Exponential family by representation theory

Koichi Tojo[1], joint work with Taro Yoshino[2]

[1]RIKEN Center for Advanced Intelligence Project, Tokyo, Japan,

[2]Graduate School of Mathematical Science, The University of Tokyo

July 29, 2020

Demonstration

Our aim

One of our aims is to suggest “good” exponential families on important spaces.

We proposed a method to construct exponential families by using representation theory in [TY18].
First, we demonstrate a family of distributions on upper half plane with Poincare metric obtained by using our method.

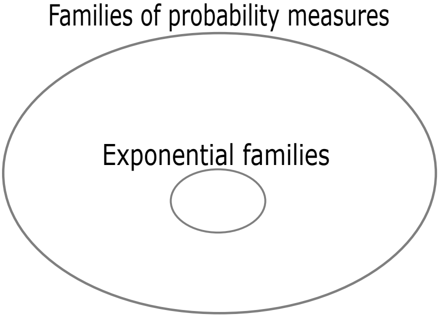
[TY18]: K. Tojo, T. Yoshino, *A method to construct exponential families by representation theory*, arXiv:1811.01394v3.

Contents

A family of probability measures and machine learning

Learning by using a family of probability measures is one of important methods in the field of machine learning.

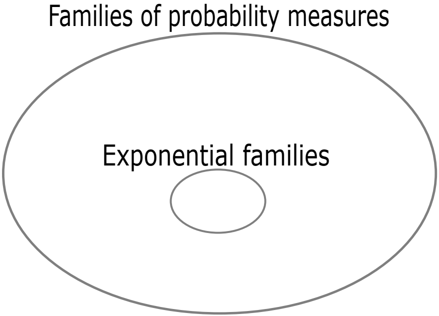
> Learning=to optimize the parameters in the family of probability measures



Exponential family

Exponential family

- Exponential families are important subject in the field of information geometry.
- Exponential families are useful for Bayesian inference.
- Exponential families include many widely used families.



Examples (exponential families)

Table: Examples of exponential families

distributions	sample sp. X
Normal	$\mathbb{R}$
Multivariate normal	$\mathbb{R}^n$
Bernoulli	$\{\pm 1\}$
Categorical	$\{1, \cdots, n\}$
Gamma	$\mathbb{R}_{>0}$
Inverse gamma	$\mathbb{R}_{>0}$
Wishart	$\mathrm{Sym}^+(n, \mathbb{R})$
Von Mises	S^1
Generalized Inverse Gauss.	$\mathbb{R}_{>0}$

Exponential family

X: manifold, $\mathcal{R}(X)$: the set of all Radon measures on X.

Definition 1.1 (exponential family).

$\emptyset \neq \mathcal{P} \subset \mathcal{R}(X)$ is an exponential family on X if there exists a triple (μ, V, T) such that

1. $\mu \in \mathcal{R}(X)$,
2. V is a finite dimensional vector space over $\mathbb{R}$,
3. $T : X \to V$, $x \mapsto T(x)$ is a continuous map,
4. For any $p \in \mathcal{P}$, there exists $\theta \in V^\vee$ such that

$$dp(x) = \exp(-\langle\theta, T(x)\rangle - \varphi(\theta))d\mu(x),$$

where $\varphi(\theta) = \log \int_{x \in X} \exp(-\langle\theta, T(x)\rangle)d\mu(x)$ (log normalizer).

We call the triple (μ, V, T) a *realization* of $\mathcal{P}$.

Example: a family of normal distributions

Example 1.2.

The following family of normal distributions is an exponential family:

$$\left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2} \right) dx \right\}_{(\sigma,m)\in\mathbb{R}_{>0}\times\mathbb{R}}$$

1. μ =Lebesgue measure,
2. $V = \mathbb{R}^2$,
3. $T\colon X = \mathbb{R} \to \mathbb{R}^2,\ x \mapsto \begin{pmatrix} x^2 \\ x \end{pmatrix}$.

Remark on exponential family

We can make too many exponential families.
In fact, for given

1. manifold X with a measure μ,
2. finite dimensional real vector space V,
3. $T\colon X \to V$ continuous,

we obtain an exponential family $\mathcal{P} := \{p_\theta\}_{\theta\in\Theta}$ on X:

$$d\tilde{p}_\theta(x) := \exp(-\langle\theta, T(x)\rangle)d\mu(x),$$

$$\theta \in \Theta := \{\theta \in V^\vee \mid \int_X d\tilde{p}_\theta < \infty\},$$

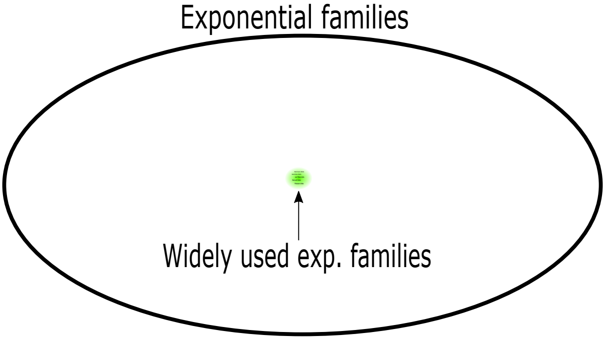
$$\varphi(\theta) := \log \int_X d\tilde{p}_\theta \quad (\theta \in \Theta),$$

$$p_\theta := e^{-\varphi(\theta)}\tilde{p}_\theta.$$

Background

Remark 1.3.

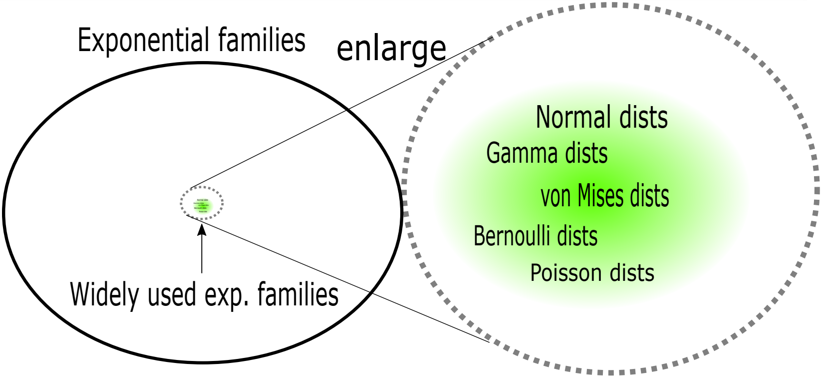
- By definition, there are too many exponential families.
- Only a small part of them are widely used.



Background

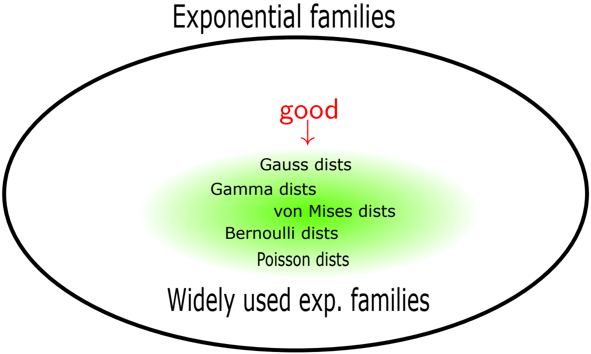
Remark 1.3.

- By definition, there are too many exponential families.
- Only a small part of them are widely used.



Motivation

We can expect there exist "good" exponential families.
We want a framework to understand "good" exponential families systematically.



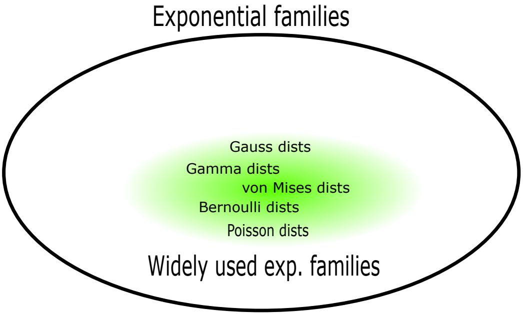
Observation 1.4.

Useful exp. families have the same symmetry as the sample spaces.

- Sample space : homogeneous space G/H
- Family : invariant under the induced G-action

$\rightsquigarrow$ Idea: use representation theory (theory of symmetry)

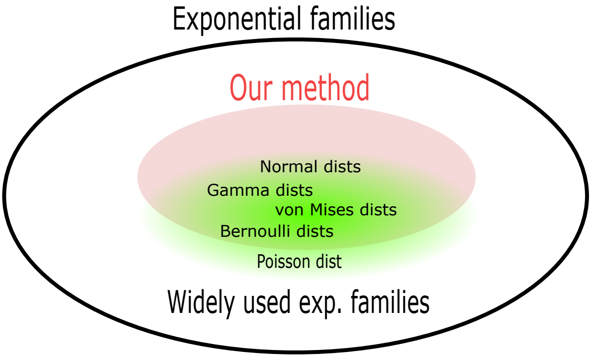
"$X = G/H$" is description focusing on symmetry of the space X.



Our method (*G*/*H*-method)

We proposed a method to construct exponential families.

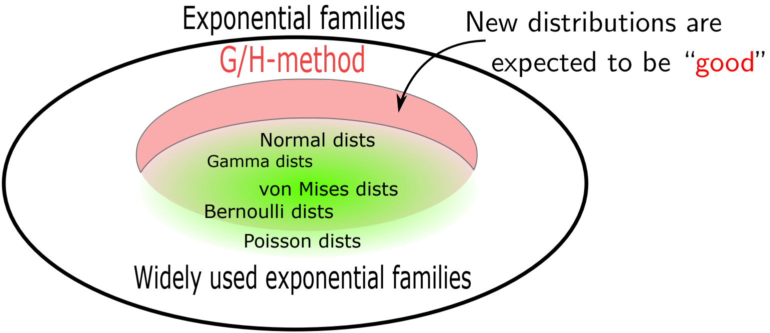
- The method generate many well-known families.
- Families obtained by the method can be classified.



Our method (G/H-method)

We proposed a method to construct exponential families.

- The method generate many well-known families.
- Families obtained by the method can be classified.



G/H-method: overview

G/H-method = a method to construct a family of probability measures on G/H from

- a finite dim. real representation $\rho : G \to GL(V)$,
- a nonzero H-fixed vector v_0.

Step 1 Take Lie groups $G \supset H$ and put $X := G/H$ (sample space)

Step 2 Inputs : finite dim. real rep. (ρ, V) and H-fixed vector

Step 3 Consider the set $\Omega(G, H)$ of all relatively G-invariant measures on X

Step 4 Make measures parameterized by $V^\vee \times \Omega(G, H)$ on X

Step 5 Normalize them.

Examples obtained by our method

Table: Examples and inputs (G, H, V, v_0) for them

Interpretation

distributions	sample sp. X	G	H	V	v_0
Normal	$\mathbb{R}$	$\mathbb{R}^\times \ltimes \mathbb{R}$	$\mathbb{R}^\times$	$\mathrm{Sym}(2,\mathbb{R})$	E_{22}
Multi. normal	$\mathbb{R}^n$	$GL(n,\mathbb{R}) \ltimes \mathbb{R}^n$	$GL(n,\mathbb{R})$	$\mathrm{Sym}(n+1,\mathbb{R})$	$E_{n+1,n+1}$
Bernoulli	$\{\pm 1\}$	$\{\pm 1\}$	$\{1\}$	$\mathbb{R}_{sgn}$	1
Categorical	$\{1,\cdots,n\}$	$\mathfrak{S}_n$	$\mathfrak{S}_{n-1}$	W	w
Gamma	$\mathbb{R}_{>0}$	$\mathbb{R}_{>0}$	$\{1\}$	$\mathbb{R}$	1
Inverse gamma	$\mathbb{R}_{>0}$	$\mathbb{R}_{>0}$	$\{1\}$	$\mathbb{R}_{-1}$	1
Wishart	$\mathrm{Sym}^+(n,\mathbb{R})$	$GL(n,\mathbb{R})$	$O(n)$	$\mathrm{Sym}(n,\mathbb{R})$	I_n
Von Mises	S^1	$SO(2)$	$\{I_2\}$	$\mathbb{R}^2$	e_1
Generalized Inv. Gauss.	$\mathbb{R}_{>0}$	$\mathbb{R}_{>0}$	$\{1\}$	$\mathbb{R}^2$	$\frac{1}{2}\begin{pmatrix}1\\1\end{pmatrix}$

Here $W = \{(x_1,\cdots,x_n) \in \mathbb{R}^n \mid \sum_{i=1}^n x_i = 0\}$,
$w = (-(n-1), 1, \cdots, 1) \in W$.

G/H-method Step 1: sample space

Representation theory: Take a symmetry G of the sample space.

Setting 2.1.

G: a Lie group,
H: a closed subgroup of G,
$X := G/H$: homogeneous space.

We construct a family $\mathcal{P} := \{p_\theta\}_{\theta \in \Theta}$ of probability measures on the homogeneous space X.

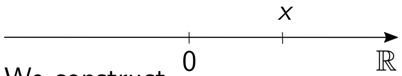
Normal distributions on $\mathbb{R}$:
$\mathbb{R}$ has symmetry of scaling and translation.

Setting 2.2.

$G = \mathbb{R}^\times \ltimes \mathbb{R}$, (scaling, translation)
$H = \mathbb{R}^\times$,
$X = G/H \simeq \mathbb{R}$.

We identify G/H with $\mathbb{R}$ by

$$G/H \ni (t, x)H \mapsto x \in \mathbb{R}$$



We construct

$$\left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx \right\}_{(\sigma,m)\in\mathbb{R}_{>0}\times\mathbb{R}}.$$

G/H-method Step 2: inputs

Representation theory:

inputs

1. V: a finite dimensional real vector space.
2. $\rho : G \to GL(V)$ is a representation.
3. H-fixed vector $v_0 \in V^H$

Normal distributions:

inputs

1. $V := \mathrm{Sym}(2, \mathbb{R})$
2. $\rho : \mathbb{R}^{\times} \ltimes \mathbb{R} \to GL(\mathrm{Sym}(2, \mathbb{R}))$
   $$\rho(t,x)S := \begin{pmatrix} t & x \\ & 1 \end{pmatrix} S \begin{pmatrix} t & \\ x & 1 \end{pmatrix}$$
3. $v_0 := \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \in V^H$

G/H-method Step 3: relatively G-invariant measures

We consider measures compatible the the symmetry.

Definition 2.3.

A measure $\mu \in \mathcal{R}(X)$ is relatively G-invariant $\overset{\text{def}}{\Longleftrightarrow}$ ${}^{\exists}\chi : G \to \mathbb{R}_{>0}$ continuous map s.t. $d\mu(gx) = \chi(g)d\mu(x)$.

Representation theory:
We put
$\Omega(G, H) := \{\mu \in \mathcal{R}(X) \mid \mu$ is relatively G-invariant $\}/\mathbb{R}_{>0}$

Normal distributions:
$\Omega(G, H) = \{dx\}$. Here dx is the Lebesgue measure, which is relatively G-invariant.

G/H-method Step 4: measure $\tilde{p}_\theta$ on X

Representation theory:
Define mesures on X parameterized by $(\xi, \mu) \in V^\vee \times \Omega(G, H)$ by

$$d\tilde{p}_\theta(x) = d\tilde{p}_{\xi,\mu}(x) := \exp(-\langle \xi, x v_0 \rangle) d\mu(x)$$

Normal distributions:
By taking an inner product on $V = \mathrm{Sym}(2, \mathbb{R})$, we obtain

$$d\tilde{p}_\theta(x) = \exp(-(\theta_1 x^2 + 2\theta_2 x + \theta_3))dx.$$

G/H-method Step 5: normalizing the measure $\tilde{p}_\theta$

Representation theory:
$\Theta := \{\theta = (\xi, \mu) \in V^\vee \times \Omega(G, H) \mid \int_X d\tilde{p}_\theta < \infty\}$
$p_\theta := e^{-\varphi(\theta)} \tilde{p}_\theta$,
$\varphi(\theta) := \log \int_X d\tilde{p}_\theta$ (log normalizer)
We obtain a family of probability measures $\{p_\theta\}_{\theta \in \Theta}$.

Normal distributions:
$\Theta = \{\theta = \begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_3 \end{pmatrix} \in \mathrm{Sym}(2, \mathbb{R}) \mid \theta_1 > 0\}$
$dp_\theta = \sqrt{\frac{\theta_1}{\pi}} \exp(-\theta_1(x + \frac{\theta_2}{\theta_1})^2) dx$
$\varphi(\theta) = \frac{\theta_2^2 - \theta_1\theta_3}{\theta_1} + \frac{1}{2} \log \frac{\pi}{\theta_1}$

G/H-method Step 5: Normal distributions

We obtain a family of probability measures on $\mathbb{R}$

$$\left\{ \sqrt{\frac{\theta_1}{\pi}} \exp\left(-\theta_1 \left(x + \frac{\theta_2}{\theta_1} \right)^2 \right) dx \right\}_{(\theta_1, \theta_2) \in \mathbb{R}_{>0} \times \mathbb{R}}.$$

By a change of variables

$$m = -\frac{\theta_2}{\theta_1}, \quad \sigma = \frac{1}{\sqrt{2\theta_1}},$$

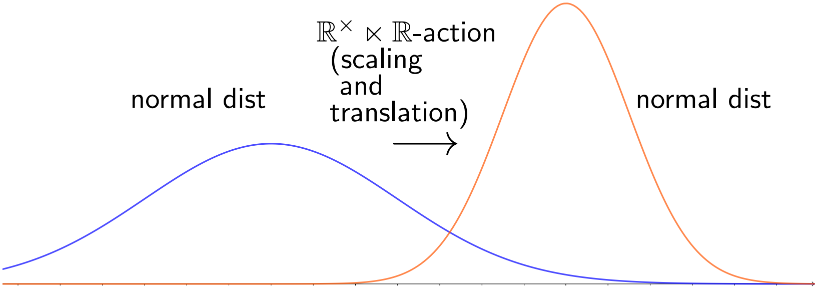
we obtain the family of normal distributions

$$\left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2} \right) dx \right\}_{(\sigma, m) \in \mathbb{R}_{>0} \times \mathbb{R}}.$$

$\mathcal{P}$ is a G-invariant exponential family

Theorem 2.4.

Any family obtained by our method is a G-invariant exponential family on G/H.



We obtain a family with the symmetry of G/H !

Question

Conversely,

Question 2.5.

Are any G-invariant exponential families on G/H obtained by our method?

$\rightsquigarrow$ Yes, under a mild assumption.

$\rightsquigarrow$ Roughly speaking,

$$\begin{aligned} &\{G\text{-invariant exponential family on } G/H\} \\ \text{``}=\text{''}&\{\text{families obtained by } G/H\text{-method}\} \end{aligned}$$

Answer to the question

Setting 2.6.

$\mathcal{P} := \{p_\theta\}_{\theta\in\Theta}$ is a G-invariant exponential family on G/H. Here Θ is the parameter space.

Theorem 2.7.

Assume

1. *G/H admits a nonzero relatively G-invariant measure,*
2. *Θ is open.*

Then, $\mathcal{P}$ is a subfamily of a certain family obtained by G/H-method.

For the details, see our paper that will appear in Information geometry, Affine Differential Geometry and Hesse Geometry: A Tribute and Memorial to Jean-Louis Koszul.

Classification problem of G-invariant exponential families

Let us consider an important homogeneous space G/H such as a sphere and a hyperbolic space, more generally symmetric spaces.

Aim

One of our aims is to classify "good" exponential families on G/H.

Problem 2.8.

Classify G-invariant exponential families on G/H.

By Theorem 2.7, this problem above is reduced to the following:

Question 2.9.

Classify families obtained by G/H-method on G/H.

First step for the classification

Question 2.10.

When do distinct pairs (V, v_0) and (V', v_0') generate the same family?

$\rightsquigarrow$ equivalence relation on the pairs of a rep. and an H-fixed vector.

Proposition 2.11.

Equivalent elements in $\tilde{\mathcal{V}}(G, H)$ generate the same family by our method.

Definition 2.12.

$$\tilde{\mathcal{V}}(G,H) := \{(V, v_0) \mid V \text{ is a fin. dim. real rep. of } G,\ v_0 \in V^H \text{ is cyclic}\}$$

Here $v_0 \in V$ is said to be cyclic if

$$\mathrm{span}\{g \cdot v_0 \mid g \in G\} = V.$$

Definition 2.13.

$(V, v_0) \sim (V', v_0') \overset{\text{def}}{\iff} {}^{\exists}\varphi \colon V \to V'$ G-linear isomorphism satisfying $\varphi(v_0) = v_0'$.

For the details, see "On a method to construct exponential families by representation theory" Geometric Science of Information. GSI2019, Lecture Notes in Computer Science, vol 11712, 147–156 (2019)

Second step for the classification

Question 2.14.

Determine the equivalence classes $\mathcal{V}(G,H) := \tilde{\mathcal{V}}(G,H)/\sim$.

We are writing a paper about this question.

Message

Aim

We want to suggest new useful families of distributions for applications by using G/H-method.

Take-home message

If you want a good family of distributions on some spaces, please tell us the spaces! We will try to propose good families of distributions on them.

Examples

1. Sphere S^n:
   - von Mises–Fisher distributions,
   - Fisher–Bingham distributions.
2. Upper half plane $\mathcal{H}$
3. Hyperbolic space H^n:
   - hyperboloid distributions.

The distributions having names above were obtained heuristically.

Remark 3.1.

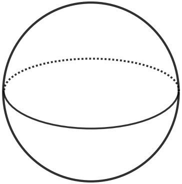
Machine learning using hyperbolic space H^n has been very active. Good exponential families on H^n has been desired.

$$H^n := \{x \in \mathbb{R}^{n+1} \mid x_1^2 + x_2^2 + \cdots + x_n^2 - x_{n+1}^2 = -1\}.$$

Example 1: Sphere

n-dimensional sphere:

$$S^n := \{x \in \mathbb{R}^{n+1} \mid x_1^2 + \cdots + x_{n+1}^2 = 1\}.$$



$G = SO(n+1)$, $H = SO(n)$, $X := G/H \simeq S^n$.

Low dimensional representations:

- $\rho : SO(n+1) \to GL(\mathbb{R}^{n+1})$ natural representation
  $\rightsquigarrow$ von Mises–Fisher distributions
- $\rho : SO(n+1) \to GL(\mathbb{R}^{n+1} \oplus \mathrm{Sym}(n+1, \mathbb{R}))$
  $\rho(g)(v, S) = (gv, gS{}^tg)$ $((v, S) \in \mathbb{R}^{n+1} \times \mathrm{Sym}(n+1, \mathbb{R}))$.
  $\rightsquigarrow$ Fisher–Bingham distributions

Higher dimensional representations:

We obtain families on S^n by G/H-method, which are not so well-known now, but appeared in the research by T. S. Cohen and M. Welling [CW15].

Related work: [CW15]

[CW15] T. S. Cohen, M. Welling, "Harmonic exponential families on manifolds" In Proceedings of the 32nd International Conference on Machine Learning(ICML), volume 37 (2015), 1757–1765

1. suggest "harmonic exponential families" on compact group or cpt homog. sp. such as $S^1 \simeq SO(2)$ and $S^2 \simeq SO(3)/SO(2)$.
2. use Fast Fourier Transform (FFT) on S^1, S^2 to perform gradient-based optimization of log-likelihood for MLE.
3. apply them to the problem of modelling the spatial distribution of significant earthquakes on the surface of the earth.

Relation between G/H-method and [CW15]

The method in [CW15]

[CW15] give a method to construct exponential families on G/H by using

- unitary representation of compact Lie group G,
- G-invariant measure on G/H.

Remark 3.2.

The method in [CW15] is the case where G is compact of G/H-method.

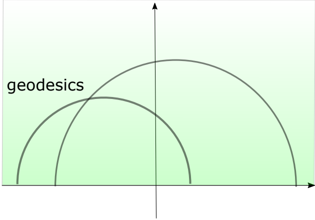
Fact 3.3.

In the case where G is compact, relatively G-invariant measure is automatically G-invariant measure.

Example 2: Upper half plane

Upper half plane $\mathcal{H} := \{z = x + iy \in \mathbb{C} \mid y > 0\}$ admits the linear fractional transformation of $SL(2,\mathbb{R})$.

$\rightsquigarrow G = SL(2,\mathbb{R})$, $H = SO(2)$, $X := G/H \simeq \mathcal{H}$.



- Low dimensional representation:
  $\rho : SL(2,\mathbb{R}) \to GL(\mathrm{Sym}(2,\mathbb{R}))$,
  $\rho(g)S = gS^tg \quad (S \in \mathrm{Sym}(2,\mathbb{R}))$.

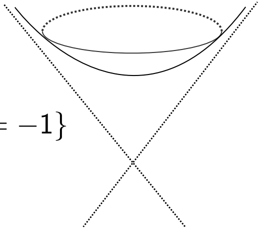
$$\rightsquigarrow \left\{ \frac{De^{2D}}{\pi} \exp\left(-\frac{a(x^2+y^2)+2bx+c}{y} \right) \frac{dxdy}{y^2} \right\}_{\begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathrm{Sym}^+(2,\mathbb{R})}$$

  Here $D = \sqrt{ac - b^2}$.

- Higher dimensional cases:
  We obtain new families by G/H-method.

This is a special case of hyperbolic space.

Example 3: Hyperbolic space



n-dim hyperbolic space:

$$H^n := \{x \in \mathbb{R}^{n+1} \mid x_1^2 + \cdots + x_n^2 - x_{n+1}^2 = -1\}$$

$G = SO_0(n,1)$, $H = SO(n)$, $X := G/H \simeq H^n$

- Low dimensional representation:
  $\rho : G \to GL(\mathbb{R}^{n+1})$ natural representation
  $\rightsquigarrow$ Hyperboloid distributions
  (O. E. Barndorff-Nielsen [BN78], J. L. Jensen [J81])
- Higher dimensional cases:
  We obtain new families by G/H-method.

References

[BN78] O. E. Barndorff-Nielsen, *Hyperbolic distributions and distribution on hyperbolae*, Scand. J. Stat. **8** (1978), 151–157.

[CW15] T. S. Cohen, M. Welling, *Harmonic exponential families on manifolds*, In Proc. of the 32nd Inter. Conf. on Machine Learning(ICML), volume 37 (2015), 1757–1765

[J81] J. L. Jensen, *On the hyperboloid distribution*, Scand. J. Statist. **8** (1981), 193–206.

[TY18] K. Tojo, T. Yoshino, *A method to construct exponential families by representation theory*, arXiv:1811.01394v3.

[TY19] K. Tojo, T. Yoshino, *On a method to construct exponential families by representation theory*, GSI2019, Lecture Notes in Computer Science, vol 11712, 147–156 (2019).

[TY20] K. Tojo, T. Yoshino, *Harmonic exponential families on homogeneous spaces*, Information geometry, Springer, to appear.