Some Results on Discrepancy and Function Approximation

Zongmin WU

Fudan University, Oct. 2010

Outline

Outline

Uniform Data Sites

UNIFORMLY DISTRIBUTED DATA

Uniformly distributed data on interval
$x_j = a + \frac{j}{n}(b - a)$ on $[a, b]$.

Uniformly distributed data on arc
$x_j = e^{ij\theta/n}$ on $[0, \theta]$.
and on circle.

with scale rotate and shift to any interval and arc.

APPROXIMATION BASED ON UNIFORM DATA

Bernstein' Polynomial Approximation

$$f_n^*(x) = \sum f(\frac{j}{n}) B_j^n(x) \to f(x)$$

where $B_j^n(x) = C_n^j x^j (1-x)^{n-j}$

Schoenberg's Model

$$\sum f(jh)\phi(\frac{x}{h} - j) \to f(x)$$

If $\phi(x)$ satisfies the Strang -Fix condition.

Wavelet representation

$$\sum \lambda_j^n \psi(2^n x - j)$$

where λ_j^n are uniformly distributed functional, one kind of sampling data.

However, there are almost no uniformly distributed point on the surface of the ball, in the ball as well as in the disc.
Therefor we can only get a quasi uniformly (almost uniformly) distributed data points.
1. For given constant $c > 1$, define the density (fill distance) $h = \max_{x\in\Omega} \min_j\{\|x - x_j\|\}$, if the number of points $\{x_j\}$ in the $\{x|\|x - \bar{x}\| < h\}$ are all less than c.
2. Very close data point may happen. $\rightarrow$ For given $c < 1$, $\min_{j\neq k} \|x_j - x_k\| > ch$.
3. Difficult to describe a measure of the scattering compare with the uniforming, $\rightarrow$ discrepancy.
4. Uniformly distributed point of probability, the point falls in to the subset S is S/Ω.

Outline

Quasi uniform Data Sites

Definition

A strictly increasing sequence of points $\{x_j^n\}_{j=0}^n$ in unit interval $[0, 1]$ are called uniformly distributed, provided that $x_j^n = \frac{j}{n}$.

Theorem

Let $B_j^n(x) = \frac{n!}{j!(n-j)!}x^j(1-x)^{n-j}$ be Bernstein Polynomials, then for any continuous function $f \in C([0, 1])$, the quasi-interpolation $B_n f(x) = \sum_{j=0}^{n} f(\frac{j}{n})B_j^n(x)$ converges to $f(x)$ uniformly on $[0, 1]$, and the error can be bounded by

$$\|B_n f(x) - f(x)\|_\infty \le \frac{3}{2}\omega(\frac{1}{\sqrt{n}})$$

Quasi uniform Data Sites

Definition

A strictly increasing sequence of points $\{x_j^n\}_{j=0}^n$ in unit interval $[0,1]$ are called quasi-uniformly distributed with discrepancy $\frac{1}{n^\beta}$, provided that

$$D_n = D_n(x_j^n) = \sup_{I \subseteq [0,1]} \left| \frac{\sharp(I)}{n} - |I| \right| \leq \mathcal{O}(\frac{1}{n^\beta})$$

where $\sharp(I)$ is the number of points $\{x_j^n\}_{j=0}^n \in I$, i.e.

$$\sharp(I) = \sum_{j=1}^{n} \chi_I(\{x_n\}),$$

Quasi uniform Data Sites

Proposition

The following two statements are equivalent.

(1) *The discrepancy D_n satisfies*

$$D_n \leq \mathcal{O}(\frac{1}{n^\beta}).$$

(2) *The following inequality*

$$|x_j^n - \frac{j}{n}| \leq \mathcal{O}(\frac{1}{n^\beta})$$

holds true for each j.

Quasi uniform Data Sites

Proof.

(1) $\to$ (2): Let $I_{x_j^n}$ denote the closed interval $[0, x_j^n]$. From (1) one can get

$$\left| |I_{x_j^n}| - \frac{\sharp(I_{x_j^n})}{n} \right| = \left| |I_{x_j^n}| - \frac{j}{n} \right| \leq D_n \leq C \cdot (\frac{1}{n^\beta}).$$

Note the fact that $|I_{x_j^n}| = x_j^n$, and it completes the proof.

(2) $\to$ (1): The proof is trivial. □

Quasi uniform Data Sites

Theorem

Let $B_j^n(x) = \frac{n!}{j!(n-j)!} x^j (1-x)^{n-j}$ *be Bernstein Polynomials, then for any continuous function* $f \in C([0,1])$, *the generalized quasi-interpolation*

$$B_n^* f(x) = \sum_{j=0}^{n} f(x_j^n) B_j^n(x)$$

converges to $f(x)$ *uniformly on* $[0,1]$, *and the error can be bounded by*

$$\|B_n^* f(x) - f(x)\|_\infty \leq (1 + \frac{1}{2}\sqrt{1 + \frac{4}{n} + \frac{4}{\sqrt{n}}})\omega(\frac{1}{\sqrt{n}}),$$

Quasi uniform Data Sites

Theorem

Let $B_j^n(x) = \frac{n!}{j!(n-j)!} x^j (1-x)^{n-j}$ *be Bernstein Polynomials, then for any continuous function* $f \in C([0,1])$*, the generalized quasi-interpolation*

$$B_n^* f(x) = \sum_{j=0}^{n} f(x_j^n) B_j^n(x)$$

converges to $f(x)$ *uniformly on* $[0,1]$*, and the error can be bounded by*

$$\|B_n^* f(x) - f(x)\|_\infty \le \omega(\frac{1}{n^\beta}) + \frac{3}{2}\omega(\frac{1}{\sqrt{n}}) \le 2 \cdot \max\{\omega(\frac{1}{n^\beta}), \frac{3}{2}\omega(\frac{1}{\sqrt{n}})\},$$

Quasi uniform Data Sites

Proof.

$$
\begin{aligned}
& |B_n^*(f,x) - f(x)| \\
\leq \quad & |\sum_{j=0}^{n} f(x_j^n) B_j^n(x) - f(x)| \\
\leq \quad & |\sum_{j=0}^{n} f(x_j^n) B_j^n(x) - f(\frac{j}{n}) B_j^n(x) + f(\frac{j}{n}) B_j^n(x) - f(x)| \\
\leq \quad & |\sum_{j=0}^{n} f(x_j^n) B_j^n(x) - f(\frac{j}{n}) B_j^n(x)| + |f(\frac{j}{n}) B_j^n(x) - f(x)| \\
\leq \quad & \omega(\frac{1}{n^\beta}) + \frac{3}{2}\omega(\frac{1}{\sqrt{n}}) \leq 2 \cdot \max\{\omega(\frac{1}{n^\beta}), \frac{3}{2}\omega(\frac{1}{\sqrt{n}})\}
\end{aligned}
$$

Quasi uniform Data Sites

Definition

An increasing sequence of points $\{x_j^n\}_{j=0}^n$ in unit interval $[0,1]$ are called uniformly distributed in probability, provided that $n+1$ points $\{x_j^n\}_{j=0}^n$ is an ordered sequence of uniform random numbers adopted from $[0,1]$.

Remark

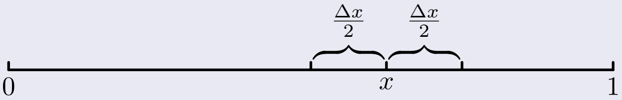
The sequence that uniformly distributed in probability must not be a quasi-uniformly distributed sequence in interval $[0,1]$. *Its discrepancy will not tend to zero as* $n \to \infty$. *The quasi-uniformly distributed sequence* $\{x_j^n\}$ *means a disturb to the sequence* $\{\frac{j}{n}\}$ *that the disturb dominated by the discrepancy.*

Quasi uniform Data Sites

Theorem

The above point x_j^n obey $(n+1)\cdot B_j^n(x)$ distribution.

Proof.



As shown in the figure, the probability that the point x_j^n falls into the interval $ball(x, \frac{\Delta x}{2})$ is $(n+1)\Delta x$, the probability that the j points $\{x_k^n\}_{k=0}^{j-1}$ fall into the interval $[0, x - \frac{\Delta x}{2})$ is $C_j^n(x - \frac{\Delta x}{2})^j$, the probability that the $n-j$ points $\{x_k^n\}_{k=j+1}^{n}$ fall into the interval $(x + \frac{\Delta x}{2}, 1]$ is $(1 - x - \frac{\Delta x}{2})^{n-j}$. Therefore, the probability density function is $(n+1)\cdot B_j^n(x)$ as Δx tends zero.

Quasi uniform Data Sites

Theorem

Assume $\frac{5}{2}\omega(\frac{1}{\sqrt{n}}) < \epsilon/2$. *Let* $B_j^n(x) = \frac{n!}{j!(n-j)!}x^j(1-x)^{n-j}$ *be Bernstein Polynomials, then for any continuous function* $f \in C([0,1])$, *the generalized quasi-interpolation that the data sampled on uniform random numbers*

$$B_n^{**}f(x) = \sum_{j=0}^{n} f(x_j^n)B_j^n(x)$$

converges to $f(x)$ *in probability on* $[0,1]$, *and the error can be bounded by*

$$P\{x||B_n^{**}f(x) - f(x)| > \epsilon\} \leq 16/\epsilon^4 n^3\omega(1/\sqrt{n})^4 \leq 16/\epsilon^4 n.$$

Quasi uniform Data Sites

Properties

The Expectation $E(x_j^n)$ *of* x_j^n *is* $\frac{j+1}{n+2}$ *and the Variance* $D(x_j^n)$ *is* $\frac{(j+1)(j+2)}{(n+2)(n+3)} - \frac{(j+1)^2}{(n+2)^2}$. *Let difference*$= E(x_j^n) - \frac{j}{n}$. *Therefore,*

$$\begin{aligned} & \int_0^1 (x - \frac{j}{n})^2 (n+1) B_j^n(x) dx \\ = \; & D(x_j^n) + \textit{difference}^2 \\ = \; & \frac{(j+1)(n-j+1)}{(n+2)^2(n+3)} + \frac{(n-2j)^2}{n^2(n+2)^2} \\ \leq \; & \frac{1}{4(n+3)} + \frac{1}{(n+2)^2} \leq \frac{1}{2(n+3)}. \end{aligned}$$

Quasi uniform Data Sites

Properties

Moreover

$$\begin{aligned}
&\int_0^1 (x-\frac{j}{n})^4 (n+1) B_j^n(x)dx \\
= &\frac{120j^4 + 24n^4 - 240j^3 n + 240j^2 n^2 - 120jn^3}{n^4(n+2)(n+3)(n+4)(n+5)} \\
&+\frac{26n^4 j + 112j^2 n^3 + 172j^3 n^2 - 86j^4 * n}{n^4(n+2)(n+3)(n+4)(n+5)} \\
&+\frac{3n^4 j^2 - 6j^3 n^3 + 3j^4 n^2}{n^4(n+2)(n+3)(n+4)(n+5)} \\
\leq &\frac{1}{(n+3)(n+5)} \leq \frac{1}{(n+3)^2}
\end{aligned}$$

Quasi uniform Data Sites

Proof.

$$
\begin{aligned}
& P\{\|\sum_{j=0}^{n} f(x_j^n)B_j^n(x) - f(x)\|_\infty > \epsilon\} \\
= \quad & P\{\|\sum_{j=0}^{n} f(x_j^n)B_j^n(x) - \sum_{j=0}^{n} f(\frac{j}{n})B_j^n(x) + \sum_{j-0}^{n} f(\frac{j}{n})B_j^n(x) - f(x)\|_\infty > \epsilon\} \\
\leq \quad & P\{\|\sum_{j=0}^{n} f(x_j^n)B_j^n(x) - \sum_{j=0}^{n} f(\frac{j}{n})B_j^n(x)\|_\infty + \frac{3}{2}\omega(\frac{1}{\sqrt{n}}) > \epsilon\} \\
\leq \quad & P\{\|\sum_{j=0}^{n} |f(x_j^n) - f(\frac{j}{n})|B_j^n(x)\|_\infty + \frac{3}{2}\omega(\frac{1}{\sqrt{n}}) > \epsilon\} \\
\leq \quad & P\{\|\sum_{j=0}^{n} \omega(|x_j^n - \frac{j}{n}|)B_j^n(x)\|_\infty + \frac{3}{2}\omega(\frac{1}{\sqrt{n}}) > \epsilon\}
\end{aligned}
$$

Quasi uniform Data Sites

Proof.

Noting the fact that $\omega(\lambda\delta) \leq (1+\lambda)\omega(\delta)$, we have

$$\begin{aligned}
& P\{\|\sum_{j=0}^{n} f(x_j^n)B_j^n(x) - f(x)\|_\infty > \epsilon\} \\
\leq\ & P\{\|\sum_{j=0}^{n}(1 + \frac{|x_j^n - \frac{j}{n}|}{\frac{1}{\sqrt{n}}})\omega(\frac{1}{\sqrt{n}})B_j^n(x)\|_\infty + \frac{3}{2}\omega(\frac{1}{\sqrt{n}}) > \epsilon\} \\
\leq\ & P\{\|\sum_{j=0}^{n} \frac{|x_j^n - \frac{j}{n}|}{\frac{1}{\sqrt{n}}}\omega(\frac{1}{\sqrt{n}})B_j^n(x)\|_\infty + \frac{5}{2}\omega(\frac{1}{\sqrt{n}}) > \epsilon\} \\
\leq\ & P\{\max_j\{\frac{|x_j^n - \frac{j}{n}|}{\frac{1}{\sqrt{n}}}\omega(\frac{1}{\sqrt{n}}) + \frac{5}{2}\omega(\frac{1}{\sqrt{n}})\} > \epsilon\} \\
\leq\ & \sum_{j=0}^{n} P\{(\frac{|x_j^n - \frac{j}{n}|}{\frac{1}{\sqrt{n}}} + \frac{5}{2})\omega(\frac{1}{\sqrt{n}}) > \epsilon\}
\end{aligned}$$

Quasi uniform Data Sites

Noting the fact that

$$\begin{aligned} & \int_{|x-\frac{j}{n}|>A} (n+1)B_j^n(x)dx \\ \leq \quad & \int_{|x-\frac{j}{n}|>A} (n+1)B_j^n(x)(x-\frac{j}{n})^4/A^4 dx \\ \leq \quad & \int_0^1 (n+1)B_j^n(x)(x-\frac{j}{n})^4/A^4 dx \\ \leq \quad & 1/A^4(n+3)^2 \end{aligned}$$

Assume $\frac{5}{2}\omega(\frac{1}{\sqrt{n}}) < \epsilon/2$, then we have

$$\begin{aligned} & \sum_{j=0}^{n} P\{(\frac{|x_j^n - \frac{j}{n}|}{\frac{1}{\sqrt{n}}} + \frac{5}{2})\omega(\frac{1}{\sqrt{n}}) > \epsilon\} \\ \leq \quad & 16/n^3\omega(\frac{1}{\sqrt{n}})^4\epsilon^4 \to 0 \quad \text{as } n \to \infty. \end{aligned}$$

Quasi uniform Data Sites

If we take $(n+1)B_j^n(t)$ as a density function of some distribution, then $\lim_{n\to\infty}(n+1)B_j^n(t) \to \delta(t-t^*)$, if $\frac{j}{n} \to t^*$.
Therefor the Bernstein's polynomial can be taken as an approximation of discrete Riemann's summation of

$$f(x) = \int_0^1 f(t)\delta(x-t)dt$$

Shortcoming: Basis computation is not very simple.
An another proof of polynomial approximation is

$$\frac{c_n}{n+1}\sum f(j/n)(1-(x-j/n)^2)^n, \int_{-1}^1 c_n(1-x^2)^n dx = 1$$

which simulate is the Riemann's summation of $f * \delta$ too.
The concept of generator: Single function contain all the DNA of

Quasi uniform Data Sites

Through same discussion as for the Bernstein's polynomial, we can generalize the approach to be

Theorem

If $W(x) > 0$ compactly supported (fast decay), $\int W(x)dx = 1$ and $\int xW(x)dx = 0$, then $nW(nx)$ is delta series tend to $\delta(x)$. If we take $\sqrt{n}W(\sqrt{n}x - j/\sqrt{n})$ as a density function of distribution η_j^n, that the expectation $E\eta_j^n = \frac{j}{n}$, the variance $D\eta_j^n = \int (x - j/n)^2 \sqrt{n}W(\sqrt{n}x - j/\sqrt{n})dx = \int x^2 W(x)dx/n$ tend to zero too, then we have three Theorem

Quasi uniform Data Sites

Theorem

If $W(x) \in \mathcal{C}^1$, *for* $f(x) \in \mathcal{C}^1$,

$$
\begin{aligned}
& |\sum f(\frac{j}{n})\sqrt{n}W(\sqrt{n}x - j/\sqrt{n}) - f(x)| \\
< \; & |\sum f(\frac{j}{n})W(\sqrt{n}x - j/\sqrt{n})/\sqrt{n} - \int f(t)\sqrt{n}W(\sqrt{n}(x-t))dt| \\
& +|\int f(t)\sqrt{n}W(\sqrt{n}(x-t))dt - f(x)| \\
< \; & \mathcal{O}(1/\sqrt{n}) + \omega(1/\sqrt{n}) = \mathcal{O}(1/\sqrt{n})
\end{aligned}
$$

Quasi uniform Data Sites

Theorem

If x_j^n is a quasi uniform distributed points with the discrepancy $|x_j^n - j/n| < 1/n^\beta$, then

$$
\begin{aligned}
&|\sum f(x_j^n)\sqrt{n}W(\sqrt{n}x - j/\sqrt{n}) - f(x)| \\
&< \mathcal{O}(1/\sqrt{n}) + \omega(1/\sqrt{n}) + \omega(1/n^\beta)
\end{aligned}
$$

Quasi uniform Data Sites

Definition

A strictly increasing sequence of points $x_0, \cdots, x_n$ is called well-positioned if

$$\left| x_j^n - \frac{j}{n} \right| \leq \mathcal{O}(\frac{1}{n^\beta}),$$

where $\beta \geq \frac{1}{2}$.

Quasi uniform Data Sites

Proposition

Let $x_0, \cdots, x_n$ be the strictly increasing sequence of points and $f(x) \in C^1(R)$. If the discrepancy $D_n \leq \frac{1}{n^\beta}$ (where $\beta \geq \frac{1}{2}$), then

$$\|f(x) - M_n(x)\| \leq \mathcal{O}(\frac{1}{\sqrt{n}})$$

where $M_n(x) = \sum f(x_j^n)\sqrt{n}W(\sqrt{n}x - j/\sqrt{n})$.

Quasi uniform Data Sites

If x_j^n is taken from a uniform distribution of probability on $[0,1]$, then for n large enough (the error of the Theorem before) less than $\epsilon/4$ and $\omega(\frac{1}{\sqrt{n}}) \sum_{j=0}^{n} W(\sqrt{n}x - j/\sqrt{n})/\sqrt{n} < \epsilon/4$, since $\omega(\frac{1}{\sqrt{n}}) \sum_{j=0}^{n} W(\sqrt{n}x - j/\sqrt{n})/\sqrt{n}$ tend to 1), then we have the theorem

Quasi uniform Data Sites

Theorem

$$
\begin{aligned}
& P\left(\left|\sum f(x_j^n) W(\sqrt{n}x - j/\sqrt{n})/\sqrt{n} - f(x)\right| > \epsilon\right) \\
\leq \; & P\left(\left|\sum (f(x_j^n) - f(j/n)) W(\sqrt{n}x - j/\sqrt{n})/\sqrt{n}\right| > 3\epsilon/4\right) \\
\leq \; & P\left(\left|\sum \left(1 + \frac{|x_j^n - \frac{j}{n}|}{\frac{1}{\sqrt{n}}}\right)\omega\left(\frac{1}{\sqrt{n}}\right) W(\sqrt{n}x - j/\sqrt{n})/\sqrt{n}\right| > 3\epsilon/4\right) \\
\leq \; & P\left(\left|\sum_{j=0}^{n} \frac{|x_j^n - \frac{j}{n}|}{\frac{1}{\sqrt{n}}}\omega\left(\frac{1}{\sqrt{n}}\right) W(\sqrt{n}x - j/\sqrt{n})/\sqrt{n}\right| > \epsilon/2\right) \\
\leq \; & P\left\{\max_j\left(\frac{|x_j^n - \frac{j}{n}|}{\frac{1}{\sqrt{n}}}\right)\omega\left(\frac{1}{\sqrt{n}}\right) > \epsilon/2\right\}
\end{aligned}
$$

Quasi uniform Data Sites

With same reasons we have

$$\begin{aligned} & \sum_{j=0}^{n} P\{(\frac{|x_j^n - \frac{j}{n}|}{\frac{1}{\sqrt{n}}})\omega(\frac{1}{\sqrt{n}}) > \epsilon/2\} \\ \leq \quad & 16/n^3\omega(\frac{1}{\sqrt{n}})^4\epsilon^4 \to 0 \ \text{ as } n \to \infty. \end{aligned}$$

Quasi uniform Data Sites

Let $t_0^n = x_0^n - 0$, $t_j^n = x_j^n - x_{j-1}^n$ for $j = 1, \cdots, n$, $t_{n+1}^n = 1 - x_n^n$, then $x_j^n = \sum_{k=0}^{j} t_k^n$.
According to the distribution of x_j^n, one can easily get the density function of the joint probability distribution of x_{j-1}^n and x_j^n is $(n+1)nC_{n-1}^{j-1}X^{j-1}(1-Y)^{n-j}$ on $0 \leq X \leq Y \leq 1$, where X, Y denote x_{j-1}^n and x_j^n respectively. We denote t_j^n by $T = Y - X$, then the density function of T is $(n+1)(1-t)^n$, proof is in next page.

Quasi uniform Data Sites

$$\int\limits_{Y=X+T} (n+1)nC_{n-1}^{j-1}X^{j-1}(1-Y)^{n-j}dl$$

$$= \int\limits_0^{1-T} \frac{(n+1)!}{(j-1)!(n-j)!}X^{j-1}[(1-X)-(Y-X)]^{n-j}\sqrt{2}dX$$

$$= \int\limits_0^{1-T} \frac{(n+1)!}{(j-1)!(n-j)!}X^{j-1}[(1-X)-T]^{n-j}\sqrt{2}dX$$

$$= \int\limits_0^{1-T} \frac{(n+1)!}{(j-1)!(n-j)!}X^{j-1}\sum_{k=0}^{n-j} C_{n-j}^k(1-X)^k(-T)^{n-j-k}\sqrt{2}dX$$

Every $x_j^n - x_{j-1}^n$ Satisfy the same distribution as $1 - x_n^n$.

Quasi uniform Data Sites

Above are already give you as you in Fudan last time.

Outline

For disk and the ball, e.g. $d = 3$

Recall the kernel based quasi interpolation that the kernel ϕ is compact supported radial function on unit disk in R^d. At first for ball in R^d. There are no uniformly distributed knots in ball or on the surface of the ball, except with some special n. Assume that the points $\{x_j^n\}_{j=1}^n$ are quasi uniform distributed in ball or on the surface of the ball with discrepancy β. The quasi interpolation scheme, here $|B_d|$ is the volume of the d dimensional unit ball or the arer of the surface

$$\frac{|B_d|}{n} \sum n^{d\gamma} f(x_j^n) \phi(n^\gamma (x - x_j^n))$$

is a Riemannian summation of the the integral

$$\int n^{d\gamma} f(t) \phi(n^\gamma (x - t)) dt$$

For disk and the ball

Take a partition of the ball with the volume of $n^{-\alpha}$ and the radius of the sell is proportional to $n^{-\alpha/d}$, (First term caused by discrepancy, the second term caused by Riemann summation)

$$\begin{aligned} & \left|\frac{|B_d|}{n}\sum n^{d\gamma}f(x_j^n)\phi(n^\gamma(x-x_j^n)) - \int n^{d\gamma}f(t)\phi(n^\gamma(x-t))dt\right| \\ < \quad & \|f\|\|\phi\|\mathcal{O}(n^{d\gamma}\cdot n^{-\beta}) + (n^{d\gamma}\|f'\|\|\phi\| + n^{(d+1)\gamma}\|f\|\|\phi'\|)\mathcal{O}(n^{-\alpha/d}) \\ \leq \quad & \mathcal{O}(n^{d\gamma-\beta}) + \mathcal{O}(n^{(d+1)\gamma-\alpha/d}) < \mathcal{O}(n^{d\gamma-\beta}) \end{aligned}$$

If $f \in C^1$ and $\phi \in C^1$ too. $\alpha/d = \beta + \gamma$ is an optimal choosing of α.

For disk and the ball

On the other side, assume the kernel $\phi(x) = C(1 - \|x\|)_+^p$

$$\begin{aligned} & |f(x) - \int n^{d\gamma} f(t)\phi(n^\gamma(x-t))dt| \\ = \ & |\int (f(x) - f(t))n^{d\gamma}\phi(n^\gamma(x-t))dt| \\ < \ & \|f''\|\mathcal{O}(n^{-2\gamma}) \ \ \text{Taylor of order2 roughly} \end{aligned}$$

since $\int t\phi(t)dt = 0$.

comments

Summarize the two part of the Error estimates, we get

$$
\begin{aligned}
& \left|\frac{|B_d|}{n}\sum n^{d\gamma} f(x_j)\phi(n^{\gamma}(x-x_j)) - f(x)\right| \\
\leq \quad & \mathcal{O}(n^{d\gamma-\beta}) + \mathcal{O}(n^{-2\gamma}) \leq \mathcal{O}(n^{-2\beta/(2+d)})
\end{aligned}
$$

if we take the optimal choosing of $\gamma = \beta/(2+d)$. These means for smaller discrepancy (β is larger), we can get more precise approximation. The order of the approximation are almost inversely proportional to the dimension d.

For disk and the ball

With the same argument, the approach is valid for the surface of the ball too, if take the all integral above on the surface.

Outline

Data on the ball and uniformly distributed in probability

Lemma

Already proved in above. The k-th point of $n+1$ uniformly distributed on $[0,1]$ in probability satisfy the distribution of $(n+1)B_n^k(x)$, where $\{B_n^k\}_{k=0}^n$ are the Bernstein polynomial of degree n. We can give a name as Bernstein distribution.

Lemma

(Generalized Bernstein distribution) For any sub domain S with the volume $B_d n^{-\alpha}$, the probability that there are k points of $\{x_j\}$ fall in to the domain S is

$$C_n^k (n^{-\alpha})^k (1 - n^{-\alpha})^{n-k}$$

Take the discrepancy in to account, the probability that the discrepancy is large than $n^{-\beta}$ in the sub set S will be

$$P(|\sharp S/n - n^{-\alpha}| > n^{-\beta}) = \sum_{|\frac{k}{n} - n^{-\alpha}| > n^{-\beta}} C_n^k (n^{-\alpha})^k (1 - n^{-\alpha})^{n-k}$$

For any given ϵ, let $\epsilon = n^{-2\beta/(d+2)}$ to get a
$\beta = -2\log(\epsilon)/(d+2)\log(n)$
If $\|f^*(x) - f(x)\| \nleq \epsilon$, then at least in one of the sub domain S, the discrepancy will large than $n^{-\beta}$. Or
$P(\|f^{**}(x) - f(x)\|_\infty > \epsilon) < 1/n\epsilon^{d+2}$ for order 2, and
$P(\|f^{**}(x) - f(x)\|_\infty > \epsilon) < 1/n^2\epsilon^{2(d+2)}$ for order 4 estimation.

Details see next pages.

Help for calculus, we use $|\frac{k}{n} - n^{-\alpha}|^2$ and $|\frac{k}{n} - n^{-\alpha}|^4$

$\frac{k}{n} B_k^n = B_{k-1}^{n-1} x$

$\frac{k}{n} B_{k-1}^{n-1} x = \frac{k-1+1}{n} B_{k-1}^{n-1} x = \frac{n-1}{n} B_{k-2}^{n-2} x^2 + \frac{1}{n} B_{k-1}^{n-1} x$

$\frac{k}{n} [\frac{n-1}{n} B_{k-2}^{n-2} x^2 + \frac{1}{n} B_{k-1}^{n-1} x] =$

$\frac{(n-1)(n-2)}{n^2} B_{k-3}^{n-3} x^3 + \frac{2(n-1)}{n^2} B_{k-2}^{n-2} x^2 + \frac{n-1}{n^2} B_{k-2}^{n-2} x^2 + \frac{1}{n^2} B_{k-1}^{n-1} x$

$\frac{k}{n} [\frac{(n-1)(n-2)}{n^2} B_{k-3}^{n-3} x^3 + \frac{3(n-1)}{n^2} B_{k-2}^{n-2} x^2 + \frac{1}{n^2} B_{k-1}^{n-1} x] =$

$\frac{(n-1)(n-2)(n-3)}{n^3} B_{k-4}^{n-4} x^4 + \frac{6}{n} \frac{(n-1)(n-2)}{n^2} B_{k-3}^{n-3} x^3 + \frac{7(n-1)}{n^3} B_{k-2}^{n-2} x^2 + \frac{1}{n^3} B_{k-1}^{n-1} x$

$\frac{1}{n^3} [(n-1)(n-2)(n-3)x^4 + 6(n-1)(n-2)x^3 + 7(n-1)x^2 + x - 4n(n-1)(n-2)x^4 - 12n(n-1)x^3 - 4nx^2 + 6n^2(n-1)x^4 + 6n^2x^3 - 3x^4]$

$= n^{-3}[x + (3n-7)x^2 + (12-6n)x^3 + (3n-6)x^4] =$

$n^{-3}[(x - x^2) + (3n-6)x(1-x)^2)] = n^{-3} x(1-x)(1 + (3n-6)(1-x))$

Theorem

Order 2: For one of the sub domain with volume $B_d n^{-\alpha}$

$$\begin{aligned} & P(|\sharp S/n - n^{-\alpha}| > n^{-\beta}) \\ = \sum_{|\frac{k}{n} - n^{-\alpha}| > n^{-\beta}} & C_n^k (n^{-\alpha})^k (1 - n^{-\alpha})^{n-k} \\ < \sum n^{2\beta} \left|\frac{k}{n} - n^{-\alpha}\right|^2 & C_n^k (n^{-\alpha})^k (1 - n^{-\alpha})^{n-k} \\ = n^{2\beta-\alpha} & (1 - n^{-\alpha})/n < n^{2\beta-\alpha-1} \end{aligned}$$

Next we use order 4.

Theorem

Order 4: For one of the sub domain with volume $B_d n^{-\alpha}$

$$\begin{aligned} & P(|\sharp S/n - n^{-\alpha}| > n^{-\beta}) \\ = & \sum_{|\frac{k}{n} - n^{-\alpha}| > n^{-\beta}} C_n^k (n^{-\alpha})^k (1-n^{-\alpha})^{n-k} \\ < & \sum n^{4\beta} |\frac{k}{n} - n^{-\alpha}|^4 C_n^k (n^{-\alpha})^k (1-n^{-\alpha})^{n-k} \\ = & 3n^{4\beta-\alpha}(1-n^{-\alpha})/n^2 < \mathcal{O}(n^{4\beta-\alpha-2}) \end{aligned}$$

Data on the ball and uniformly distributed in probability

Therefore the probability that the discrepancy atleast in one of the subdomain of the volume $B_d n^{-\alpha}$ are larger than $n^{-\beta}$ will be less than $n^{2\beta-1-\alpha} n^{\alpha}$.

$$P(|\frac{|B_d|}{n}\sum n^{d\gamma} f(x_j)\phi(n^{\gamma}(x-x_j)) - f(x)| > \mathcal{O}(n^{-2\beta/(2+d)})) < n^{2\beta-1}$$

Or use order 4:

$$P(|\frac{|B_d|}{n}\sum n^{d\gamma} f(x_j)\phi(n^{\gamma}(x-x_j)) - f(x)| > \mathcal{O}(n^{-2\beta/(2+d)})) < 3n^{4\beta-2}$$

Theorem

Infact. the integral taken only in the support of the $\phi(n^{\gamma}x)$, *we have* $n^{-\gamma}n^{\alpha}$ *sud domain with the volume* $B_d n^{-\alpha}$ *for numerical integral. For any given* x, *the quasi interpolation with large error possesses samll probability*

$$\begin{aligned} < n^{2\beta-1} &\text{ can be sharp to } < n^{2\beta+\gamma-\alpha-1} \\ < n^{4\beta-2} &\text{ can be sharp to } < n^{4\beta+\gamma-\alpha-2} \end{aligned}$$

? tobe check again, if it is right.

Data on the ball and uniformly distributed in probability

Data on the ball and uniformly distributed in probability

Above are the second time to Sun XinPing.

Outline

Learning from the view point of the approximation

In the learning theory, the key feature is to estimate the density function or the distribution function from the sampling data $\{x_j, y_j\}$, after that we can estimate the conditional expectation $E(y|x)$ to simulate the underlaying function for learning or for the classification.

What in the learning theory doing is to assume that the underlaying function is in some given function space and then to find the approximation to minimize the mixed energy function. In most of the cases, the energy is defined more artificially and short of the physical meaning.

Learning from the view point of the approximation

Begin with the univariate problem that using Galton's Plate to simulate t

If the sampling points $\{\xi_j^n\}_{j=1}^n$ are taken from a probability distribution with density function $p(x)$, then we can estimate the density function $p(x)$ via Galton's model, which is the base ground of the large number theorem, we use $\sum q(x-\xi_j^n)$ to approximate the density function $p(x)$, where $q(x)$ is the ball function, therefor $q(x-\xi_j^n)$ are balls fall at the point ξ_j^n respects to the distribution of $p(x)$.

Learning from the view point of the approximation

The ball can be replaced by other particles "dirt" too, that is the reason why we use the function q. If $\int q(x)dx = 1$. To keep the total volume of the balls are keeping equal to 1, the balls should shrink or deflate as more and more balls fall in to the Galton's plant. Following are some of the models.

Learning from the view point of the approximation

- $p(x) \sim \sum q(x - \xi_j^n)/n$, deflate only in hight.
- $p(x) \sim \sum q(n(x - \xi_j^n))$, deflate only in width.
- $p(x) \sim \sum q(\sqrt{n}(x - \xi_j^n))/\sqrt{n}$, deflate both in hight and width in same scale. (for multi variate $x \in R^d$ case, $p(x) \sim \sum q(n^{\frac{1}{d+1}}(x - \xi_j^n))/n^{\frac{1}{d+1}})$
- Or even with more complicated model: $p(x) \sim \sum q(n^p(x - \xi_j^n))/n^{1-pd})$, deflate both in hight and width but not in same scale.

Learning from the view point of the approximation

Let $X \in R$ denotes a real valued random variable with $p_0(x)$ be its probability density function and $p_1(x)$ as their probability distribution function respectively. Moreover, let $p_{k+1}(x)$ be the integral of $p_k(x)$ to get $p_2(x)$,$p_3(x)$ etc.

At first, we should point out that the density function or the distribution do not form a function space, that is only a half supper plan, since $p_0(x) \geq 0$ and the convex combination is still a density function.

Learning from the view point of the approximation

With the notations mentioned above, we can define three types of discrepancy between two random variables. (We use the L_2 type of the norm for reason of calculation)

$$D_0 = \int |p_0(x) - q_0(x)|^2 dx$$

not suitable, since p_0 may be δ function, the integral will not exists.

$$D_1 = \int |p_1(x) - q_1(x)|^2 dx$$

Not always integrable, depend on the decay of the $p_1(x) - q_1(x)$. Valid if p_0, q_0 compactly supported.

$$D_2 = \int |p_2(x) - q_2(x)|^2 dx$$

not suitable, since the integral is certainly not integrable

Examples

Here are some examples to show the problem of the norm definition.

- Let $p_0(x) = 2\chi_{[0,\frac{1}{2}]}(x)$ and $q_0(x) = 2\chi_{[\epsilon,\epsilon+\frac{1}{2}]}(x)$, then L_∞ norm of $p_1 - q_1$ tends to zero as $\epsilon \to 0$, but the L_∞ norm of $p_0 - q_0$ does not.
- Let $p_0(x) = \delta(x)$ and $q_0(x) = \delta(x - \epsilon)$, then L_2 norm of $p_1 - q_1$ tends to zero as $\epsilon \to 0$, but the L_∞ norm of $p_1 - q_1$ does not.

Wasserstein Metric

Intuitively, the distribution is viewed as a unit amount of "dirt" piled on the M (simply R), the difference of the two distribution is the minimal cost of turning one pile in to the other, which is assumed to be the amount of the dirt that needs to be moved times the distance it has to be moved—-earth mover's distance.

Because it is a minimal problem, Difficulty to calculate.

Perhaps

$$|\int x(p_1(x) - q_1(x))dx|$$

can be served as a substituter.

Based on L_p

Nevertheless the L_2 norm of the error of the function $p_1(x) - q_1(x)$ is more suitable. We assume the density function possesses fast decay (compact supported), the error can be defined as to be a weighted L_2 norm of $p_1(x) - q_1(x)$ that

$$\|p_1(x) - q_1(x)\|^2 = \int w(x)(p_1(x) - q_1(x))^2 dx$$

Learning from the view point of the approximation

p_0, p_1, p_2 are all in a half supper plane, only positive weighted average is allowed (linea combination is not allowed).
A approach is to move $p_0(x)$ to $p_0(x) - \delta(x)$ (still only positive linear combination is allowed)
误差应该是与某$\delta(x - E_0)$的距离即

$$D(p) + (E(p) - E_0)^2$$

点空间中可以定义原点，譬如delta（x）或高斯分布（位置与方差？）。

First we consider the problem of the point space via vector space.
By using the δ distribution.
$p_0(x) = \delta(x - e_0) + (p_0(x) - \delta(x - e_0))$, where $e_0 = E_{p_0}(x)$.
Therefor we can define the norm of the $p_0(x)$ that
$\|p_0(x) - \delta(x - e_0)\|^2$, 这样原点是动态的？

Other extension of the definition of the norm

The reasonable definition of the difference of two distribution $p_1(x) - q_1(x)$ is the L_1 -norm

$$\|p_1(x) - q_1(x)\|_1 = \int_{-\infty}^{\infty} |p_1(x) - q_1(x)| dx$$

From the Lesbuge integral it is equivalent to

$$\int_{-\infty}^{\infty} |p_1(x) - q_1(x)| dx = \int_0^1 |p_1^{-1}(c) - q_1^{-1}(c)| dc$$

However

$$\int_{-\infty}^{\infty} |p_1(x) - q_1(x)|^2 dx \neq \int_0^1 |p_1^{-1}(c) - q_1^{-1}(c)|^2 dc$$

The last formula means the variance of the point, which the distribution archive the same value of c, more reasonable to

Learning from the view point of the approximation

Note the facts that

$$\widehat{p_2'} = i\omega\widehat{p_2} = \widehat{p_1}$$

and

$$\widehat{p_1'} = i\omega\widehat{p_1} = \widehat{p_0},$$

moreover we have

$$\widehat{p_0} = i\omega\widehat{p_1} = -\omega^2\widehat{p_2}$$

The D_0-discrepancy will certainly lead to a L_{2,D_0}-norm and a L_{2,D_0}-inner product.

$$\begin{aligned} < p_0, q_0 >_{L_{2,D_0}} &= \int p_0(x) \cdot q_0(x) dx = \int \widehat{p_0}(\omega) \cdot \widehat{q_0}(\omega) d\omega \\ &= \int \omega^2 \widehat{p_1}(\omega) \cdot \widehat{q_1}(\omega) d\omega = \int \omega^4 \widehat{p_2}(\omega) \cdot \widehat{q_2}(\omega) d\omega \end{aligned}$$

Learning from the view point of the approximation

Noticing that

$$\widehat{|x|^{\alpha}} = \omega^{-\alpha-1},$$

we have

$$\begin{aligned} < p_0, q_0 >_{L_{2,D_0}} &= \int p_0(x)\cdot q_0(x)dx = \int \widehat{p_0}(\omega)\cdot\widehat{q_0}(\omega)d\omega \\ &= \int \frac{\widehat{p_1}(\omega)\cdot\widehat{q_1}(\omega)}{\omega^2}d\omega = \int \frac{\widehat{p_2}(\omega)\cdot\widehat{q_2}(\omega)}{\omega^4}d\omega \end{aligned}$$

From the example that the norm for de difference of two distributions function should better be $\|p_1(x) - q_1(x)\|_2$, therefor the respected inner product should be

$$< p_1, q_1 >_{L_{2,D_0}} = \int p_1(x) \cdot q_1(x) dx = \int \widehat{p}_1(\omega) \cdot \widehat{q}_1(\omega) d\omega$$

However the distributions function p_1, q_1 is usually do not integrable, since $\lim_{x\to\infty} p_1(x) = 1$ 这里是问题的关键，需要办法

Using the knowledge of **Reproducing Kernel Hilbert Space**, we know that $|x|^3$ is the reproducing kernel of p_2(这里不知道该怎么表述比较合适。)

Learning from the view point of the approximation

Consider unit interval $[0,1]$. For such an interval $I \subseteq [0,1]$ and a sequence $\{x_j\}_{j=1}^n$, $x_j \in I$.

From the definition of three kinds of discrepancy and two examples, using the weak-star discrepancy to measure the difference of the two probability P_1 and P_2 on $[0,1]$:

$$\int_0^1 |P_1(t) - P_2(t)|dt,$$

is more suitable in applications. Obviously, it is a L_1-norm of $P_1 - P_2$. For the convenience of the later study, we use L_2-norm instead.

Learning from the view point of the approximation

Using Parseval's equality,

$$\int [P_1(x) - P_2(x)]^2 dx = \int \frac{[\hat{p}_1(\omega) - \hat{p}_2(\omega)]^2}{\omega^2} d\omega.$$

Then an inner product can be derived from the norm,

$$< p(x), q(x) >= \int \frac{\hat{p}(\omega) \cdot \hat{q}(\omega)}{\omega^2} d\omega = \int \frac{\hat{p}(\omega) \cdot \hat{q}(\omega)}{\hat{G}(\omega)} d\omega$$

and become a reproducing kernel Hilbert space. The kernel can be represented as to be $G(x) = \frac{1}{|x|^3}$

Learning from the view point of the approximation

We now place ourselves in the world of random variables and probability spaces. Let $X_1 \in R$ and $X_2 \in R$ denote two real valued random variables with $P_1(x)$ and $P_2(x)$ as their probability distribution function respectively. Moreover, let $p_1(x)$ and $p_2(x)$ be their probability density function respectively. With the notations mentioned above, we can define three types of discrepancy between two random variables.

Learning from the view point of the approximation

Definition

Strong discrepancy D_s is defined to be

$$D_s = \sup_{x \in [0,1]} |p_1(x) - p_2(x)|$$

weak discrepancy D_w is defined to be

$$D_w = \sup_{x \in [0,1]} |P_1(x) - P_2(x)|$$

weak-star discrepancy D_{w*} is defined to be

$$D_{w*} = \int_0^1 |P_1(t) - P_2(t)| dt,$$

Learning from the view point of the approximation

For the error of two density function of distributions.
1. The maximum norm $\|p_1(x) - p_2(x)\|_\infty$ can not describe the error very good. For example $p_1(x) = \mathrm{sign}(\sin(nx))_+$ and $p_2(x) = \mathrm{sign}(\sin(nx - h))_+$. Obviously $h \to 0$, we get convergence. However the L_∞ norm of the error is always 1.
2. The maximum norm of $\|P_1 - P_2\| < \pi h/n$

Learning from the view point of the approximation

Consider the Galton's needle plant for simulating the Gaussian distribution by drooping balls. If the sampling points $\{x_j\}_{j=1}^{n}$ is come from the distribution of $p(x)$, we use $\sum q(\sqrt{n}(x-x_j))/\sqrt{n}$ to simulate the density function $p(x)$. If we choose ball, then $q(x) = c\sqrt{(}1-x^2)$. The problem is that, which $q(x)$ will approximate the density function $p(x)$ better.

Then we want

$$\| \int p(x) - \sum \sum q(\sqrt{n}(x-x_j))/\sqrt{n} \|$$

to be minimized. If we use L_2 to the the weak discrepancy as the error measurement.

$$Err = \int (P(x) - \sum Q(\sqrt{n}(x-x_j))/n)^2 dx$$

Learning from the view point of the approximation

Outline

Changing Basis Functions

Thanks for Your Attentions!

L_2-norm, $\int (p_1(x) - q_(x))^2 dx$,
Remark:p_1 is in a half-supper plane, (Only positive convex combination is still a distribution)
Derive a inner product $< p, q >= \int p_1(x) q_1(x) dx$, however $\int p_1^2(x) dx < \infty$ is not satisfied.
Let $p_1(x) := p_1(x) - \int_{-\infty}^{x} \delta(x) dx$, now the function is in L_2 (especially valid for compactly supported $p(x)$, which is mostly exist in application).
Now $\|p_1(x)\|_{L_2}^2$ means the bias (expectation^2 + variance) to the δ-distribution.

Dirt moving: $p_1(x) = \int_{-\infty}^{x} p(x)dx$. To minimize the energy of moving the dirt, the moving queue should be not crossover, then the dirt on $p_1(x) = c$ should move to $q_1(x)c$, therefore the total energy can be written in

$$\int |p_1^{-1}(y) - q_1^{-1}(y)|dy = \int |x - q_1^{-1}(p_1(x))|p(x)dx$$

Generalize to L_2, $\int (p_1^{-1}(y))^2 dy = \int x^2 p(x)dx$, (expectation^2 + variance).

$\phi(x)$ standard dirt. $\int \phi(x)dx = 1$, $\phi(x) \geq 0$.
Model: $p_n(x) = \sum \phi(n^p(x - x_j)) \cdot n^p/n$, where x_j sampling from the distribution of $p(x)$.
Fourier: $\{\frac{1}{n}\sum e^{ix_j w}\} \cdot \hat{\phi}(\frac{w}{n^p}) \to \hat{p}$.

boundary distribution:
$p(x) = \int p(x,y)dy$, $p(y) = \int p(x,y)dx$
$p_1(x) = \int_{-\infty}^{x} p(x)dx$,

$$p_1(y) = \int_{-\infty}^{y} p(y)dy$$

.
The dirt of $p_1(x) = c$ should be more to the $q_1(x) = c$, adopted from the univariate approach.
On the $p_1(x) = c$, again adopt the univariate approach more $p(x,y)/p(x)$ to $q(x,y)/q(x)$.
These perhaps not the shortest way to more the dirt(second step may be not the line but curve), however is a simulation of the best way of moving of the dirt.

$$k(x,y)=\sum b_j(x)b_j(y), \quad \{b_j(x)\}\text{authornomal} \ \ \text{basis}$$

$$f(x) =< f(t), k(x,t) >$$

$$f(x+c) =< f(t+c), k(x,t) >=< f(t), k(x,t-c) >=< f(t), k(x+c,t) >$$

Polynomial generator $(1+xy)^d$,
$\sum \lambda_j \lambda_k (1+x_j x_k)^d = \sum C_d^l \sum x_j^{2(l-1)} \cdot (\sum \lambda_j x_j)^2$
For reproducing kernel

$$Lf =< f(t), L_x k(x-t) >=< f(t), L_t^* k(x-t) >=< L_t f(t), k(x-t) >$$

$$f(x) = \int g(t)\phi(\frac{x-t}{c})/c dt$$

A convolution with a standard δ, as $c \to 0$. Taken Fourier transform

$$\hat{g}(t) = \hat{f}(t)/\hat{\phi}(ct)$$

Intuitively: $\hat{\phi}(ct)$ should be large where the $\hat{f}(t)$ are large, subject to the integrable of the integral.

A simple simulation:

$$\int t|\hat{f}(t)|dt = \int t|\hat{\phi}(ct)|dt$$

to get the best paraments c.

To bandlimited function the condition of the integrable can not be considered.

More accurately

$x' = p(x)$, using polynomial $p(x)$ to approximate the behaviors. What means the small perturbation of $p(x)$, depend on the choosing of the basis.
In different display of basis, the coefficients, is not same distributed, for some basis, one of the coefficient may be very small and can be ignored. By using the concept of frame, one can fine the large coefficients.