Abstract

Résumé:
We introduce a novel kernel density estimator for a large class of symmetric spaces and prove a minimax rate of convergence as fast as the minimax rate on Euclidean space. We prove a minimax rate of convergence proven without any compactness assumptions on the space or Hölder-class assumptions on the densities. A main tool used in proving the convergence rate is the Helgason-Fourier transform, a generalization of the Fourier transform for semisimple Lie groups modulo maximal compact subgroups. This paper obtains a simplified formula in the special case when the symmetric space is the 2-dimensional hyperboloid.

Collection

Kernel Density Estimation on Symmetric SpacesDena Asta
We introduce a novel kernel density estimator for a large class of symmetric spaces and prove a minimax rate of convergence as fast as the minimax rate on Euclidean space. We prove a minimax rate of convergence proven without any compactness assumptions on the space or Hölder-class assumptions on the densities. A main tool used in proving the convergence rate is the Helgason-Fourier transform, a generalization of the Fourier transform for semisimple Lie groups modulo maximal compact subgroups. This paper obtains a simplified formula in the special case when the symmetric space is the 2-dimensional hyperboloid.

Authors

Dena Asta

Kernel Density Estimation on Symmetric Spaces

Metrics

Downloaded: 123
Viewed: 4
Size: 3.79 MB
Type: application/pdf
URI: bitcache://fc0c9959d31ac45d4320403aa2b7bea5d87a71db

License

La SEE (Société de l'Electricité, de l'Electronique et des Technologies de l'Information et de la Communication – Association reconnue d'utilité publique, régie par la loi du 1er juillet 1901) met à la disposition de ses adhérents et des abonnés à ses publications, un ensemble de documents numériques accessibles à partir de son portail des publications. Ces documents incluent notamment les articles des revues REE, 3 EI et e-STA disponibles sous forme numérique ainsi que des publications additionnelles regroupées dans l'espace eREE. Les présentes conditions précisent les conditions de diffusion et d'utilisation de ces documents et des informations qu’ils contiennent. L’accès à ces documents, qu’il se fasse de façon gratuite ou dans le cadre d’abonnements ou d’achats faits à titre onéreux, implique l’acceptation sans restriction de ces dispositions.

Droits de propriété et de diffusion des contenus téléchargés sur le portail des publications

Les contenus rendus accessibles sur le portail des publications sont, en règle générale, protégés par le droit d'auteur. En tant que producteur, et le cas échéant d’auteur, des informations rassemblées dans les contenus accessibles par ce portail, SEE se réserve l'exclusivité des droits de copie et de diffusion de tout ou partie de ces contenus.

Les contenus sont rendus accessibles à titre individuel, pour les besoins de la personne en
détenant des droits d'accès en cours de validité. Aussi, la modification, la reproduction et/ou la diffusion via Internet ou le Web, intranet, extranet ou toute autre forme numérique ou imprimée, de tout ou partie des contenus téléchargés sont interdites. Une tolérance est consentie quant à la reproduction d’extraits limités de ces contenus, dans le cadre de travaux ou d’activités auxquels ils sont utiles, à la condition que l’origine de ces reproductions partielles soit mentionnée de façon lisible et sans ambiguïté. Figurent en particulier : la REE (ou toute autre revue accessible sur le portail) en tant que la source, la référence de la publication et le nom de l’auteur (s’il figure dans la revue).

Ces dispositions s’appliquent également aux figures, illustrations, logos ou images.

**Publication externe des contenus du portail des publications**

Tout extrait des contenus du portail destiné à être utilisé dans des publicités, des communiqués de presse ou du matériel de promotion nécessite un accord préalable écrit de la SEE. Une version préliminaire du document proposé contenant ces extraits doit accompagner chacune de ces demandes. SEE se réserve le droit de refuser un tel usage externe pour quelque raison que ce soit.

**Responsabilités**

La SEE apporte tout le soin possible à la préparation des informations délivrées dans les contenus produits. Cependant elle ne peut être tenue pour responsable d’aucune perte ou frais qui pourrait résulter d’imprécisions, d’inexactitudes, d’erreurs ou de possibles omissions portant sur des informations publiées, ni des résultats obtenus par l’utilisation et la pratique des informations délivrées.

**Utilisation des informations recueillies lors du téléchargement de contenu**

Le portail des publications est susceptible d’utiliser des « cookies » afin notamment de permettre l’utilisation de paniers d’achat et de personnaliser les parcours sur le site. SEE se réserve la possibilité d’utiliser les informations recueillies lors des téléchargements pour ses besoins internes et notamment pour l’amélioration de ses services, sans qu’elles puissent être cédées à des partenaires commerciaux. Conformément à la loi “informatique et libertés” du 6 janvier 1978, chaque utilisateur du portail dispose d’un droit d’accès et de rectification aux informations qui le concernent. Pour exercer ce droit, les utilisateurs doivent s’adresser à SEE – 17 rue de l’amiral Hamelin – 75783 Paris Cedex 16, par simple lettre ou en utilisant le formulaire de contact disponible sur son site.

Paris, le 28 avril 2013

**Sponsors**

**Organizers**
Dena Marie Asta  
Department of Statistics  
Ohio State University  
Supported by NSF grant DMS-1418124 and NSF Graduate Research Fellowship under grant DGE-1252522.

Kernel Density Estimation on Symmetric Spaces  
2 Geometric Methods for Statistical Analysis

Classical statistics assumes data is unrestricted on Euclidean space. Exploiting the geometry of the data leads to faster and more accurate tools.

\[ \text{var}[X] = E[X^2] - E[X]^2 \]

Implicit geometry in non-Euclidean data explicit geometry in networks

Motivation: Non-Euclidean Data  
3 Normal Distributions  
sphere  
Diffusion Tensor Imaging

Material Stress, Gravitational Lensing, Directional Headings  
3x3 symmetric positive definite matrices

Nonparametric Methods: Non-Euclidean Data

Classical non-parametric estimators assume Euclidean structure. Sometimes the given data has other geometric structure to exploit. Kernel density estimator, kernel regression, conditional density estimator.

Motivation: Non-Euclidean Distances

Normal Distributions  
sphere  
Diffusion Tensor Imaging  
Euclidean distances are often not the right notion of distance between data points. Material Stress, Gravitational Lensing, Directional Headings  
3x3 symmetric positive definite matrices  
3x3 symmetric positive definite matrices  
hyperboloid

Motivation: Non-Euclidean Distances

Euclidean distances are often not the right notion of distance between data points. Material Stress, Gravitational Lensing, Directional Headings. Distance between directional headings should be shortest path-length.

Motivation: Non-Euclidean Distances

Euclidean distances are often not the right notion of distance between data points. Hyperboloid mean standard deviation. An isometric representation of the hyperboloid is the Poincare Half-Plane. Each point in either model represents a normal distribution. Distance is the Fisher Distance, which is similar to KL-Divergence.

Motivation: Non-Euclidean Distances

Euclidean distance not the right distance. Euclidean volume not the right volume. We want to minimize risk for density estimation on a (Riemannian) manifold. Estimator based on n samples true density manifold volume measure based on intrinsic distance. Existing estimators:

\[ \text{E} f M(\text{f}^\star M f n) \frac{d\mu}{d\nu} \]

Exploiting Geometry: Symmetries

Symmetries = geometry. Symmetries make the smoothing of data (convolution by a kernel) tractable. Other spaces call for other symmetries.

Exploiting symmetries to convolve

Kernel density estimation is about convolving a kernel with the data. More general spaces, depending on their geometry, we will require symmetries other than translations.

\[ \text{density on the space of translations on Rn} \]

Identify t
with $T_t$ and interpret $g$ as a density on the space of $T_t$'s. Kernel density estimation is about convolving a kernel with the data. More general spaces, depending on their geometry, we will require symmetries other than translations... density on $\mathbb{R}^n$ density on the space of translations on $\mathbb{R}^n$ $\hat{f}_h(X_1,\ldots,X_n) = Kh \bowtie \text{empirical}(X_1,\ldots,X_n)$ \cite{Asta2014}

$\hat{f}_h(C)(X_1,\ldots,X_n) = Kh \bowtie \text{empirical}(X_1,\ldots,X_n)$ \cite{Terras1985}

$\text{Empirical} = \frac{1}{n} \sum_{i=1}^{n} f(X_i)$

$G$-Kernel Density Estimator: general form density on group of symmetries $G$ "empirical density" on symmetric space $X$ Bandwidth and cutoff parameters sample observations We can use harmonic analysis on symmetric spaces to define and analyze this estimator. \cite{Asta2014}.