23 June 2010 at Royal Holloway

Randomness criteria in terms of f-divergences

Akio Fujiwara (Osaka University)

A joint research with Masahiro Nakamura

Outline

- What is an f-divergence?
- What is a randomness criterion?
- Randomness criteria in terms of f-divergences

Part I: f-divergence

Divergence function

A divergence function is a convex function

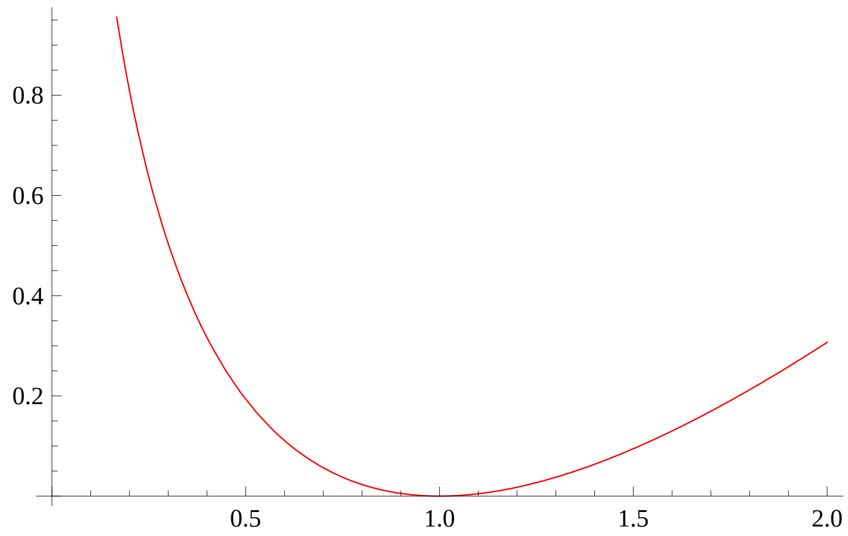
$$f : (0, \infty) \to [0, \infty)$$

satisfying

$$f(x) = 0 \iff x = 1$$

Its transpose is defined by

$$f^*(x) := xf\left(\frac{1}{x}\right)$$



f-divergence

Given a divergence function f, the f-divergence between probability measures P, Q on $(\Omega, \mathcal{F})$ is:

$$D_f(P\|Q) := \int_\Omega p(\omega)\, f\left(\frac{q(\omega)}{p(\omega)}\right) d\mu(\omega)$$

Here we understand that

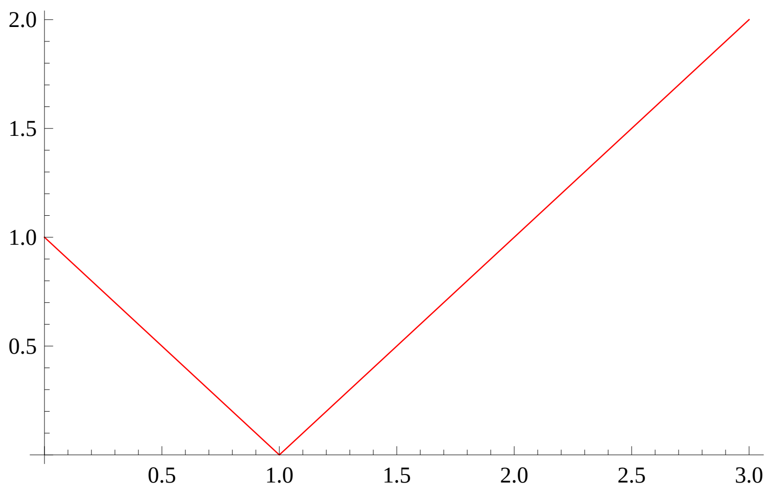
$$0 \cdot f\left(\frac{0}{0}\right) = 0$$

and

$$0 \cdot f\left(\frac{a}{0}\right) = \lim_{x \downarrow 0} x\, f\left(\frac{a}{x}\right) = a f^*(0) \qquad (a > 0)$$

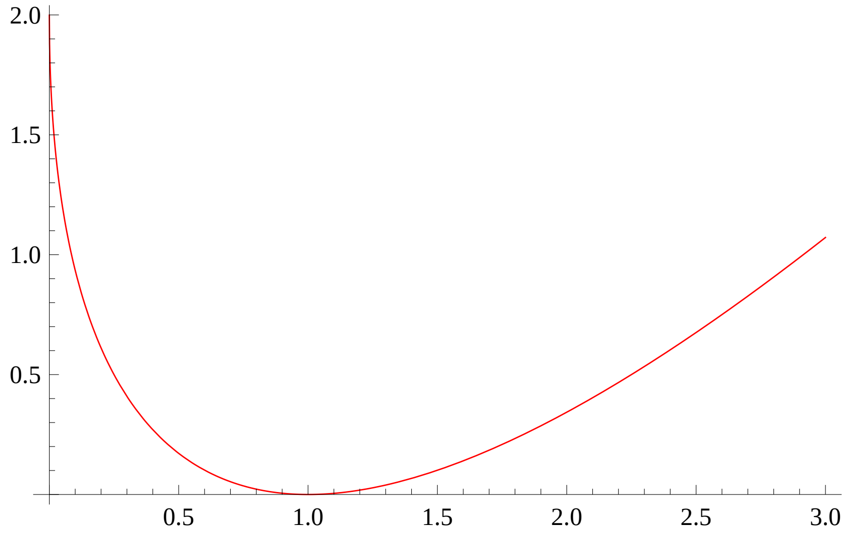
Example 1: Variational distance

$$
\begin{aligned}
D_{|1-x|}(P\|Q) &= \int_\Omega |p(\omega) - q(\omega)|\, d\mu(\omega) \\
&= \{P(E) - Q(E)\} + \{Q(E^c) - P(E^c)\} \\
&\qquad\qquad (E := \{p \geq q\}) \\
&= \|P - Q\|_1
\end{aligned}
$$



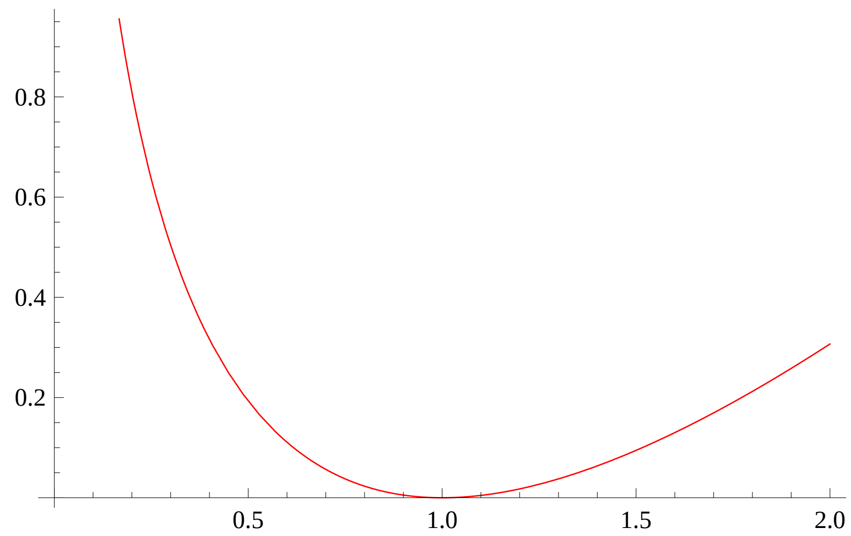
Example 2: Hellinger distance

$$D_{2(1-\sqrt{x})^2}(P\|Q) = 2\int_\Omega (\sqrt{p(\omega)} - \sqrt{q(\omega)})^2 \, d\mu(\omega)$$



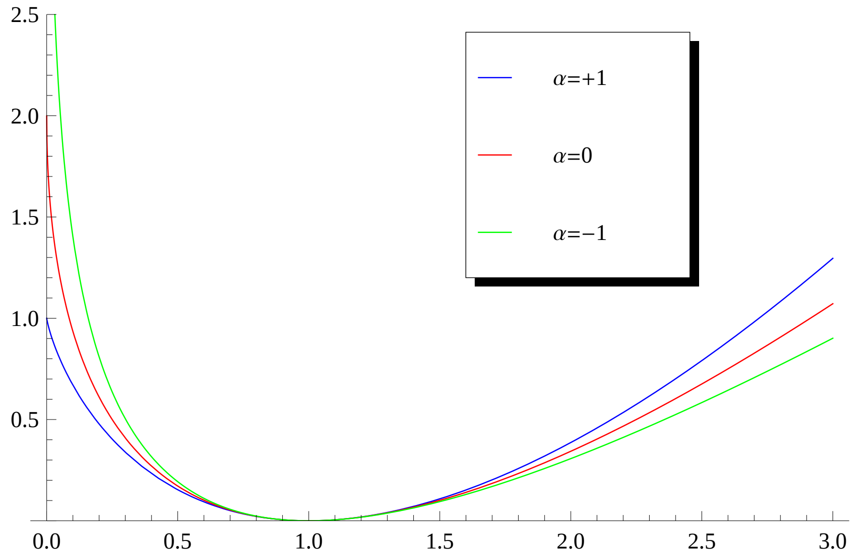
Example 3: Kullback-Leibler divergence

$$D_{(x-1-\log x)}(P\|Q) = \int_\Omega p(\omega) \log\left(\frac{p(\omega)}{q(\omega)}\right)\, d\mu(\omega)$$



Example 4: α-divergence

$$f^{(\alpha)}(x) := \begin{cases} \dfrac{4}{1-\alpha^2}\left[\dfrac{1-\alpha}{2} + \dfrac{1+\alpha}{2}x - x^{\frac{1+\alpha}{2}}\right], & \alpha \neq \pm 1 \\ 1 - x + x\log x, & \alpha = +1 \\ -1 + x - \log x, & \alpha = -1 \end{cases}$$



$$D^{(\alpha)}(P\|Q) := \begin{cases} \dfrac{4}{1-\alpha^2}\left(1 - \int_\Omega p(\omega)^{\frac{1-\alpha}{2}} q(\omega)^{\frac{1+\alpha}{2}}\right) d\mu, & \alpha \neq \pm 1 \\ \int_\Omega q(\omega) \log \dfrac{q(\omega)}{p(\omega)}\, d\mu & \alpha = +1 \\ \int_\Omega p(\omega) \log \dfrac{p(\omega)}{q(\omega)}\, d\mu & \alpha = -1 \end{cases}$$

For example:

- 0-divergence = Hellinger distance
- (-1)-divergence = Kullback-Leibler divergence

Topological properties

- For a set $\mathcal{P}$ of probability measures on $(\Omega, \mathcal{F})$,

$$U_f(P, \varepsilon) := \{Q \in \mathcal{P} \; ; \; D_f(P \| Q) < \varepsilon\}$$

is called the f-neighborhood of radius ε of P.

- The weakest topology on $\mathcal{P}$ containing the family

$$\mathcal{U}_f := \bigcup_{P \in \mathcal{P}} \{U_f(P, \varepsilon) \; ; \; \varepsilon > 0\} .$$

is called the topology τ_f induced by the f-divergence.

- Does the subbase $\mathcal{U}_f$ form a base of τ_f?

Csiszár's theorem

Let $f(0) := \lim_{x \downarrow 0} f(x)$ and $f^*(0) := \lim_{x \downarrow 0} f^*(x) = \lim_{x \to \infty} \dfrac{f(x)}{x}$.

Theorem [Csiszár, 1967].

[1] If $f(0) + f^*(0) < \infty$, then $\mathcal{U}_f$ forms a basis of τ_f.

[2] If $f(0) + f^*(0) = \infty$, then $\mathcal{U}_f$ does not in general form a basis of τ_f.

[3] If $f(0) + f^*(0) < \infty$ and $g(0) + g^*(0) < \infty$, then $\tau_f = \tau_g$.

[4] If $f(0) + f^*(0) < \infty$ and $g(0) + g^*(0) = \infty$, then $\tau_f \prec \tau_g$.

Topological properties of α-divergences

Corollary [Csiszár, 1967].

[1] For $|\alpha| < 1$, the α-topology induced by the α-divergence is equivalent to the metric topology induced by the variational distance.

[2] For $|\alpha| \geq 1$, the α-topology is stronger than the variational distance topology.

Measure theoretic properties

Theorem [Vajda, 1972].

[1] $D_f(P\|Q) \leq f(0) + f^*(0)$.

[2] $D_f(P\|Q) = f(0) + f^*(0)$ if $P \perp Q$.

[3] Suppose $f(0) + f^*(0) < \infty$. Then

$$D_f(P\|Q) = f(0) + f^*(0) \iff P \perp Q$$

Part II: Randomness criteria

Notations

- $\mathcal{A}$: a finite set (e.g. $\mathcal{A} = \{0, 1\}$)

  $\mathcal{A}^\infty$: the set of one-sided infinite sequences

- $\mathcal{F}$: σ-algebra on $\mathcal{A}^\infty$ generated by $\{\Gamma_s \ ; \ s \in \mathcal{A}^*\}$

- $\mathcal{P}(\mathcal{A}^\infty)$: the set of computable probability measures

Martin-Löf randomness

- Given a $P \in \mathcal{P}(\mathcal{A}^\infty)$, a sequence $\{M_k\}_{k=1}^{\infty}$, each M_k being a family of cylinder sets, is called a sequential Martin-Löf P-test if

  1. the set $\{(k, s) \in \mathbb{N} \times \mathcal{A}^*; \Gamma_s \in M_k\}$ is r.e., and

  2. $P\left(\bigcup M_k\right) < 2^{-k}$ for each k.

- We say a sequential Martin-Löf P-test $\{M_k\}_{k=1}^{\infty}$
  - rejects $\omega \in \mathcal{A}^\infty$ if $\omega \in \bigcup M_k$ infinitely often
  - accepts $\omega \in \mathcal{A}^\infty$ if $\omega \notin \bigcup M_k$ eventually

Martin-Löf randomness (cont'd)

- A sequence $\omega \in \mathcal{A}^\infty$ is
  - P-nonrandom if it is rejected by a sequential P-test.
  - P-random if there is no P-test that rejects ω.
- $\mathsf{R}_{\mathsf{ML}}(P)$: the set of P-random infinite sequences.

Basic properties of random sequences

- $P(\mathsf{R}_{\mathsf{ML}}(P)) = 1$.

- $\mathsf{R}_{\mathsf{ML}}(P) \cap \mathsf{R}_{\mathsf{ML}}(Q) = \emptyset \iff P \perp Q$

- $\mathsf{R}_{\mathsf{ML}}(P) \subset \mathsf{R}_{\mathsf{ML}}(Q) \implies P \ll Q$

Characterization of $\mathsf{R}_{\mathsf{ML}}(P) \cap \mathsf{R}_{\mathsf{ML}}(Q)$

- Likelihood ratio test:

  Let $\omega \in \mathsf{R}_{\mathsf{ML}}(P)$. Then $\omega \in \mathsf{R}_{\mathsf{ML}}(Q)$ if and only if

$$0 < \lim_{n\to\infty} \frac{Q(\omega^n)}{P(\omega^n)} < \infty$$

Characterization of $\mathsf{R}_{\mathsf{ML}}(P) \cap \mathsf{R}_{\mathsf{ML}}(Q)$

- Likelihood ratio test:

  Let $\omega \in \mathsf{R}_{\mathsf{ML}}(P)$. Then $\omega \in \mathsf{R}_{\mathsf{ML}}(Q)$ if and only if

$$0 < \lim_{n\to\infty} \frac{Q(\omega^n)}{P(\omega^n)} < \infty$$

- Randomness criterion [Vovk, 1987]:

  Let $\omega \in \mathsf{R}_{\mathsf{ML}}(P)$. Then $\omega \in \mathsf{R}_{\mathsf{ML}}(Q)$ if and only if $Q(\omega^n) \neq 0$ for all n and

$$\sum_{i=1}^{\infty} D^{(0)}\Big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\Big) < \infty$$

Application: Kakutani dichotomy

When P and Q are product measures satisfying $P \overset{loc}{\ll} Q$, then either $P \ll Q$ or $P \perp Q$ holds, and

(i) $P \ll Q \iff \mathsf{R}_{\mathsf{ML}}(P) \subset \mathsf{R}_{\mathsf{ML}}(Q)$

(ii) $P \perp Q \iff \mathsf{R}_{\mathsf{ML}}(P) \cap \mathsf{R}_{\mathsf{ML}}(Q) = \emptyset$

Randomness criteria in terms of α-divergences?

- Vovk's theorem:

  Let $\omega \in \mathsf{R}_{\mathsf{ML}}(P)$. Then $\omega \in \mathsf{R}_{\mathsf{ML}}(Q)$ if and only if $Q(\omega^n) \neq 0$ for all n and

  $$\sum_{i=1}^{\infty} D^{(0)}\Big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\Big) < \infty$$

- Can this be extended to other α-divergences?

Randomness criteria in terms of α-divergences?

- Vovk's theorem:

  Let $\omega \in \mathsf{R}_{\mathsf{ML}}(P)$. Then $\omega \in \mathsf{R}_{\mathsf{ML}}(Q)$ if and only if $Q(\omega^n) \neq 0$ for all n and

$$\sum_{i=1}^{\infty} D^{(0)}\big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\big) < \infty$$

- Can this be extended to other α-divergences?

$\longrightarrow$ Yes!

Randomness criteria in terms of α-divergences

- [F, 2008]

  Let $\omega \in \mathrm{R}_{\mathrm{ML}}(P)$. Then $\omega \in \mathrm{R}_{\mathrm{ML}}(Q)$ if and only if $Q(\omega^n) \neq 0$ for all n and

$$\sum_{i=1}^{\infty} D^{(\alpha)}\Big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\Big) < \infty$$

  for some (in fact, any) $\alpha \in (-1, 1)$.

Randomness criteria in terms of α-divergences

- [F, 2008]

  Let $\omega \in \mathrm{R}_{\mathrm{ML}}(P)$. Then $\omega \in \mathrm{R}_{\mathrm{ML}}(Q)$ if and only if $Q(\omega^n) \neq 0$ for all n and

  $$\sum_{i=1}^{\infty} D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\big) < \infty$$

  for some (in fact, any) $\alpha \in (-1, 1)$.

- NB: α cannot be extended to $\alpha = \pm 1$ and beyond.

Example

Let P, Q be product measures, $P = \prod P_n$, $Q = \prod Q_n$, on $\{0,1\}^\infty$ defined by

$$P_n(0) = \frac{1}{2n^2} \quad \text{and} \quad Q_n(0) = e^{-n}.$$

Then

$$\begin{aligned}
\frac{1}{4} D^{(0)}(P_n \| Q_n) &= 1 - \sum_{x \in \{0,1\}} \sqrt{P_n(x) Q_n(x)} \\
&\leq 1 - \sum_{x \in \{0,1\}} P_n(x) Q_n(x) = \frac{1}{2n^2} + e^{-n} - \frac{e^{-n}}{n^2}
\end{aligned}$$

Consequently, $\sum_n D^{(0)}(P_n \| Q_n) < \infty$, and

$$\mathsf{R}_{\mathsf{ML}}(P) = \mathsf{R}_{\mathsf{ML}}(Q)$$

Example (cont'd)

On the other hand

$$\begin{aligned} D^{(-1)}(P_n \| Q_n) &\geq Q_n(0) - P_n(0) + P_n(0) \log \frac{P_n(0)}{Q_n(0)} \\ &= e^{-n} - \frac{1}{2n^2} + \frac{1}{2n^2} \log \left(\frac{e^n}{2n^2} \right) = O\left(\frac{1}{n} \right) \end{aligned}$$

Consequently

$$\sum_n D^{(-1)}(P_n \| Q_n) = \infty = \sum_n D^{(+1)}(Q_n \| P_n)$$

Thus the KL-divergence does not give a randomness criterion.

Merging of opinions

Randomness criterion is closely related to the issue of Merging of opinions.

- **Theorem** [Blackwell and Dubins, 1962]

  If $P \ll Q$, then

  $$\lim_{n\to\infty} \|P(X^\infty|\mathcal{F}_{n-1}) - Q(X^\infty|\mathcal{F}_{n-1})\|_1 = 0 \qquad (P\text{-a.s.})$$

- **Corollary**

  If $P \ll Q$, then for each $\omega_n \in \mathcal{A}$,

  $$\lim_{n\to\infty} |P(\omega_n|\mathcal{F}_{n-1}) - Q(\omega_n|\mathcal{F}_{n-1})| = 0 \qquad (P\text{-a.s.})$$

- Randomness criterion is an "individual" version of merging of opinions.

Randomness criterion

Let $\omega \in \mathrm{R}_{\mathrm{ML}}(P)$. Then $\omega \in \mathrm{R}_{\mathrm{ML}}(Q)$ if and only if $Q(\omega^n) \neq 0$ for all n and

$$\sum_{i=1}^{\infty} D^{(\alpha)}\big(P(\,\cdot\,|\omega^{i-1}) \| Q(\,\cdot\,|\omega^{i-1})\big) < \infty$$

for any $\alpha \in (-1, 1)$.

- This has a game theoretic counterpart [Vovk, 2009].

Competitive Testing Protocol [Vovk, 2009]

Players: Reality, Forecasters I and II, Sceptics I and II

Protocol:

$\mathcal{K}^{\mathrm{I}}(\varepsilon) := 1$ and $\mathcal{K}^{\mathrm{II}}(\varepsilon) := 1$.

FOR $n = 1, 2, \dots$:

Forecaster I announces P_n^{I}.

Forecaster II announces P_n^{II}.

Sceptic II announces $f_n^{\mathrm{II}} : \mathcal{A} \to \mathbb{R}_+$ such that $E_{P_n^{\mathrm{II}}}[f_n^{\mathrm{II}}] = 1$.

Sceptic I announces $f_n^{\mathrm{I}} : \mathcal{A} \to \mathbb{R}_+$ such that $E_{P_n^{\mathrm{I}}}[f_n^{\mathrm{I}}] = 1$.

Reality announces $\omega_n \in \mathcal{A}$.

$\mathcal{K}^{\mathrm{I}}(\omega^n) := \mathcal{K}^{\mathrm{I}}(\omega^{n-1}) \cdot f_n^{\mathrm{I}}(\omega_n)$ and $\mathcal{K}^{\mathrm{II}}(\omega^n) := \mathcal{K}^{\mathrm{II}}(\omega^{n-1}) \cdot f_n^{\mathrm{II}}(\omega_n)$.

END FOR

Merging of opinions in game theoretic probability [Vovk, 2009]

Let $\alpha \in (-1, 1)$, and suppose that Forecaster II is timid, in that $P_n^{\mathrm{I}} \ll P_n^{\mathrm{II}}$ for all n.

(i) If $\sum_{n=1}^{\infty} D^{(\alpha)}(P_n^{\mathrm{I}} \| P_n^{\mathrm{II}}) = \infty$, then Sceptics I and II have a joint strategy guaranteeing that at least one of them will become infinitely rich.

(ii) If $\sum_{n=1}^{\infty} D^{(\alpha)}(P_n^{\mathrm{I}} \| P_n^{\mathrm{II}}) < \infty$ and Sceptic II becomes infinitely rich, then Sceptic I have a strategy guaranteeing that he will become infinitely rich.

To put it differently...

- **Notation**

$$\omega \in \mathsf{R}(\{P_n^{\mathrm{I}}\}) \iff \sup_n \mathcal{K}^{\mathrm{I}}(\omega^n) < \infty \text{ for any strategy } \{f_n^{\mathrm{I}}\}$$

$$\omega \in \mathsf{R}(\{P_n^{\mathrm{II}}\}) \iff \sup_n \mathcal{K}^{\mathrm{II}}(\omega^n) < \infty \text{ for any strategy } \{f_n^{\mathrm{II}}\}$$

- **Theorem** [Vovk, 2009]

Suppose that Forecaster II is timid, and that $\omega \in \mathsf{R}(\{P_n^{\mathrm{I}}\})$. Then $\omega \in \mathsf{R}(\{P_n^{\mathrm{II}}\})$ if and only if

$$\sum_{n=1}^{\infty} D^{(\alpha)}(P_n^{\mathrm{I}} \| P_n^{\mathrm{II}}) < \infty$$

for any $\alpha \in (-1, 1)$.

Observation

- Randomness criteria and game theoretic merging of opinions are represented in terms of α-divergences ($|\alpha| < 1$).

- $|\alpha| < 1 \iff f^{(\alpha)}(0) + f^{(\alpha)*}(0) < \infty$.

- All f-divergences with $f(0) + f^*(0) < \infty$ generate the same topology as the variational distance.

- Blackwell-Dubins theorem is represented in terms of the variational distance.

Conjecture

Let $\omega \in \mathsf{R}_{\mathsf{ML}}(P)$. Then $\omega \in \mathsf{R}_{\mathsf{ML}}(Q)$ if and only if $Q(\omega^n) \neq 0$ for all n and

$$\sum_{i=1}^{\infty} D_f\big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\big) < \infty$$

for any f satisfying $f(0) + f^*(0) < \infty$.

Conjecture

Let $\omega \in \mathsf{R}_{\mathsf{ML}}(P)$. Then $\omega \in \mathsf{R}_{\mathsf{ML}}(Q)$ if and only if $Q(\omega^n) \neq 0$ for all n and

$$\sum_{i=1}^{\infty} D_f\Big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\Big) < \infty$$

for any f satisfying $f(0) + f^*(0) < \infty$.

$\Downarrow$

Wrong!

Counterexample

Let P, Q be product measures on $\{0,1\}^\infty$ defined by

$$P_n(0) := \frac{1}{2} \quad \text{and} \quad Q_n(0) := \frac{1}{2}\left(1 + \frac{1}{n+1}\right).$$

Then

$$D^{(0)}(P_n \| Q_n) = 4 - 4 \sum_{x \in \{0,1\}} \sqrt{P_n(x) Q_n(x)}$$

$$= 4 - 2\left\{\sqrt{1 + \frac{1}{n+1}} + \sqrt{1 - \frac{1}{n+1}}\right\} \leq \frac{4 - 2\sqrt{2}}{(n+1)^2}$$

Consequently, $\sum_{n=1}^{\infty} D^{(0)}(P_n \| Q_n) < \infty$, and

$$\mathsf{R}_{\mathsf{ML}}(P) = \mathsf{R}_{\mathsf{ML}}(Q)$$

Counterexample (cont'd)

On the other hand,

$$\|P_n - Q_n\|_1 = \sum_{x \in \{0,1\}} |P_n(x) - Q_n(x)| = \frac{1}{n+1}$$

Consequently, $\sum_{n=1}^{\infty} \|P_n - Q_n\|_1 = \infty$.

Thus the variational distance does not give a randomness criterion.

Toward randomness criteria in terms of f-divergences

- We say a divergence function $f(x)$ quasi-symmetric if

$$\limsup_{x \to 1} \frac{f^*(x)}{f(x)} < \infty$$

- Remark:
  - This is equivalent to $\liminf_{x \to 1} \frac{f^*(x)}{f(x)} > 0$
  - $f(x)$ is quasi-symmetric $\iff$ $f^*(x)$ is quasi-symmetric
  - $f^{(\alpha)}(x)$ and $f(x) = |1 - x|$ are quasi-symmetric

New randomness criterion

Let $\omega \in \mathsf{R}_{\mathsf{ML}}(P)$. Then $\omega \in \mathsf{R}_{\mathsf{ML}}(Q)$ if and only if $Q(\omega^n) \neq 0$ for all n and

$$\sum_{i=1}^{\infty} D_f\Big(P(\,\cdot\,|\omega^{i-1})\|Q(\,\cdot\,|\omega^{i-1})\Big) < \infty$$

for any quasi-symmetric divergence function f satisfying $f(0) + f^*(0) < \infty$ and

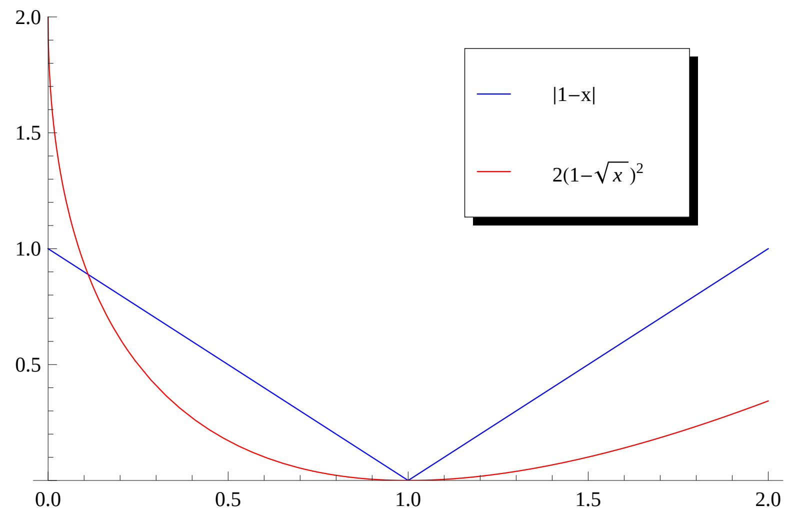
$$0 < \liminf_{x\to 1} \frac{f(x)}{(1-\sqrt{x})^2} \leq \limsup_{x\to 1} \frac{f(x)}{(1-\sqrt{x})^2} < \infty.$$

Remark and Example

- If $f(x)$ is twice differentiable at $x = 1$, the last condition is equivalent to $0 < f''(1) < \infty$.

- Variational distance:

$$\lim_{x \to 1} \frac{|1 - x|}{(1 - \sqrt{x})^2} = \infty.$$



Concluding remarks

- Randomness criteria can be extended to a certain class of f-divergences.

- It is not clear if "quasi-symmetricity" can be dispensed with.