Statistical Inference and the Parallel Transport of Probability

Dalton A R Sakthivadivel

Based on 2203.08119 (+ 2204.05084, 2205.11543)

5th June 2022

Stony Brook University

Maximum entropy

Consider a probability density p over some space X satisfying the diffusion process

$$\frac{\partial}{\partial t} p = -\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x} J p\right) + D \frac{\partial^2}{\partial x^2} p$$

with $D = \text{const}$ and $J : X \to \mathbb{R}$ a measurable function.

Maximum entropy

Consider a probability density p over some space X satisfying the diffusion process

$$\frac{\partial}{\partial t} p = -\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x} J p\right) + D \frac{\partial^2}{\partial x^2} p$$

with $D = \text{const}$ and $J : X \to \mathbb{R}$ a measurable function.

When $\partial_t p = 0$,

Maximum entropy

Consider a probability density p over some space X satisfying the diffusion process

$$\frac{\partial}{\partial t}p = -\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}Jp\right) + D\frac{\partial^2}{\partial x^2}p$$

with $D = \text{const}$ and $J : X \to \mathbb{R}$ a measurable function.

When $\partial_t p = 0$,

$$p = \arg\max_p \left[-\int p \ln p - \lambda\left(\int Jp + C\right)\right]$$

via integration by parts.

Maximum entropy

Consider a probability density p over some space X satisfying the diffusion process

$$\frac{\partial}{\partial t} p = -\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x} J p\right) + D \frac{\partial^2}{\partial x^2} p$$

with $D = \text{const}$ and $J : X \to \mathbb{R}$ a measurable function.

When $\partial_t p = 0$,

$$p = \arg\max_p \left[-\int p \ln p - \lambda \left(\int J p + C \right) \right]$$

via integration by parts.

Discussed by Jordan, Kinderlehrer, and Otto (1998); Markowich and Villani (2000)

Gradient ascent

We are interested in the Euler-Lagrange equation

Gradient ascent

We are interested in the Euler-Lagrange equation

Entropy is a functional of the form

$$S[p] = \int \mathcal{L}(p(x))\,\mathrm{d}x$$

where $\mathcal{L} = -p \ln p - Jp$.

Gradient ascent

We are interested in the Euler-Lagrange equation

Entropy is a functional of the form

$$S[p] = \int \mathcal{L}(p(x))\,\mathrm{d}x$$

where $\mathcal{L} = -p \ln p - Jp$.

Gradient ascent is given by

$$\begin{aligned} \frac{\partial}{\partial p}\big(- p \ln p - \lambda J p\big) &= 0 \\ - \ln p - \lambda J &= 0. \end{aligned}$$

Gradient ascent

We are interested in the Euler-Lagrange equation

Entropy is a functional of the form

$$S[p] = \int \mathcal{L}(p(x))\,\mathrm{d}x$$

where $\mathcal{L} = -p \ln p - Jp$.

Gradient ascent is given by

$$\begin{aligned} \frac{\partial}{\partial p}\big(- p \ln p - \lambda J p \big) &= 0 \\ - \ln p - \lambda J &= 0. \end{aligned}$$

Stationary solution is the desired Gibbs measure, $\exp\{-\lambda J\}$.

The role of J in maximising entropy

The function J in our diffusion process can be interpreted as a penalty on states, since $p(x) \propto -J(x)$

The role of J in maximising entropy

The function J in our diffusion process can be interpreted as a penalty on states, since $p(x) \propto -J(x)$

We can imagine a probability density as a surface

$$x \mapsto (x, p(x)),$$

and therefore as a section s of a line bundle $E \xrightarrow{\pi} X$ with typical fibre $\mathbb{R}_{>0}$.

The role of J in maximising entropy

The function J in our diffusion process can be interpreted as a penalty on states, since $p(x) \propto -J(x)$

We can imagine a probability density as a surface

$$x \mapsto (x, p(x)),$$

and therefore as a section s of a line bundle $E \xrightarrow{\pi} X$ with typical fibre $\mathbb{R}_{>0}$.

The role of J in maximising entropy

The function J in our diffusion process can be interpreted as a penalty on states, since $p(x) \propto -J(x)$

We can imagine a probability density as a surface

$$x \mapsto (x, p(x)),$$

and therefore as a section s of a line bundle $E \xrightarrow{\pi} X$ with typical fibre $\mathbb{R}_{>0}$. Hence J is also a constraint on the shape of the graph of s.

A quick word on parallel transport

Question

Given a path in a base space $\varphi : x_0 \to x_1$ and a constraint on $(x, p(x))$, what is $s(\varphi)$?

A quick word on parallel transport

Question

Given a path in a base space $\varphi : x_0 \to x_1$ *and a constraint on* $(x, p(x))$*, what is* $s(\varphi)$*?*

A *horizontal lift* of such a φ should lie on $(x, p(x))$

A quick word on parallel transport

Question

Given a path in a base space $\varphi : x_0 \to x_1$ and a constraint on $(x, p(x))$, what is $s(\varphi)$?

A *horizontal lift* of such a φ should lie on $(x, p(x))$

- Parallel transport is like a constraint on motion of $p(x_0)$ across to $p(x_1)$. Parallel graphs or 'horizontal lifts' are constant with respect to some penalty J

A quick word on parallel transport

Question

Given a path in a base space $\varphi : x_0 \to x_1$ and a constraint on $(x, p(x))$, what is $s(\varphi)$?

A *horizontal lift* of such a φ should lie on $(x, p(x))$

- Parallel transport is like a constraint on motion of $p(x_0)$ across to $p(x_1)$. Parallel graphs or 'horizontal lifts' are constant with respect to some penalty J
- Loci of equiprobable states $\exp\{-J\} = q$ are precisely level sets $J(x) = c$; probability of those states is given by the lift of those level sets

A quick word on parallel transport

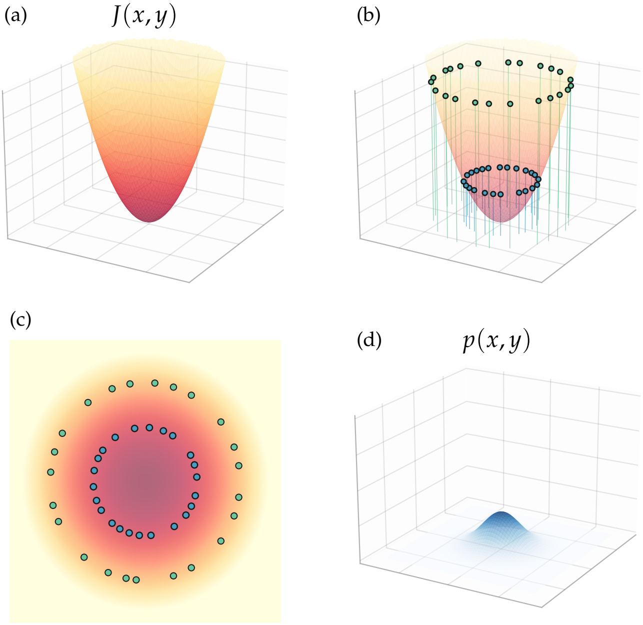
Question

Given a path in a base space $\varphi : x_0 \to x_1$ and a constraint on $(x, p(x))$, what is $s(\varphi)$?

A *horizontal lift* of such a φ should lie on $(x, p(x))$

- Parallel transport is like a constraint on motion of $p(x_0)$ across to $p(x_1)$. Parallel graphs or 'horizontal lifts' are constant with respect to some penalty J
- Loci of equiprobable states $\exp\{-J\} = q$ are precisely level sets $J(x) = c$; probability of those states is given by the lift of those level sets

So s should consist of parallel transport lines



Adapted from arXiv:2205.11543. Credit to Brennan Klein.

The stationary solution to CME is parallel transport

Recall:

$$\ln p(x) = -\lambda J(x)$$

maximises entropy subject to constraint $\mathbb{E}[J] = C$. A brief formal computation has:

The stationary solution to CME is parallel transport

Recall:

$$\ln p(x) = -\lambda J(x)$$

maximises entropy subject to constraint $\mathbb{E}[J] = C$. A brief formal computation has:

$$\begin{aligned}
\mathrm{d} \ln p(x) &= -\lambda \, \mathrm{d}J(x) \\
\frac{\partial}{\partial x} p(x) \, \mathrm{d}x &= -\lambda \frac{\partial}{\partial x} J(x) \, \mathrm{d}x \, p(x) \\
\mathrm{d}p(x) &= -\, \mathrm{d}J(x) \, p(x)
\end{aligned}$$

The stationary solution to CME is parallel transport

Recall:

$$\ln p(x) = -\lambda J(x)$$

maximises entropy subject to constraint $\mathbb{E}[J] = C$. A brief formal computation has:

$$\begin{aligned}
\mathrm{d}\ln p(x) &= -\lambda\,\mathrm{d}J(x) \\
\frac{\partial}{\partial x} p(x)\,\mathrm{d}x &= -\lambda \frac{\partial}{\partial x} J(x)\,\mathrm{d}x\, p(x) \\
\mathrm{d}p(x) &= -\,\mathrm{d}J(x)\, p(x)
\end{aligned}$$

$\therefore$ maximising entropy yields the solution to parallel transport with connection valued in $\mathbb{R}$.

A quick sanity check

Stationary solution to constrained maximum entropy:
$p(x) = \exp\{-\lambda J(x)\}$

A quick sanity check

Stationary solution to constrained maximum entropy:
$p(x) = \exp\{-\lambda J(x)\}$

Solution to covariant differentiation of a section s in connection $\mathrm{d}J(x)$: $s(x) = \exp\left\{-\lambda \int_k^x \mathrm{d}J(\tilde{x})\right\} = \exp\{-\lambda J(x)\}$

A first idea of gauge theory

Suppose S is a functional on the space of p's, $\Gamma(E)$ and let G be a set. A *gauge* is a quantity g in some Lie group G such that

$$S(\rho(g)\Gamma) = 0$$

for all $g \in G$. A choice of gauge is a choice of one such g. Our $\rho(g)$ is a G-valued transition function t.

A first idea of gauge theory

Suppose S is a functional on the space of p's, $\Gamma(E)$ and let G be a set. A *gauge* is a quantity g in some Lie group G such that

$$S(\rho(g)\Gamma) = 0$$

for all $g \in G$. A choice of gauge is a choice of one such g. Our $\rho(g)$ is a G-valued transition function t.

A gauge is also called a frame

A first idea of gauge theory

Suppose S is a functional on the space of p's, $\Gamma(E)$ and let G be a set. A *gauge* is a quantity g in some Lie group G such that

$$S(\rho(g)\Gamma) = 0$$

for all $g \in G$. A choice of gauge is a choice of one such g. Our $\rho(g)$ is a G-valued transition function t.

A gauge is also called a frame

Motivation: changes of frame on TX, where

$$V = V^i \frac{\partial}{\partial x^i} \text{ or } \tilde{V}^i \frac{\partial}{\partial y^i}$$

related by the Jacobian matrix

$$\tilde{V}^j = \frac{\partial y^j}{\partial x^i} V^i$$

A gauge-theoretic interpretation

Define a space of Gibbs measures where a frame therein is given by $\exp\{-J\}$ for some $J : X \to \mathbb{R}$.

A gauge-theoretic interpretation

Define a space of Gibbs measures where a frame therein is given by $\exp\{-J\}$ for some $J : X \to \mathbb{R}$.

What does a change in frame do to the constrained entropy functional?

A gauge-theoretic interpretation

Define a space of Gibbs measures where a frame therein is given by $\exp\{-J\}$ for some $J : X \to \mathbb{R}$.

What does a change in frame do to the constrained entropy functional?

Under $p \mapsto e^{-J'} p$, $J \mapsto J' + J$,

$$\mathcal{L} = -e^{-J} p \ln e^{-J} p - (J' + J) e^{-J} p$$
$$= e^{-J} p(- \ln p - J).$$

A gauge-theoretic interpretation

Define a space of Gibbs measures where a frame therein is given by $\exp\{-J\}$ for some $J : X \to \mathbb{R}$.

What does a change in frame do to the constrained entropy functional?

Under $p \mapsto e^{-J'}p$, $J \mapsto J' + J$,

$$\mathcal{L} = -e^{-J}p \ln e^{-J}p - (J' + J)e^{-J}p$$

$$= e^{-J}p(-\ln p - J).$$

What is the variation? Root of

$$e^{-J}(-\ln p - J)$$

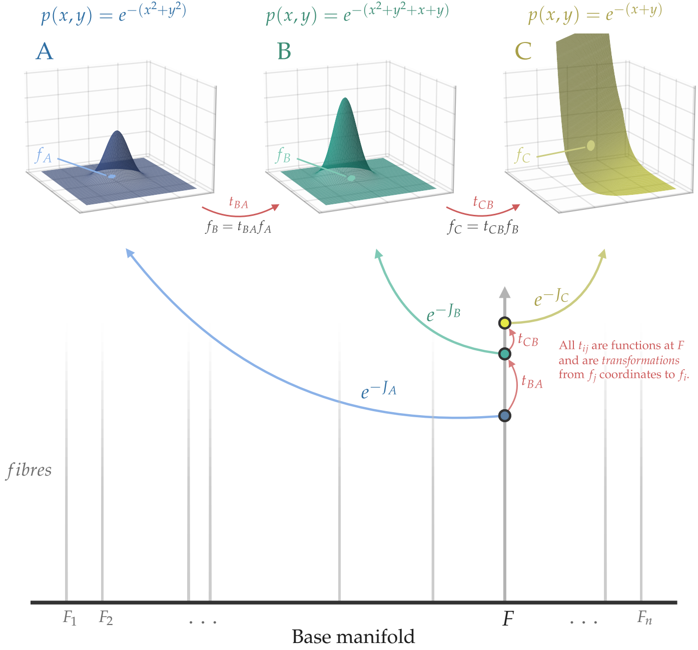
is equal to the root of

$$-\ln p - J.$$

Gibbs frames

Theorem

Let P be the principal bundle associated to E. Since level sets of p are constant with respect to J, there exists some group element $\exp\{J\}$ *for p whose logarithmic derivative is $\mathrm{d}J$; moreover, there exists a principal bundle of such Gibbs frames, which is P.*



Adapted from arXiv:2205.11543. Credit to Brennan Klein.

- The entropy functional appears to describe the parallel transport of probability across the state space

- The entropy functional appears to describe the parallel transport of probability across the state space
- The notion of what is ‘probabilistically parallel’ is given by constraints

- The entropy functional appears to describe the parallel transport of probability across the state space
- The notion of what is ‘probabilistically parallel’ is given by constraints
- In this case, parallel motion of probability over data is literally constrained to lie in a particular vector field, the distribution of $\mathrm{d}J$

- The entropy functional appears to describe the parallel transport of probability across the state space
- The notion of what is ‘probabilistically parallel’ is given by constraints
- In this case, parallel motion of probability over data is literally constrained to lie in a particular vector field, the distribution of $\mathrm{d}J$
- Change in Gibbs frame changes the expression of an assignment in probability

- The entropy functional appears to describe the parallel transport of probability across the state space
- The notion of what is 'probabilistically parallel' is given by constraints
- In this case, parallel motion of probability over data is literally constrained to lie in a particular vector field, the distribution of $\mathrm{d}J$
- Change in Gibbs frame changes the expression of an assignment in probability
- Vertical $\oplus$ horizontal flows