# Distance to a measure to compare samples of points

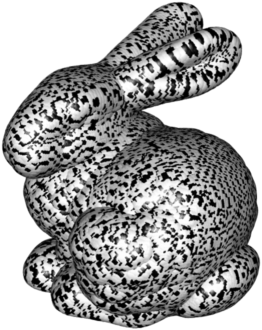
Claire Bréchetteau

Université Paris-Sud 11, Laboratoire de Mathématiques d'Orsay and Inria Select Team – Inria Saclay, Data Shape Team

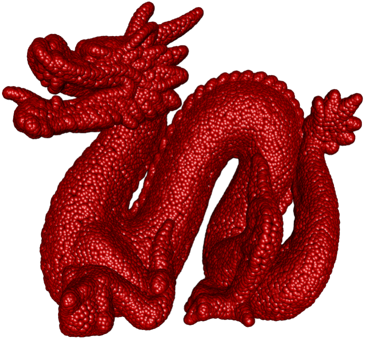
Under the supervision of
Pascal Massart (Université Paris-Sud 11) and Frédéric Chazal (Inria Saclay)

23 juin 2016

# Can a dragon pretend to be a bunny ?



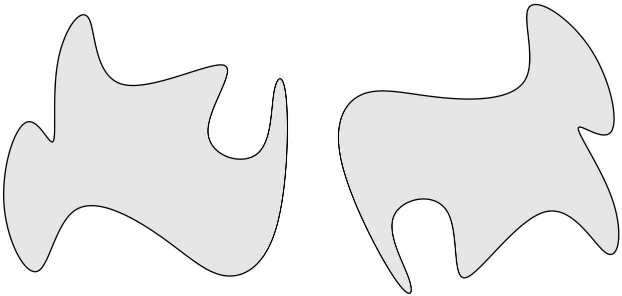
Bunny



Dragon

Data from the Stanford 3D Scanning Repository

A **metric measure space** (mm-space) is a triple $(\mathcal{X}, \mathrm{d}, \mu)$ s.t. : $(\mathcal{X}, \mathrm{d})$ is a metric space and $\mu$ a borel measure supported on $\mathcal{X}$.

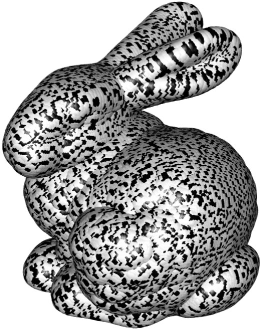


Two mm-spaces $(\mathcal{X}, \mathrm{d}, \mu)$ and $(\mathcal{Y}, \mathrm{d}', \nu)$ are **isomorphic** if : $\exists\, \phi : \mathcal{X} \mapsto \mathcal{Y}$ a one-to-one isometry, s.t. for all borel set $A$,

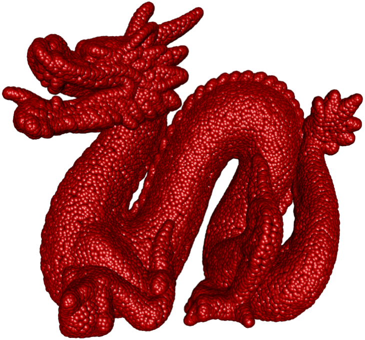
$$\nu(\phi(A)) = \mu(A).$$

How to build a test of level $\alpha > 0$ to test the null

$H_0(\mathcal{X})$ : "$(\mathcal{Y}, \mathrm{d}', \nu)$ is isomorphic to $(\mathcal{X}, \mathrm{d}, \mu)$" ?



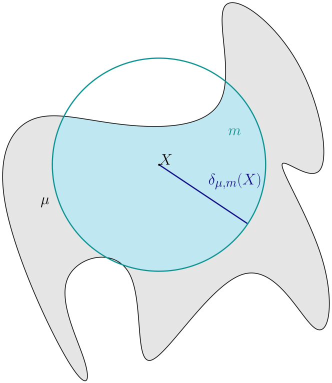
$(\mathcal{X}, \mathrm{d}, \mu)$



$(\mathcal{Y}, \mathrm{d}', \nu)$

The **distance to a measure** [Chazal, Cohen-Steiner, Mérigot 2009] is defined for all $x \in \mathcal{X}$ and $m \in [0,1]$ by :

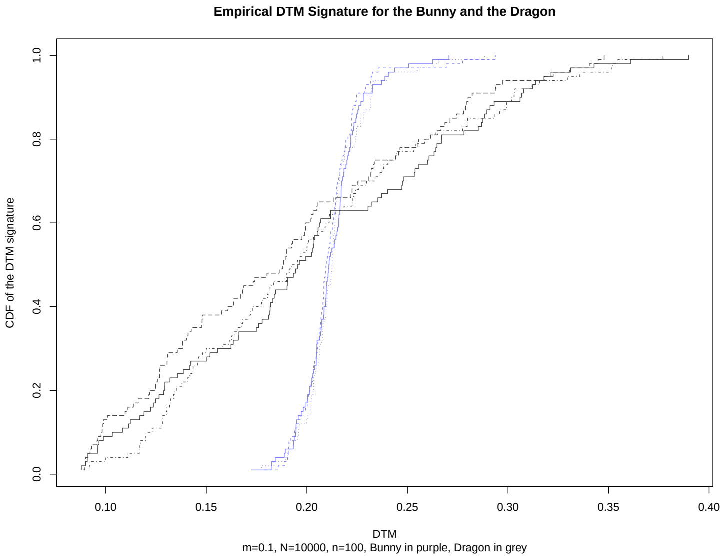
$$\mathrm{d}_{\mu,m}(x) = \frac{1}{m}\int_0^m \delta_{\mu,l}(x)\mathrm{d}l.$$




If $X \sim \mu$, the **distance-to-a-measure signature** $\mathrm{d}_{\mu,m}(\mu)$, is the distribution of $\mathrm{d}_{\mu,m}(X)$.

*N*-samples on the Bunny and the Dragon

The **empirical distance-to-a-measure signature** is defined by $\mathrm{d}_{\widehat{\mu}_{N-n},m}(\widehat{\mu}_n)$ from two independent $(N-n)$ and $n$-samples of law $\mu$.




Our test statistic : $T = \sqrt{n} W_1(\mathrm{d}_{\widehat{\mu}_{N-n},m}(\widehat{\mu}_n), \mathrm{d}_{\widehat{\nu}_{N-n},m}(\widehat{\nu}_n))$,
$\sqrt{n}$ times the area between the two cdf.

Under the null, we approximate
$\mathcal{L}(\sqrt{n}W_1(\mathrm{d}_{\widehat{\mu}_{N-n},m}(\widehat{\mu}_n), \mathrm{d}_{\widehat{\nu}_{N-n},m}(\widehat{\nu}_n)))$ by the bootstrap law :
$\mathcal{L}^*(\sqrt{n}W_1(\mathrm{d}_{\widehat{\mu}_N,m}(\mu_n^*), \mathrm{d}_{\widehat{\nu}_N,m}(\nu_n^*))|\widehat{\mu}_N, \widehat{\nu}_N)$.