Statistics for high and $\infty$–dimensional data

Exercise 1.1

Lasse Vuursteen

February 13, 2018
Exercise 1.1

Let \( f, g \) be positive probability densities with respect to some measure \( \mu \). Show that \( \text{KL}(f, g) \geq 0 \) and that
\( \text{KL}(f, g) = 0 \iff f = g \ \mu\)-almost everywhere. Hint: lower bound the KL-divergence by the squared Hellinger distance.
Exercise 1.1

Let $f, g$ be positive probability densities with respect to some measure $\mu$. Show that $\text{KL}(f, g) \geq 0$ and that $\text{KL}(f, g) = 0 \iff f = g \mu$–almost everywhere. Hint: lower bound the KL-divergence by the squared Hellinger distance.

Solution:
Exercise 1.1

Let $f, g$ be positive probability densities with respect to some measure $\mu$. Show that $\text{KL}(f, g) \geq 0$ and that $\text{KL}(f, g) = 0 \iff f = g \mu$–almost everywhere. Hint: lower bound the KL-divergence by the squared Hellinger distance.

Solution:

Let $\text{KL}(f, g)$ be defined as on page 5. That is,

$$\text{KL}(f, g) := \int f \log \left( \frac{f}{g} \right) d\mu$$
Exercise 1.1

Let $f, g$ be positive probability densities with respect to some measure $\mu$. Show that $\text{KL}(f, g) \geq 0$ and that $\text{KL}(f, g) = 0 \iff f = g \ \mu$–almost everywhere. Hint: lower bound the KL-divergence by the squared Hellinger distance.

Solution:

Let $\text{KL}(f, g)$ be defined as on page 5. That is,

$$\text{KL}(f, g) := \int f \log \left( \frac{f}{g} \right) d\mu$$

This is well defined if $\mu(\{g = 0\}) = 0$. 
Showing $\text{KL}(f, g) \geq 0$

Let the underlying probability space be denoted by $(\Omega, \mathcal{F}, \mu)$. Define the distribution $F(E) := \int_E f \, d\mu$, $E \in \mathcal{F}$. Write

$$\int f \log(fg) \, d\mu = \int \log(fg) \, dF = \int -\log(gf) \, dF \geq -\log \left( \int gf \, d\mu \right) = 0.$$
Showing $\text{KL}(f, g) \geq 0$

Let the underlying probability space be denoted by $(\Omega, \mathcal{F}, \mu)$. Define the distribution $F(E) := \int_E f \, d\mu$, $E \in \mathcal{F}$.

Write

$$\int f \log\left(\frac{f}{g}\right) \, d\mu = \int \log\left(\frac{f}{g}\right) \, dF = \int -\log\left(\frac{g}{f}\right) \, dF$$
Showing $\text{KL}(f, g) \geq 0$

Let the underlying probability space be denoted by $(\Omega, \mathcal{F}, \mu)$. Define the distribution $F(E) := \int_E f \, d\mu$, $E \in \mathcal{F}$.

Write
\[
\int f \log\left(\frac{f}{g}\right) \, d\mu = \int \log\left(\frac{f}{g}\right) \, dF = \int -\log\left(\frac{g}{f}\right) \, dF
\]

Note that the mapping $\mathbb{R} \to \mathbb{R}, \ x \mapsto -\log(x)$ is convex.
Let the underlying probability space be denoted by \((\Omega, \mathcal{F}, \mu)\). Define the distribution \(F(E) := \int_E f \, d\mu, \ E \in \mathcal{F}\).

Write

\[
\int f \log \left( \frac{f}{g} \right) \, d\mu = \int \log (\frac{f}{g}) \, dF = \int -\log (\frac{g}{f}) \, dF
\]

Note that the mapping \(\mathbb{R} \to \mathbb{R}, x \mapsto -\log(x)\) is convex.

By Jensen’s inequality,

\[
\int -\log (\frac{g}{f}) \, dF \geq -\log \int (\frac{g}{f}) \, dF = -\log \int f (\frac{g}{f}) \, d\mu = 0.
\]
Showing $\text{KL}(f, g) \geq 0$

Let the underlying probability space be denoted by $(\Omega, \mathcal{F}, \mu)$. Define the distribution $F(E) := \int_E f \, d\mu$, $E \in \mathcal{F}$.

Write

\[
\int f \log \left( \frac{f}{g} \right) \, d\mu = \int \log \left( \frac{f}{g} \right) \, dF = \int -\log \left( \frac{g}{f} \right) \, dF
\]

Note that the mapping $\mathbb{R} \to \mathbb{R}, x \mapsto -\log(x)$ is convex. By Jensen’s inequality,

\[
\int -\log \left( \frac{g}{f} \right) \, dF \geq -\log \int \left( \frac{g}{f} \right) \, dF = -\log \int f \left( \frac{g}{f} \right) \, d\mu = 0.
\]

Conclude: $\text{KL}(f, g) \geq 0$. 

Showing $\text{KL}(f, g) = 0 \iff f = g \mu$–a.e.

Easy part: "$\iff$" Let $f = g \mu$–a.e.. Then

$$\int f \log \left( \frac{f}{g} \right) d\mu = \int f \log \left( \frac{f}{g} \right) 1_{\{f=g\}} d\mu = 0.$$ 

More difficult: "$\implies$" part. Following the hint, it will be shown that

$$\int f \log \left( \frac{f}{g} \right) d\mu \geq \int (\sqrt{f} - \sqrt{g})^2 d\mu = d_{H^2}(f, g)$$
Showing $\text{KL}(f, g) = 0 \iff f = g \mu$–a.e.

Easy part: ”$\iff$” Let $f = g \mu$–a.e.. Then

$$\int f \log\left(\frac{f}{g}\right) \, d\mu = \int f \log\left(\frac{f}{g}\right) 1_{\{f=g\}} \, d\mu = 0.$$  

More difficult: ”$\Rightarrow$” part. Following the hint, it will be shown that

$$\int f \log\left(\frac{f}{g}\right) \, d\mu \geq \int (\sqrt{f} - \sqrt{g})^2 \, d\mu = d_{H^2}(f, g)$$

d$_{H^2}$(f, g) is the metric called squared Hellinger distance.
Showing $\text{KL}(f, g) \geq d_{H^2}(f, g)$

Use the inequality $-\log(x) \geq 2 - 2\sqrt{x}$, to obtain:

$$\int f \log\left(\frac{f}{g}\right) \, d\mu = \int -f \log\left(\frac{g}{f}\right) \, d\mu \geq \int f \left(2 - 2\sqrt{\frac{g}{f}}\right) \, d\mu$$

$$= 1 + 1 - 2 \int \sqrt{f} \sqrt{g} \, d\mu$$
Showing $\text{KL}(f, g) \geq d_{H^2}(f, g)$

Use the inequality $-\log(x) \geq 2 - 2\sqrt{x}$, to obtain:

$$
\int f \log\left(\frac{f}{g}\right) \, d\mu = \int -f \log\left(\frac{g}{f}\right) \, d\mu \geq \int f \left(2 - 2\sqrt{\frac{g}{f}}\right) \, d\mu
$$

$$
= 1 + 1 - 2 \int \sqrt{f} \sqrt{g} \, d\mu
$$

$$
= \int f + g - 2\sqrt{f} \sqrt{g} \, d\mu
$$
Showing $\text{KL}(f, g) \geq d_{H^2}(f, g)$

Use the inequality $-\log(x) \geq 2 - 2\sqrt{x}$, to obtain:

$$\int f \log\left(\frac{f}{g}\right) \, d\mu = \int -f \log\left(\frac{g}{f}\right) \, d\mu \geq \int f \left(2 - 2\sqrt{\frac{g}{f}}\right) \, d\mu$$

$$= 1 + 1 - 2 \int \sqrt{f} \sqrt{g} \, d\mu$$

$$= \int f + g - 2\sqrt{f} \sqrt{g} \, d\mu$$

$$= \int (\sqrt{f} - \sqrt{g})^2 \, d\mu = d_{H^2}(f, g)$$
Showing $\text{KL}(f, g) \geq d_{H^2}(f, g)$

Use the inequality $-\log(x) \geq 2 - 2\sqrt{x}$, to obtain:

$$\int f \log\left(\frac{f}{g}\right) d\mu = \int -f \log\left(\frac{g}{f}\right) d\mu \geq \int f \left(2 - 2\sqrt{\frac{g}{f}}\right) d\mu$$

$$= 1 + 1 - 2 \int \sqrt{f} \sqrt{g} d\mu$$

$$= \int f + g - 2\sqrt{f} \sqrt{g} d\mu$$

$$= \int (\sqrt{f} - \sqrt{g})^2 d\mu = d_{H^2}(f, g)$$
Showing $\text{KL}(f, g) \geq d_{H^2}(f, g)$

Use the inequality $-\log(x) \geq 2 - 2\sqrt{x}$, to obtain:

$$\int f \log\left(\frac{f}{g}\right) \, d\mu = \int -f \log\left(\frac{g}{f}\right) \, d\mu \geq \int f \left(2 - 2\sqrt{\frac{g}{f}}\right) \, d\mu$$

$$= 1 + 1 - 2 \int \sqrt{f} \sqrt{g} \, d\mu$$

$$= \int f + g - 2\sqrt{f} \sqrt{g} \, d\mu$$

$$= \int (\sqrt{f} - \sqrt{g})^2 \, d\mu = d_{H^2}(f, g)$$

Conclude:
$$\text{KL}(f, g) = 0 \implies d_{H^2}(f, g) = 0 \implies f = g \mu\text{-a.e.}$$
Conclusion

It has now been shown that

$$\text{KL}(f, g) \geq 0$$
Conclusion

It has now been shown that

\[ \text{KL}(f, g) \geq 0 \]
\[ \text{KL}(f, g) = 0 \iff f = g \mu\text{-a.e.} \]
Conclusion

It has now been shown that
\[ \text{KL}(f, g) \geq 0 \]
\[ \text{KL}(f, g) = 0 \iff f = g \text{ } \mu\text{-a.e.} \]
Together this implies that KL(f, g) has a (\mu-a.e.) unique minimum.
Conclusion

It has now been shown that

\[ KL(f, g) \geq 0 \]
\[ KL(f, g) = 0 \iff f = g \mu\text{-a.e.} \]

Together this implies that \( KL(f, g) \) has a (\( \mu\text{-a.e.} \)) unique minimum.

Thank you for your attention!