情報幾何勉強会第一回でいただいた質問と回答 · 補足

科学計算総合研究所 谷村慈則

質問 1

P から Q への f-divergence の定義は Wikipedia などを見ると

$$D_f(P\|Q) = \int_\Omega f\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)\mathrm{d}Q$$

になっていて、スライドの

$$D_f(Q\|P) = \int_\Omega f\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)\mathrm{d}Q$$

食い違っている。また、「P から Q への」という語感からは $P \ll Q$ ではなく $P \gg Q$ であるべき。

回答　勉強会中はご指摘いただいた通りに修正すると申し上げましたが、後日調べたところ P, Q の順番についてはいくつか流派があるようです。Wikipedia とそこで参照されている Csiszár の論文ではご指摘いただいた通りの順番なのですが、私が普段参照している以下の文献ではスライドの通りの順番になっていました。

- 甘利 俊一「情報幾何学の新展開」サイエンス社
- Shun-ichi Amari, “Information Geometry and Its Applications,” Springer
- N.Ay, J.Jost, H..Lê and L.Schwachhöfer, “Information Geometry,” Springer

そこで、当勉強会では今後、以下のような対応を取ることにします。

- 私が普段参照している文献とノーテーションがズレるとややこしいので、申し訳ありませんが当勉強会では $D_f(Q\|P)$ の方のノーテーションを取ることにします。
- 「P から Q への」という言葉を使うとややこしくなるので、この言葉遣いは廃止します。実際、上記の 3 つの文献ではこの言葉遣いは使用していません。
- 以下の理由から、仮定を強めて「P と Q は互いに絶対連続」とします。
  - 次回以降、互いに絶対連続でない確率測度の f-divergence を取る機会がない。
  - $f(0) = \infty$ と評価することにすると、$P \gg Q$ でなければ $D_f(Q\|P)$ は定義できても発散するため、考えてもしょうがない（絶対連続性の十分条件が得られるため意味のない考察ではないが、当勉強会の本筋から外れるため、あまり深入りしない）。

質問 2

情報幾何では幾何学として何を扱うのか、また、統計学的に何ができるのか。

回答　まず最初に、私自身が勉強中であるため、自信を持ってこれができるという回答はいまのところできません。その前提の下で、現在私が考えている内容をお答えします。

幾何学としては双対接続の幾何学を扱います。Riemann 計量が中心的な役割を果たしますが、接続は Levi-Civita 接続ではなく双対接続を扱います。Levi-Civita 接続ではなく双対接続を扱うのは、確率論の逆問題たる統計学において、我々がこれから扱う評価関数が f-divergence である

ことに起因します。f-divergence は距離と違って対称律を満たさないため、非対称性が双対接続という形になって現れてくるわけです。

情報幾何について私がいまのところ期待している効用としては、座標変換に依らない量の切り出しが挙げられます。これは単に数学的にきれいというだけでなく、具体的に座標を取った際に計算を楽にする効果もあります。例えば、ユークリッド空間における体積要素の極座標表示を考える際には、毎回一から計算をするより、一般の Riemann 多様体での公式

$$v = \sqrt{|G|}\, \mathrm{d}x^0 \wedge \cdots \wedge \mathrm{d}x^{n-1}$$

を用いたほうが速いです。情報幾何を用いれば、統計的な考察においても同様に、色々な計算を楽にすることができるのではないかと期待しています。

質問 3

Wasserstein 距離との関連性について

回答　Wasserstein 距離は互いに絶対連続ではない確率測度同士を測る指標として自然であるのは確かで、実際に画像処理や 3DCG の文脈で用いられています。ただし、Wasserstein 距離は Riemann 幾何学の最適輸送理論から出てきた概念なので、統計学的なモチベーションから考え出された概念ではありません。Wasserstein 距離を用いるにあたっては、以下の点で注意が必要です。

- Wasserstein 距離を用いるときには、可測空間 $(\Omega, \mathcal{F})$ も距離空間であることが要請される。
- Wasserstein 距離そのものが最適化問題の解として定式化されるため、与えられた 2 点間の距離を計算するだけでも工夫が必要。

質問 4

平均エントロピーと相対エントロピーの関係性

回答　$\Omega = \tilde{n} = \{0, 1, \ldots, n-1\}$ の場合に限定して考えると、以下の命題が成り立ちます。

命題 5

$\tilde{n}$ の一様分布を P_0 とおく。P は $\tilde{n}$ 上の確率測度で $P(\{i\}) > 0$ $(i \in \tilde{n})$ を満たすとする。
このとき、$H(P_0) - H(P) = D_{KL}(P \| P_0)$ が成り立つ。

もう少し大胆な言い方をすると、

平均エントロピーは数え上げ測度や Lebesgue 測度との近さを測る指標である

ということができます。