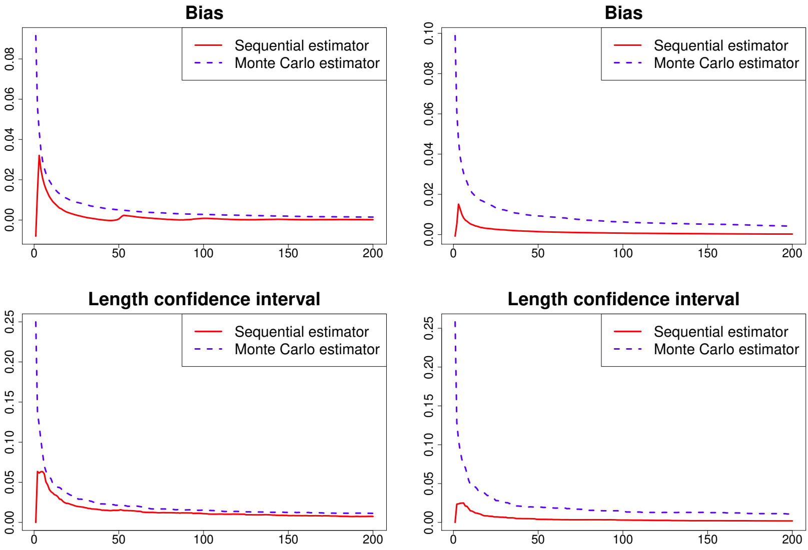
Left: $p = 10^{-2}$. **Right:** $p = 10^{-4}$

$d = 5$



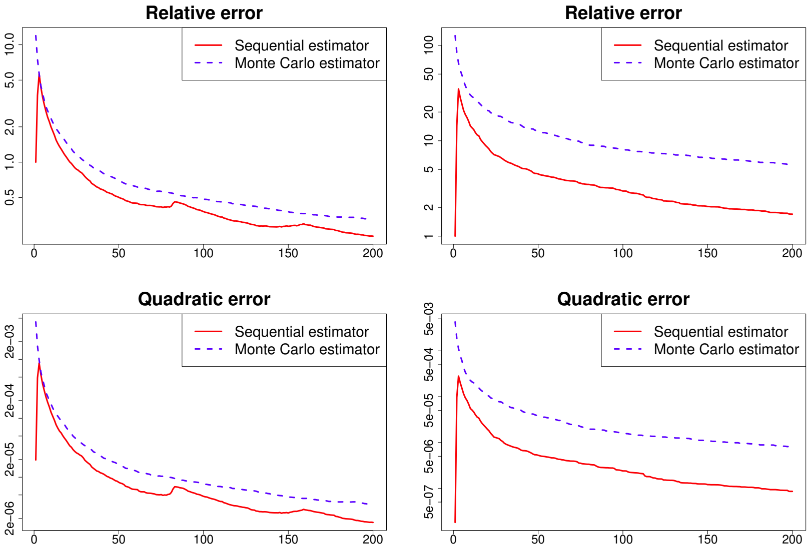
Left: $p = 10^{-2}$. **Right:** $p = 10^{-4}$

$d = 5$



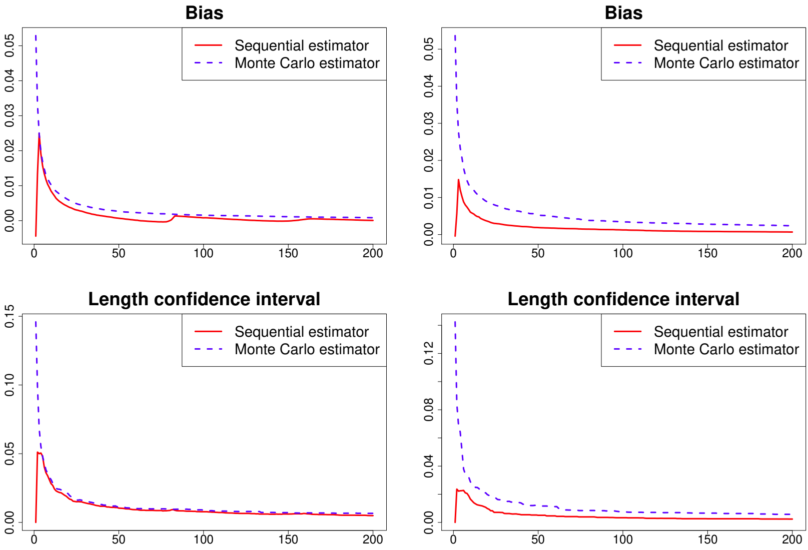
Left: $p = 10^{-2}$. **Right:** $p = 10^{-4}$

$d = 7$

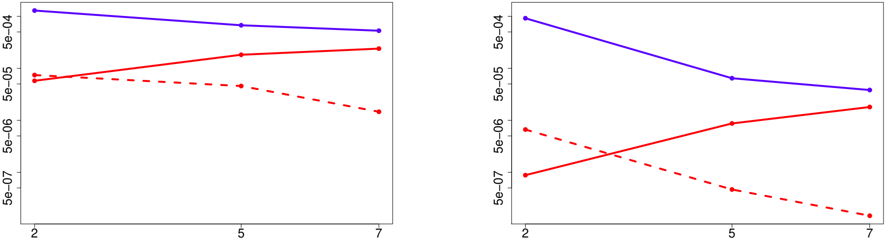


Left: $p = 10^{-2}$. **Right:** $p = 10^{-4}$

$d = 7$



Left: $p = 10^{-2}$. **Right:** $p = 10^{-4}$



Left: $p = 10^{-2}$. **Right:** $p = 10^{-4}$

Conclusion

- A central limit theorem can be obtained to control the estimator ?

- Instead a uniform sequential sampling, use a sequential importance sampling to accelerate the estimation of $F(q)$ then to apply on quantile estimation

- Comparing this method with existing method (Guyader, Morio...)