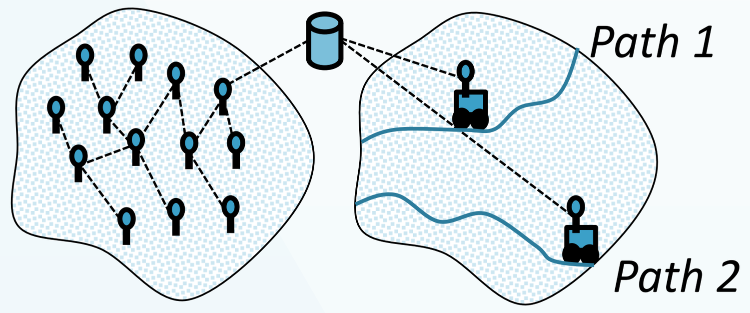
Distribution Learning of a Random Spatial Field with a Location-Unaware Mobile Sensor

Meera Pai and Animesh Kumar

Electrical Engineering, Indian Institute of Technology, Bombay

Introduction

- Learning the (statistical) distribution of physical fields from observed values is a fundamental task in applications like, environment monitoring, smart city and pollution control.



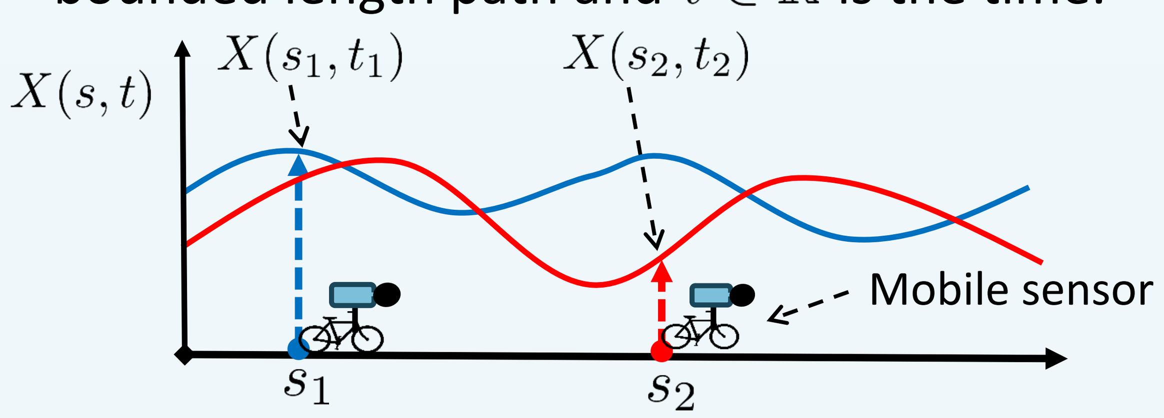
-fixed sensor -mobile sensor -remote processing node

Spatial field sensing: (a) with fixed grid of sensors. (b) with mobile sensors.

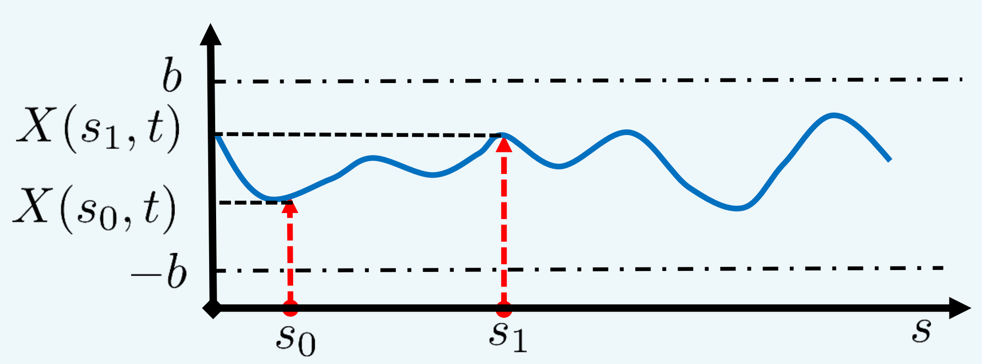
- Mobile sensing is a way to increase spatial sampling density however their location uncertainty is a challenge.
- Since localization methods (like GPS and wireless outdoor localization) require extra energy and hardware, we aim to learn the statistical distribution of spatial fields as a function of space using location unaware mobile sensors.

Spatial field

- The field of interest is $X(s,t)$ where $s \in \mathcal{P}$ is a bounded length path and $t \in \mathbb{R}$ is the time.



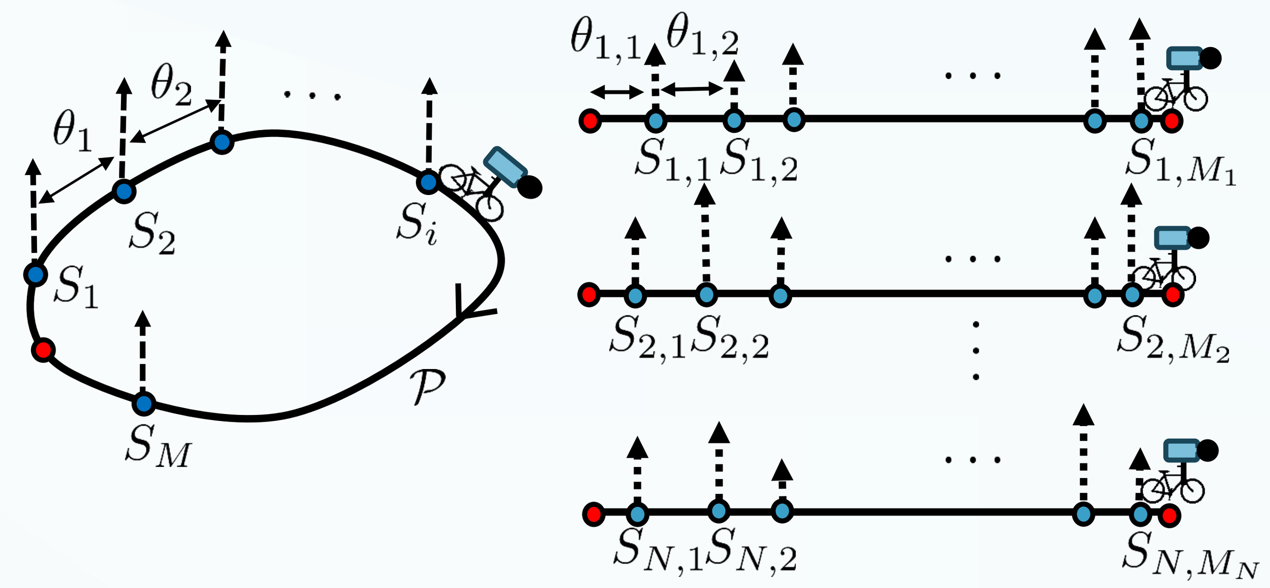
- The field is assumed to be time varying, bounded and Lipschitz continuous.



$|X(s_1,t) - X(s_0,t)| \leq \alpha |s_1 - s_0|$

Sensing model

- The spatial field is sampled using a location unaware mobile sensor.
- N trials of the mobile sensing are performed.

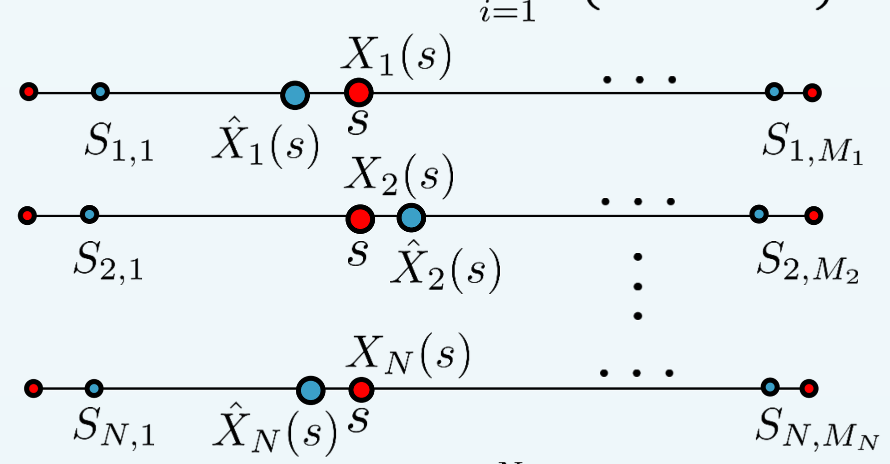


- The unknown sampling locations $S_1, S_2, \ldots, S_M$ are modeled using an unknown renewal process.
- The random variables $\theta_1, \theta_2, \ldots$ are independent and identically distributed.
- It is assumed that $0 < \theta \leq \frac{\lambda}{n}$ and $\mathbb{E}(\theta) = \frac{1}{n}$ where $\lambda > 1$ is finite.
- average sampling rate is n per meter and λ represents maximum sensor speed.

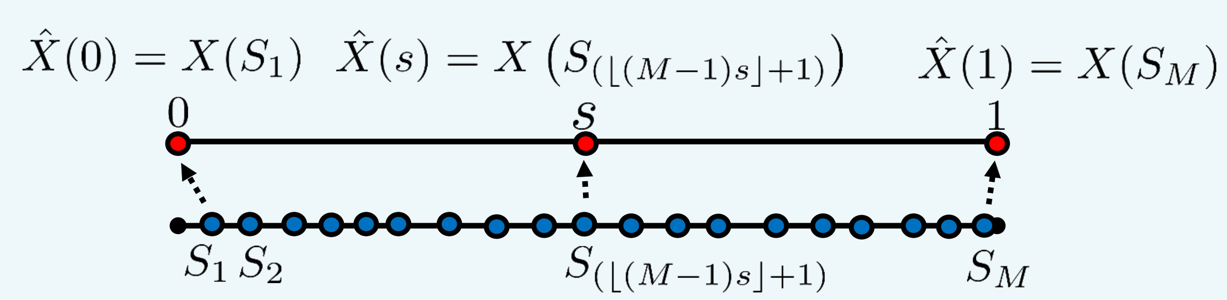
Distribution Learning method

- We wish to estimate the spatial field distribution at location s, $F_{X(s)}(x) = \mathbb{P}(X(s) \leq x)$ using samples from the location unaware mobile sensor.

$$F_{X(s)}(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\left\{X_i(s) \leq x\right\}$$



$$F_{\hat{X}(s)}(x) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}\left\{\hat{X}_i(s) \leq x\right\}$$



- The error between $F_{X(s)}$ and its estimate $F_{\hat{X}(s)}(x) = \mathbb{P}(\hat{X}(s) \leq x)$ depends on error in location estimation $|S_{\lfloor (M-1)s+1 \rfloor} - s|$ and smoothness of the field i.e. α.

- The mean squared error between location s and its estimate $S_{\lfloor (M-1)s \rfloor + 1}$ is

$$\mathbb{E}\left[\left|S_{l(M,s)} - s\right|^2\right] \leq ((n+\lambda-1)s(1-s) + C)\frac{\lambda^2}{n^2},$$

and the error in field estimation is

$$\mathbb{P}\left(|X(s) - X(S_{\lfloor (M-1)s \rfloor + 1})| > \varepsilon\right) \leq \frac{\alpha^2}{\varepsilon^2}((n+\lambda-1)s(1-s) + C)\frac{\lambda^2}{n^2}.$$

This upper bound depends on s and has a maximum at $s = 1/2$.

> *Theorem 1:* For every $x \in \mathbb{R}, s \in [0,1]$ and for any $\varepsilon > 0$,
> $$|F_{X(s)}(x) - F_{\hat{X}(s)}(x)| \leq \varepsilon . \max\left(f_{X(s)}(x)\right) + \frac{\alpha^2}{\varepsilon^2}((n+\lambda-1)s(1-s) + C)\frac{\lambda^2}{n^2}.$$

- The supremum of error in field estimation over $s \in [0,1]$ is

$$\mathbb{P}\left(\sup_{s\in[0,1]} |X(s) - X(S_{\lfloor (M-1)s \rfloor + 1})| > \varepsilon\right) \leq \frac{32}{\beta}\frac{\alpha^2}{\varepsilon^2}(n+\lambda-1)\frac{\lambda^2}{n^2}.$$

> *Theorem 2:* For every $x \in \mathbb{R}, s \in [0,1]$ and for any $\varepsilon > 0$,
> $$\sup_{s\in[0,1]} \left|F_{\hat{X}(s)}(x) - F_{X(s)}(x)\right| \leq \varepsilon . \max\left(f_{X(s)}(x)\right) + \frac{32}{\beta}\frac{\alpha^2}{\varepsilon^2}(n+\lambda-1)\frac{\lambda^2}{n^2}$$

- β is the symmetrisation constant.
- If $\varepsilon = \frac{1}{n^{1/3}}$

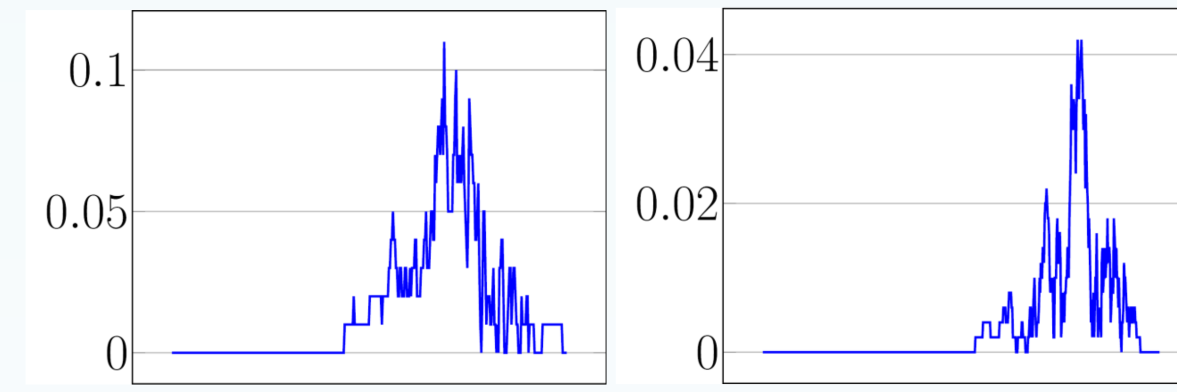
$$|F_{X(s)}(x) - F_{\hat{X}(s)}(x)| \leq (\max\left(f_{X(s)}(x)\right) + \alpha^2 s(1-s))\frac{\lambda^2}{n^{1/3}} + \alpha^2((\lambda-1)s(1-s) + C)\frac{\lambda^2}{n^{4/3}}.$$

Simulation

- To validate our distribution learning method we consider acoustic levels varying along a 1-D path.
- The spatial field is simulated by

$$X(s,t) = \left|1000 + \sum_{r=1}^{10} A_r(t)\cos(2\pi f_r(t) s)\right|.$$

- The sampling locations are modelled by a renewal process where inter-sample intervals are modelled by a Rayleigh distribution with parameter $\frac{1}{n}\sqrt{\frac{2}{\pi}}$.



(a) for N=100 and n=100 *(b) for N=100 and n=1000*

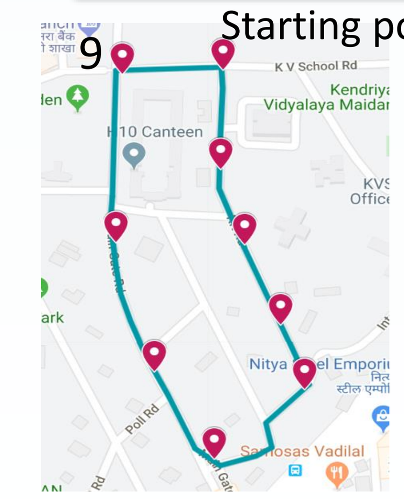
$\left|\hat{F}_{X(s)}(\frac{1}{2}) - \hat{F}_{\hat{X}(s)}(\frac{1}{2})\right|$

Experiment

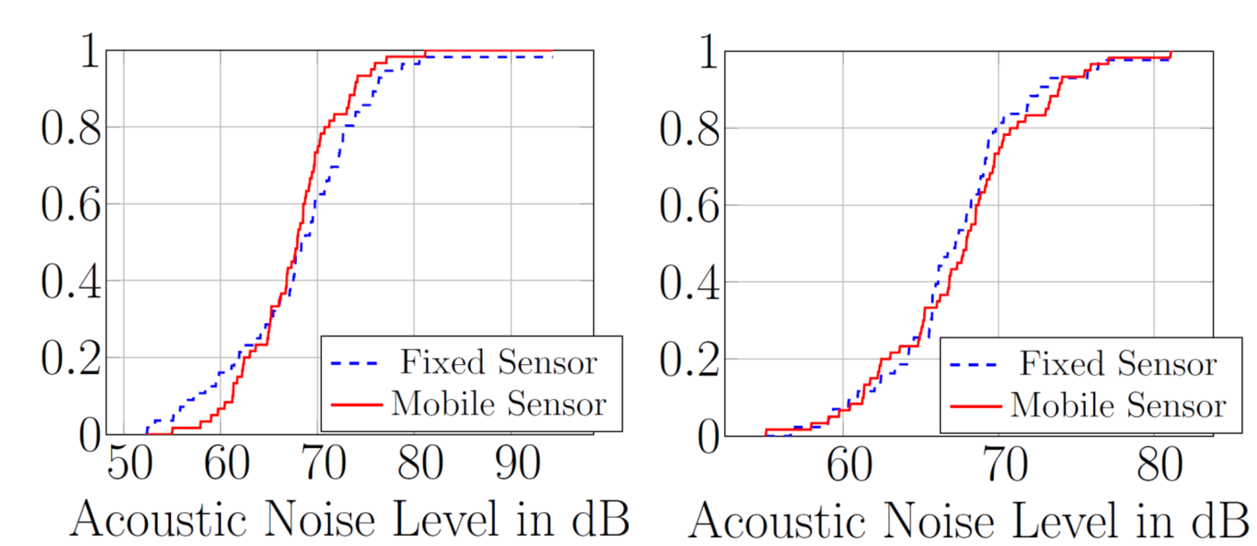
- Acoustic noise levels are measured along the path shown in the map using the sound meter by BAFX products (Model no: BAFX3608).

Specifications of sound meter

Range: 30-130 dB	*Sampling rate: 1 per sec*
Memory: 4700 readings	*Accuracy: ± 1.5 dB*



Dataset is collected on the path shown in the map.
Dataset 1: 43 trials of mobile sensing by walk, Samples of static sensors
Dataset 2: 43 trials of mobile sensing using a bicycle , Samples of static sensor



(a) location 9 using dataset 2 *(b) location 9 using dataset 1*

$\hat{F}_{X(s)}(s)$ vs $\hat{F}_{\hat{X}(s)}(s)$

Conclusion

- We proposed a data-driven method for learning the statistical distribution of a Lipschitz continuous spatial field along a path.
- We showed that the accuracy of the distribution-learning method increases with the spatial sampling rate of the mobile sensor.

Acknowledgements

We would like to express our gratitude towards Google for their generous support towards conference travel. We would also like to thank NeurIPS for funding the stay during the conference.