2018 年 第 2 問

ジャンル: 確率
難易度: Hard

> **📖 問題の概要**
> 確率密度を、累積分布関数の表現で考える問題。慣れると簡単ですが、最初は難しいかもしれません。 最後の距離はワッサースタイン計量と呼ばれるものの一種です。

(1)

KL-divergence の定義は覚えておこう。

$$
\begin{aligned}
D(P_1 \mid\mid P_2) &= \int_{-\infty}^{\infty} p_1(x) \log\left(\sqrt{\frac{\sigma_2^2}{\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{(x-\mu_2)^2}{2\sigma_2^2}\right)\right) \mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{2} \log\left(\frac{\sigma_2^2}{\sigma_1^2}\right) p_1(x) + \left(-\frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{(x-\mu_2)^2}{2\sigma_2^2}\right) p_1(x) \, \mathrm{d}x \\
&= E\left[\frac{1}{2} \log\left(\frac{\sigma_2^2}{\sigma_1^2}\right) - \frac{(x-\mu_1)^2}{2\sigma_1^2} + \frac{(x-\mu_2)^2}{2\sigma_2^2}\right] \\
&= \frac{1}{2} \log\left(\frac{\sigma_2^2}{\sigma_1^2}\right) - \frac{1}{2} + \frac{1}{2\sigma_2^2}(\sigma_1^2 + \mu_1^2) - \frac{1}{2\sigma_2^2} 2\mu_2\mu_1 + \frac{1}{2\sigma_2^2}\mu_2^2 \\
&= \frac{1}{2}\left(-1 + \log\left(\frac{\sigma_2^2}{\sigma_1^2}\right) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{\sigma_2^2}\right)
\end{aligned}
$$

(2)

X, Y は非負の確率変数であり、有限な二次モーメントを持つ。 これは、$P(X \geq \infty) = 0$ を意味することに注意。後々、部分積分の際に使う。

(2-1)

連続な密度関数を持つなら、密度関数は分布関数の微分であることを使う。つまり、$p(x, y) = \frac{\partial^2}{\partial x \partial y} P(\{X \leq x\} \cap \{Y \leq y\})$。

$$
\begin{aligned}
E[XY] &= \int_0^\infty \int_0^\infty xy\, p(x,y) \,\mathrm{d}x \,\mathrm{d}y \\
&= \int_0^\infty \int_0^\infty xy \frac{\partial^2}{\partial x \partial y} P(\{Y \leq y\} - P(\{X \geq x\} \cap \{Y \leq y\})) \,\mathrm{d}x \,\mathrm{d}y \\
&= \int_0^\infty \int_0^\infty -xy \frac{\partial^2}{\partial x \partial y} P(\{X \geq x\} \cap \{Y \leq y\}) \,\mathrm{d}x \,\mathrm{d}y \\
&= \int_0^\infty y \int_0^\infty -x \frac{\partial^2}{\partial x \partial y} P(\{X \geq x\} \cap \{Y \leq y\}) \,\mathrm{d}x \,\mathrm{d}y \\
&= \int_0^\infty y \int_0^\infty \frac{\partial}{\partial y} P(\{X \geq x\} \cap \{Y \leq y\}) \,\mathrm{d}x \,\mathrm{d}y \\
&= \int_0^\infty \int_0^\infty y \frac{\partial}{\partial y} (P(X \geq x) - P(\{X \geq x\} \cap \{Y \geq y\})) \,\mathrm{d}y \,\mathrm{d}x \\
&= \int_0^\infty \int_0^\infty P(\{X \geq x\} \cap \{Y \geq y\}) \,\mathrm{d}y \,\mathrm{d}x
\end{aligned}
$$

と、部分積分を駆使することで示すことができる。あまりに大変である。しかも、これは同時確率分布 P_{XY} が連続であるという仮定が必要である。

ので、もう少し良い方法を紹介する。ルベーグ・スティルチェス積分を使う。ルベーグ・スティルチェス積分とは、$\int_0^\infty g(x) \,\mathrm{d}F(x)$ のような記法で書かれることが多いルベーグ積分の拡張で、確率論を勉強しているときに一度は目にしたことがあるかもしれない。これの素晴らしい点は、連続確率変数と離散確率変数の両方を同時に扱えることである。したがって、不連続な確率密度関数についても扱える (ほんと？ ルベーグの分解定理より、絶対連続、特異連続な分布、離散分布に一意に分解できて、和の積分を積分の和にできる性質があるので、それで積分したらいいので良さそう)。さて、いくつかの準備の後ルベーグ・スティルチェス積分の練習をしてみよう。

Theorem 1 (Jensen の不等式): f は凸関数であり、X は確率変数であるとする。このとき、$E[f(X)] \geq f(E[X])$ が成り立つ。

これの重要な系として、以下が成り立つ。

Corollary 1: $E[X^2] \geq E[|X|]^2$

Proof: Jensen の不等式で $f(x) = x^2$ として、$|X|^2 = X^2$ より。 □

Theorem 2 (フビニ・トネリの定理): 絶対値を付けた累次積分のどちらかが計算でき、有限の値ならば、絶対値を外した重積分と累次積分は有限の値で一致する。

問題文の条件より、$E[X^2] < \infty$ なので、$E[|X|] < \infty$ がわかった。Y についても同様。 よって、$E[|XY|] \leq E[|X|]E[|Y|] < \infty$ が分かったので、フビニ・トネリの定理より重積分は入れ替えてよい。これを後で使う。

まずは練習で、有限な二次モーメントを持つ非負変数 X に対する $E[X]$ をスティルチェス積分で計算してみよう。

$$\begin{aligned}E[X] &= \int_0^\infty x\,\mathrm{d}F(x)\\ &= \int_0^\infty \int_0^x \mathrm{d}y\,\mathrm{d}F(x)\\ &\text{ここで重積分の変数変換を行う}\\ &= \int_0^\infty \int_y^\infty \mathrm{d}F(x)\,\mathrm{d}y\\ &= \int_0^\infty P(X \geq y)\,\mathrm{d}y\end{aligned}$$

さて、これを参考にして有限な二次モーメントを持つ非負変数 X, Y に対して $E[XY]$ を計算してみよう。

$$\begin{aligned}E[XY] &= \int_0^\infty \int_0^\infty xy\ \mathrm{d}F(x,y)\\ &= \int_0^\infty \int_0^\infty \int_0^x \int_0^y \mathrm{d}z\,\mathrm{d}w\,\mathrm{d}F(x,y)\\ &\text{ここで重積分の変数変換を行う}\\ &= \int_0^\infty \int_0^\infty \int_z^\infty \int_w^\infty \mathrm{d}F(x,y)\,\mathrm{d}z\,\mathrm{d}w\\ &= \int_0^\infty \int_0^\infty P(X \geq z, Y \geq w)\,\mathrm{d}z\,\mathrm{d}w\end{aligned}$$

このようにして、一般的かつ簡潔に示すことができた。知っておくと便利かもしれない。

(2-2)

$E[X^2] = \int\int x^2\,\mathrm{d}F(x,y) = \int x^2\,\mathrm{d}F(x) = \int_0^1 \left(F^{-1}(U)\right)^2 \mathrm{d}U$ ここで、最後は $U = F(x)$ と変数変換した。 初等的に書くと、$E[X^2] = \int\int P_{XY} x^2\,\mathrm{d}x\,\mathrm{d}y = \int P_1 x^2\,\mathrm{d}x = \int \left(F^{-1}(U)\right)^2 \mathrm{d}U$ となる。

上を踏まえたうえで式を展開すると、$E[XY] \leq E[F^{-1}(U)G^{-1}(G)]$ を示せばよいことがわかる。この式は理解しやすい。非負変数の、積の期待値の代わりに、値が大きいところ同士をかけて積分したほうが積分値が大きくなるという自然なことを意味している。

(2-1) を使って示す方針でいく。正直この問題はかなり難しい。

$$
\begin{aligned}
E[F^{-1}(U)G^{-1}(U)] &= \int\int P\{F^{-1}(U) \geq x, G^{-1}(U) \geq y\}\ \mathrm{d}x\,\mathrm{d}y \\
&= \int\int P\{U \geq F(x), U \geq G(y)\}\ \mathrm{d}x\,\mathrm{d}y \\
&= \int\int P\{U \geq \max(F(x), G(y))\}\ \mathrm{d}x\,\mathrm{d}y \\
&= \int\int 1 - \max(F(x), G(y))\ \mathrm{d}x\,\mathrm{d}y \\
&= \int\int \min(1 - F(x), 1 - G(y))\ \mathrm{d}x\,\mathrm{d}y \\
&= \int\int \min(P(X \geq x), P(Y \geq y))\ \mathrm{d}x\,\mathrm{d}y \\
&= \int\int P(X \geq x, Y \geq y)\ \mathrm{d}x\,\mathrm{d}y \\
&= E[XY]
\end{aligned}
$$

(3)

これは距離にユークリッドノルムを採用した、$p = 2$ の場合のワッサースタイン計量 W_2 らしい。wikipedia に詳しく書いてある。

さて、$E\left[\left(F^{-1}(U) - G^{-1}(U)\right)^2\right]$ を計算したいのだが、$E[F^{-1}(U)G^{-1}(U)]$ の計算をどうすればいいのか困ってしまう。ここで、二つはどちらもガウス分布であることに注目すると、

$$
\frac{F^{-1}(U) - \mu_1}{\sigma_1} = \frac{G^{-1}(U) - \mu_2}{\sigma_2}
$$

なる関係式が成り立つ。二つの分布をそれぞれ標準正規分布に直したら、同じ値になるため従う。 これより、$G^{-1}(U) = \frac{\sigma_2}{\sigma_1}(F^{-1}(U) - \mu_1) + \mu_2$ なので、これを代入して計算すればよい。

$$
\begin{aligned}
W(P_1, P_2)^2 &\leq 2\sigma_2^2 D(P_1 \parallel P_2) \\
\Leftrightarrow (\mu_1 - \mu_2)^2 + (\sigma_1 - \sigma_2)^2 &\leq \sigma_2^2 \left(-1 + \log\left(\frac{\sigma_2^2}{\sigma_1^2}\right) + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{\sigma_2^2} \right)
\end{aligned}
$$

を示していく。 式を整理すると、

$$
\begin{aligned}
(\sigma_1 - \sigma_2)^2 &\leq -\sigma_2^2 + \sigma_1^2 + \sigma_2^2 \log\left(\frac{\sigma_2^2}{\sigma_1^2}\right) \\
\Leftrightarrow 2\sigma_2^2 - 2\sigma_1\sigma_2 &\leq 2\sigma_2^2 \log\left(\frac{\sigma_2}{\sigma_1}\right) \\
\Leftrightarrow \qquad \sigma_2 - \sigma_1 &\leq \sigma_2 \log\left(\frac{\sigma_2}{\sigma_1}\right) \\
\Leftrightarrow \qquad 0 &\leq \frac{\sigma_2}{\sigma_1} \log\left(\frac{\sigma_2}{\sigma_1}\right) - \frac{\sigma_2}{\sigma_1} + 1
\end{aligned}
$$

となる。最後の式で $\frac{\sigma_2}{\sigma_1} = x > 0$ と置いて、$f(x) = x \log x - x + 1$ とすると $f'(x) = \log x$ なので、$f(x)$ は $x = 1$ で最小値を取ることがわかる。$f(1) = 0$ なので、不等式が示された。また、等号は $\frac{\sigma_2}{\sigma_1} = 1 \Leftrightarrow \sigma_2 = \sigma_1$ のときのみである。

ちなみに、ワッサースタイン計量で調べると対数ソボレフノルムとか最適輸送問題とか出てきたので、これは鈴木大慈先生が作問者で間違いなさそう。