

Recap

- Transition Kernel: (X, \mathcal{F}) and (Y, \mathcal{G}) are measurable spaces.

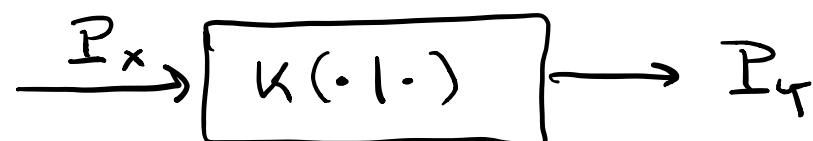
$K(\cdot | \cdot)$ is a transition kernel from space one to space two if:

$$\textcircled{1} \quad K(\cdot | x) \in P_{\mathcal{G}}(Y) \quad \begin{matrix} \downarrow \text{set of all probability measures over} \\ \text{space } Y \text{ wrt } \mathcal{G} \end{matrix} \quad \forall x \in X$$

$$\textcircled{2} \quad K(B | \cdot) : X \rightarrow \mathbb{R} \quad \forall B \in \mathcal{G}$$

is a rv wrt (X, \mathcal{F}) space

Interpretation: A transition kernel transforms a $P_x \in P_X(x)$ into $P_y \in P_{\mathcal{G}}(y)$



$$P_y(B) := \mathbb{E}_{P_x}[K(B|x)] = \int_X K(B|x) dP_x(x) \quad \forall B \in \mathcal{G}$$

There is a close relation between transition kernels and joint distributions that will give rise to decompositions such as $P_{XY} = P_X P_{Y|X}$

Proposition: Let (X, \mathcal{F}) and (Y, \mathcal{G}) be measurable spaces.

$P_x \in P_{\mathcal{F}}(X)$ and $K(\cdot | \cdot)$ be a transition kernel from (X, \mathcal{F}) to (Y, \mathcal{G}) .

Then, $\exists!$ probability measure P_{xy} on $(X \times Y, \mathcal{F} \otimes \mathcal{G})$ such that

$$P_{xy}(A, B) = \int_A K(B|x) dP_x(x) \quad \forall (A, B) \in \mathcal{F} \otimes \mathcal{G}$$

Conversely, given a probability space $(X \times Y, \mathcal{F} \otimes \mathcal{G}, P_{xy})$ $\exists!$ pair comprising $P_x \in P_{\mathcal{F}}(X)$ and a transition $K(\cdot | \cdot)$ s.t. the previous equation holds.

Comment

Henceforth, we will write $P_{xy} = P_x P_{y|x} = P_y P_{x|y}$ while understanding this equality wrt the previous proposition.

Conditional Expectations

Have

$$(X, Y) \sim P_{XY} \in \mathcal{P}(X \times Y)$$

Definition:

$$\mathbb{E}[Y|X=x_0] := \int_Y y dP_{Y|X}(y|x_0)$$

a number,
deterministic,
function of x_0

$$\mathbb{E}[Y|X] := \int_Y y dP_{Y|X}(y|X)$$

a random
variable

Proposition: Let $(X, Y) \sim P_{XY} \in \mathcal{P}(X \times Y)$. $\exists!$ measurable function $h: X \rightarrow Y$ such that

$$(i) \quad h(x) = \mathbb{E}[Y|X=x] \quad \forall x \in X$$

$$(ii) \quad h(X) = h \circ X = \mathbb{E}[Y|X]$$

Example/Exercise: $X = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases} \quad \text{if } Z \sim N(0, \sigma^2)$

$$Y \triangleq X + Z. \quad \mathbb{E}[X|Y] ? = \dots = \tanh\left(\frac{Y}{\sigma^2}\right) = h(Y)$$

Theorem: Law of Total Expectation

Let $(X, Y) \sim P_{XY} = P_X P_{Y|X} \circ P(X \times Y)$. Then

$$E_{P_Y}[Y] = E_{P_X}\left[E_{P_{Y|X}}[Y|X]\right]$$

$$= \int_X h(x) dP_X(x)$$

$$= \int_X E[Y|X=x] dP_X(x)$$

$$\left(E_{P_Y}[Y] = E_{P_X}\left[E_{P_{Y|X}}[Y|X]\right] = \sum_{x \in X} P_X(x) E_{P_{Y|X=x}}[Y|X=x] \right)$$

$$= \sum_{x \in X} P_X(x) \left(\sum_{y \in Y} y \cdot P_{Y|X=x}(y|x) \right)$$

f -divergences

Goal: We want to develop a way to measure a reasonable notion of "distance" between probability distributions.

Definition (Divergence): Let δ be a functional

$$\delta: \mathcal{P}(X) \times \mathcal{P}(X) \rightarrow \mathbb{R}_{\geq 0} \cup \{\infty\}$$

We will call δ a divergence if $\delta(P, Q) = 0 \Leftrightarrow P = Q$

Definition (Metric): If a divergence δ also satisfies

$$(i) \delta(P, Q) = \delta(Q, P) \quad \text{Symmetry}$$

$$(ii) \delta(P, Q) \leq \delta(P, R) + \delta(R, Q) \quad \forall P, Q, R \in \mathcal{P}(X)$$

↑ Triangle
Inequality

then δ is a metric

Why do f -divergences?

Aside: Generative Modeling

Target P

Model Class $\{Q_\theta\}_{\theta \in \mathcal{X}}$

"Learn a Q_θ that is a good proxy of P .."

Formulate: Pick δ divergence

$$\inf_{\Theta \in \mathcal{H}} \delta(Q_\theta, P)$$

AMAZING

Motivates Generative adversarial networks!

A large class of divergences falls under the framework of f-divergences. This includes

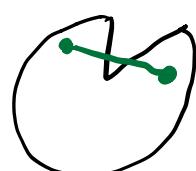
- Kullback Leibler distance
- Total Variation distance
- Hellinger distance
- χ^2 divergence
- Lecoin distance

Primer: Convexity

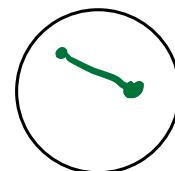
A subset K of a vector space V is convex if

$$\alpha x + (1-\alpha)y \in K \quad \forall x, y \in K; \alpha \in [0, 1]$$

Non-Convex



Convex



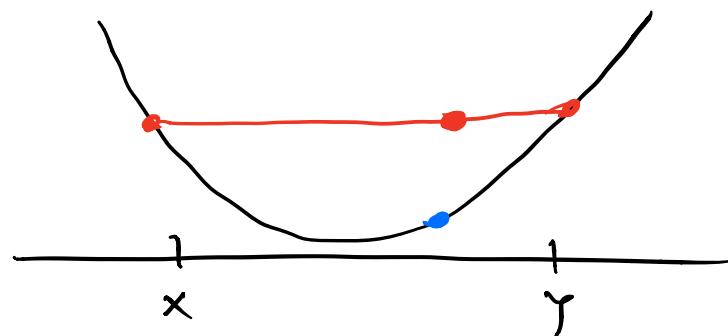
Example: $P(X)$ is convex

$$(\alpha P_1 + (1-\alpha) P_2)(A) \leq \alpha P_1(A) + (1-\alpha) P_2(A)$$

Definition: A function $f: K \rightarrow \mathbb{R}$, where $K \subseteq \mathbb{R}^d$ is convex if it satisfies

$$\boxed{f(\alpha x + (1-\alpha)y)} \leq \boxed{\alpha f(x) + (1-\alpha)f(y)} \quad \forall x, y \in K, \alpha \in [0,1]$$

Illustration



Comments: ① f is strictly convex if the above inequality is STRICT

② f is concave if $-f$ is convex

Def(epigraph):

The epigraph of a function $f: K \rightarrow \mathbb{R}$ is

$$\text{epi}(f) := \{(x, y) \in K \times \mathbb{R} \mid y \geq f(x)\}$$

Prop: $f: K \rightarrow \mathbb{R}$ is convex $\Leftrightarrow \text{epi}(f)$ a convex set

Prop (Operations that conserve convexity): Let $f_1, f_2: \mathbb{R}^d \rightarrow \mathbb{R}$ be convex. Then:

- ① $f_1 + f_2$ is convex
- ② $\max\{f_1, f_2\}$ is convex
- ③ $A \in \mathbb{R}^{n \times d}$, $g: \mathbb{R}^d \rightarrow \mathbb{R}$, $g(x) = f_1(Ax)$ is convex

Proposition: $f: K \rightarrow \mathbb{R}$, $K \subseteq \mathbb{R}^d$ is convex iff its Hessian is Positive Semi Definite.

$$\text{Hess}(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_d} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_d \partial x_1} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{bmatrix}$$

Note: $A \in \mathbb{R}^{n \times n}$ is PSD if $x^T A x \geq 0 \quad \forall x \in \mathbb{R}^n$

Corollary: $f: \mathbb{R} \rightarrow \mathbb{R}$

① f convex iff $\frac{\partial^2 f}{\partial x^2} \geq 0$

② f strictly convex iff $\frac{\partial^2 f}{\partial x^2} > 0$

Theorem (Jensen's Inequality)

Let $X \sim P_x \in \mathcal{P}(x)$ and $f: X \rightarrow \mathbb{R}$. Then

① If f is convex

$$\mathbb{E}[f(x)] \geq f(\mathbb{E}[x])$$

② If f is strictly convex

$$\mathbb{E}[f(x)] > f(\mathbb{E}[x])$$

unless X is almost surely constant

③ If f is concave

$$\mathbb{E}[f(x)] \leq f(\mathbb{E}[x])$$