Software

We are pleased to make available two software repositories accompanying this analysis.

- Package `compx` for the `R` programming language implements computation of the information measures $H(Y)$, $I(X,Y)$, and $J(X,Y)$, as well as a method for information-theoretic clustering. Access `compx` at

  `https://github.com/PhilChodrow/compx`.

  To install in `R`, install the package `devtools` at

  `https://cran.r-project.org/web/packages/devtools/index.html`

  and run the command

  `devtools::install_github(‘‘PhilChodrow/compx’’)`

  in the `R` console.

- The analysis repository for this project, including data acquisition and processing; core computations; and figure generation. We hope that this repository will provide useful examples of how to use `compx` for others aiming to replicate and extend our results. Download the project files at

  `https://github.com/PhilChodrow/spatial_complexity`.

Relation of mutual and Fisher informations

Let X be a continuous random variable taking values in $\mathbb{R}^n$, and let Y be a discrete random variable define on finite alphabet $\mathcal{Y}$. Suppose further that $p(y|x) > 0$ and that $p(y|x)$ is differentiable as a function of x for all $x, y \in \mathbb{R}^n \times \mathcal{Y}$. Fix $x_0 \in \mathbb{R}^n$, and define $B_r \triangleq B_r(x_0) = \{x \in R^n \mid \|x - x_0\| \leq r\}$. Additionally, define the *local mutual information* in B_r as the mutual information between X and Y where X is restricted to B_r:

$$I_r(x_0) \triangleq \mathbb{E}_X[D[p(\cdot|X)\|p(\cdot|X \in B_r)]|X \in B_r] \tag{1}$$

$$= \int_{B_r} p(x|X \in B_r) D[p(\cdot|x)\|p(\cdot|X \in B_r)] d^n x \ . \tag{2}$$

where $D[p\|q] \triangleq \sum_y p(y) \log \frac{p(y)}{q(y)}$ is the Kullback-Leibler divergence of q from p.

Theorem 1. *Under the stated conditions,*

$$\lim_{r \to 0} \frac{I_r(x_0)}{r^2} = \frac{n}{2(n+2)} \text{trace } J_Y(x_0) \ . \tag{3}$$

where the Fisher information matrix J_Y is given by

$$J_Y(x) \triangleq \mathbb{E}_Y \left[\nabla_x S_Y(x) \nabla_x S_Y(x)^T\right] \tag{4}$$

$$S_y(x) \triangleq \log p(y|x) \ . \tag{5}$$

The proof of Theorem 2 proceeds by the application of a number of Taylor approximations, in tandem with a fundamental relationship of information geometry. We first expand out $I_r(x_0)$ explicitly as

$$I_r(x_0) = \int_{B_r} p(x|X \in B_r) D[p(\cdot|x) \| p(\cdot|X \in B_r)] d^n x \ . \tag{6}$$

Lemma 1. *The following approximation relationships hold for the components of* (6)*:*

(a) $p(X \in B_r, Y) = p(x_0, Y) v(B_r) + O(r^{n+2})$

(b) $p(Y|X \in B_r) = p(Y|x_0) + e_y$ *where the error terms* e_y *satisfy* $e_y \in O(r^2)$ *and* $\sum_{y \in \mathcal{Y}} e_y = 0$.

(c) $p(x|X \in B_r) = \frac{1+O(r)}{v(B_r)}$

Proof. For each approximation, we directly apply Taylor expansions about $X = x_0$.

(a) We have

$$\begin{aligned}
p(X \in B_r, Y) &= \int_{B_r} p(x, Y)\ d^n x && (7)\\
&= \int_{B_r} p(x_0, Y) + \frac{\partial p(x_0, Y)}{\partial x}(x - x_0) + O(\|x - x_0\|^2)\ d^n x && (8)\\
&= p(x_0, Y) v(B_r) + \frac{\partial p(x_0, Y)}{\partial x} \int_{B_r} (x - x_0)\ d^n x && (9)\\
&\quad + O\left(\int_{B_r} \|x - x_0\|^2\ d^n x \right) && (10)\\
&= p(x_0, Y) v(B_r) + O(r^{n+2}) \ , && (11)
\end{aligned}$$

where the middle term vanishes due to spherical symmetry.

(b) The fact that the error terms e_y must satisfy $\sum_{y \in \mathcal{Y}} e_y = 0$ follows from the fact that $p(Y|X \in B_r)$ must be a valid probability distribution over $\mathcal{Y}$. We'll now show that $e_y \in O(r^2)$. First,

$$\begin{aligned}
p(X \in B_r) &= \sum_{y \in \mathcal{Y}} p(X \in B_r, y) && (12)\\
&= \sum_{y \in \mathcal{Y}} \left[p(x_0, y) v(B_r) + O(r^{n+2}) \right] && (13)\\
&= p(x_0) v(B_r) + O(r^2); && (14)
\end{aligned}$$

from part (a). Next,

$$\begin{aligned}
p(Y|X \in B_r) &= \frac{p(X \in B_r, Y)}{p(X \in B_r)} && (15)\\
&= \frac{p(x_0, Y) v(B_r) + O(r^{n+2})}{p(x_0) v(B_r) + O(r^{n+2})} && (16)\\
&= p(Y|x_0) + O(r^2) \ , && (17)
\end{aligned}$$

which completes this part of the argument.

(c) First,

$$p(X \in B_r) = \int_{B_r} p(x)\, d^n x \tag{18}$$

$$= \int_{B_r} \left[p(x_0) + \nabla p(x_0)(x - x_0) + O(r^2) \right] \, d^n x \tag{19}$$

$$= p(x_0) v(B_r) + O(r^{n+2}) \, , \tag{20}$$

where the middle term again vanishes through spherical symmetry. Thus, for $x \in B_r$, we have

$$p(x|X \in B_r) = \frac{p(x)}{p(X \in B_r)} \tag{21}$$

$$= \frac{p(x_0) + \nabla p(x_0)(x - x_0) + O(r^2)}{p(x_0) v(B_r) + O(r^{n+2})} \tag{22}$$

$$= \frac{1 + O(r)}{v(B_r)} \, . \tag{23}$$

□

Lemma 2. *The following approximation holds for the divergence factor in the integral* (6)

$$D[p(\cdot|x) \| p(\cdot|X \in B_r)] = D[p(\cdot|x) \| p(\cdot|x_0)] + O(r^3) \tag{24}$$

Proof. We compute directly:

$$
\begin{aligned}
D[p(\cdot|x)\|p(\cdot|X \in B_r)] &= \sum_{y\in\mathcal{Y}} p(y|x)\log\frac{p(y|x)}{p(y|X \in B_r)} && (25)\\
&= -H[Y|X=x] - \sum_{y\in\mathcal{Y}} p(y|x)\log p(y|X\in B_r) && (26)\\
&= -H[Y|X=x] - \sum_{y\in\mathcal{Y}} p(y|x)\log\left(p(y|x_0)+e_y\right) \\
& && \text{(from Lemma 1)}\\
&= -H[Y|X=x] - \sum_{y\in\mathcal{Y}} p(y|x)\left[\log p(y|x_0) + \frac{e_y}{p(y|x_0)} + O(e_y^2)\right] \\
& && (27)\\
&= D[p(\cdot|x)\|p(\cdot|x_0)] + \sum_{y\in\mathcal{Y}} \frac{p(y|x)}{p(y|x_0)} e_y \\
& && \text{(quadratic terms negligible)}\\
&= D[p(\cdot|x)\|p(\cdot|x_0)] && (28)\\
&\quad + \sum_{y\in\mathcal{Y}} \left(1+\frac{1}{p(y|x_0)}\nabla p(y|x_0)(x-x_0)+O(r^2)\right) e_y && (29)\\
&= D[p(\cdot|x)\|p(\cdot|x_0)] + \sum_{y\in\mathcal{Y}} \left[e_y + O(r^3)\right] && (e_y \in O(r^2))\\
&= D[p(\cdot|x)\|p(\cdot|x_0)] + O(r^3) && (\textstyle\sum_{y\in\mathcal{Y}} e_y = 0)
\end{aligned}
$$

□

Lemma 3. *For any positive-semidefinite matrix $A \in \mathbb{R}^{n\times n}$,*

$$\int_{B_r} \langle x-x_0, A(x-x_0)\rangle \, d^n x = \frac{n}{n+2} r^2 v(B_r)\mathrm{trace}(A)$$

Proof. Since A is positive-semidefinite, there exist an orthonormal matrix P and a diagonal matrix D such that $A = P^T D P$. Furthermore, the entries of D are the eigenvalues $\{\lambda_i\}$ of A. Then,

$$
\begin{aligned}
\int_{B_r} \langle x-x_0, A(x-x_0)\rangle \, d^n x &= \int_{B_r} \left\langle x-x_0, P^T D P(x-x_0)\right\rangle d^n x && (30)\\
&= \int_{B_r} \langle P(x-x_0), DP(x-x_0)\rangle \, d^n x\,. && (31)
\end{aligned}
$$

We can regard P as a reparameterization of B_r; since $\det P = 1$, we have

$$
\begin{aligned}
\int_{B_r} \langle P(x-x_0), DP(x-x_0)\rangle \, d^n x &= \int_{B_r} \langle x-x_0, D(x-x_0)\rangle \, d^n x && (32)\\
&= r^n \int_{B_n} \langle rx, rDx\rangle \, d^n x && (33)\\
&= r^{n+2} \int_{B_n} \langle x, Dx\rangle \, d^n x\,, && (34)
\end{aligned}
$$

where B_n is the unit n-ball. We also let $S_n(r)$ be the n-sphere of radius r. Continuing,

$$r^{n+2} \int_{B_n} \langle x, Dx \rangle \, d^n x = r^{n+2} \int_{B_n} \sum_{i=1}^{n} x_i^2 \lambda_i \, d^n x \tag{35}$$

$$= r^{n+2} \sum_{i=1}^{n} \lambda_i \int_{B_n} x_i^2 \, d^n x \tag{36}$$

$$= \frac{r^{n+2}}{n} \sum_{i=1}^{n} \lambda_i \int_{B_n} \|x\|^2 \; d^n x \tag{37}$$

(spherical symmetry)

$$= \frac{r^{n+2}}{n} \mathrm{trace}(A) \int_{B_n} \|x\|^2 \; d^n x \tag{38}$$

(spherical symmetry)

$$= \frac{r^{n+2}}{n} \mathrm{trace}(A) \int_{\rho \in [0,1]} \rho^2 S_{n-1}(\rho) d\rho \tag{39}$$

$$= \frac{r^{n+2}}{n} \mathrm{trace}(A) \int_{\rho \in [0,1]} \rho^{n+1} S_{n-1}(1) d\rho \tag{40}$$

$$= \frac{r^{n+2}}{n} \mathrm{trace}(A) \frac{1}{n+2} S_{n-1}(1) \tag{41}$$

$$= \frac{r^2}{n+2} \mathrm{trace}(A) n r^n v(B_n(1)) \tag{42}$$

$$= \frac{n}{n+2} r^2 v(B_r) \mathrm{trace}(A) \, , \tag{43}$$

as was to be shown. □

Fact. *The Kullback-Leibler divergence and the Fisher information J_Y are related according to the approximation*

$$D[p(\cdot|x) \| p(\cdot|x_0)] = \frac{1}{2} \langle x - x_0, J_Y(x_0)(x - x_0) \rangle + O(\|x - x_0\|^3) \tag{44}$$

We are finally ready to prove Theorem 2. Computing directly, we have

$$I_r(x_0) \triangleq \mathbb{E}_X[D[p(\cdot|X)\|p(\cdot|X \in B_r)]|X \in B_r] \; . \tag{45}$$

$$= \int_{B_r} p(x|X \in B_r) D[p(\cdot|x)\|p(\cdot|X \in B_r)] d^n x \tag{46}$$

$$= \int_{B_r} \left[\frac{1+O(r)}{v(B_r)}\right] D[p(\cdot|x)\|p(\cdot|X \in B_r)] d^n x \tag*{(Lemma 1(c))}$$

$$= \left[\frac{1+O(r)}{v(B_r)}\right] \int_{B_r} \left(D[p(\cdot|x)\|p(\cdot|x_0)] + O(r^3)\right) d^n x \tag*{(Lemma 2)}$$

$$= \left[\frac{1+O(r)}{v(B_r)}\right] \int_{B_r} \left(\frac{1}{2}\langle x - x_0, J_Y(x_0)(x-x_0)\rangle + O(\|x-x_0\|^3) + O(r^3)\right) d^n x \tag{47}$$

$$= \frac{1}{2}\left[\frac{1+O(r)}{v(B_r)}\right] \int_{B_r} \left(\langle x - x_0, J_Y(x_0)(x-x_0)\rangle + O(r^3)\right) d^n x \tag{48}$$

$$= \frac{1}{2}\left[\frac{1+O(r)}{v(B_r)}\right] \left(\frac{n}{n+2} r^2 v(B_r) \mathrm{trace}(J_Y(x_0)) + v(B_r) O(r^3)\right) \tag{49}$$

$$= r^2 \frac{n}{2(n+2)} \left[1 + O(r)\right] \left(\mathrm{trace}(J_Y(x_0)) + O(r^3)\right) \; . \tag{50}$$

Dividing through by r^2 and computing the limit as $r \to 0$ proves the result.

Computational Methods and Assumptions

In this section, we provide a specification of the computational procedure used to estimate $J(X, Y) = \mathbb{E}_x[J_Y(X)]$ using blockgroup level data from the U.S. Census.

For fixed Census blockgroup i, let P_i be the population, let A_i be the area, let $\rho_i = P_i/A_i$ be the population density, and let $p_Y^i(y)$ be the observed proportion of racial group y. For hex k in our hexagonal grid, let N_k be the set of overlapping Census blockgroups. We also define $p_I^k(i) = \rho_i / \sum_{i \in N_k} \rho_i$ as the estimated proportion of population within hex k residing in blockgroup i. This definition embodies a computationally-simplifying assumption that each blockgroup in N_k overlaps hex k with equal area. Finally, $p_Y^k(y) = \sum_{i \in N_k} p_I^k(i) p_Y^i(y)$ is the estimated overall racial composition of hex k. Then, we estimate the mutual information in hex k as

$$I(k) = \sum_{i \in N_k} p_I^k(i) D[p_Y^i(\cdot)\|p_Y^k(\cdot)] \; . \tag{51}$$

Using (**??**), the estimated Fisher information is

$$J(k) \approx \frac{4I(k)}{r^2} \tag{52}$$

where r is the grid radius. The estimated population in hex k is $P_k = A_k \sum_{i \in N_k} \rho_i$, where A_k is the cell area. We finally estimate $\mathbb{E}_X[J(X)]$ as

$$J(X, Y) = \mathbb{E}_X[J_Y(X)] \approx \frac{1}{\sum_k P_k} \sum_k P_k J(k) \tag{53}$$

Mutual and Fisher Informations: Part 2

Let X be a continuous random variable taking values in $\mathbb{R}^n$, and let Y be a discrete random variable define on finite alphabet $\mathcal{Y}$. Suppose further that $p(y|x) > 0$ and that $p(y|x)$ is differentiable as a function of x for all $x, y \in \mathbb{R}^n \times \mathcal{Y}$. Fix $x_0 \in \mathbb{R}^n$, and define $B_r \triangleq B_r(x_0) = \{x \in R^n \mid \|x - x_0\| \leq r\}$. Additionally, define the *local mutual information* in B_r as the mutual information between X and Y where X is restricted to B_r:

$$I_r(x_0) \triangleq \mathbb{E}_X[D[p(\cdot|X)\|p(\cdot|X \in B_r)]|X \in B_r] \tag{54}$$

$$= \int_{B_r} p(x|X \in B_r) D[p(\cdot|x)\|p(\cdot|X \in B_r)] d^n x \ . \tag{55}$$

where $D[p\|q] \triangleq \sum_y p(y) \log \frac{p(y)}{q(y)}$ is the Kullback-Leibler divergence of q from p.

Theorem 2. *Under the stated conditions,*

$$\lim_{r \to 0} \frac{I_r(x_0)}{r^2} = \frac{n}{2(n+2)} \text{trace } J_Y(x_0) \ . \tag{56}$$

where the Fisher information matrix J_Y is given by

$$J_Y(x) \triangleq \mathbb{E}_Y \left[\nabla_x S_Y(x) \nabla_x S_Y(x)^T \right] \tag{57}$$

$$S_y(x) \triangleq \log p(y|x) \ . \tag{58}$$

We first define the following functions. For set A, let

$$f_A(x) \triangleq D[p(y|x)\|p(y|x \in A)] \ . \tag{59}$$

Additionally, let

$$q_r(x) \triangleq p(x|X \in B_r) \ . \tag{60}$$

We can then write the local mutual information as

$$I_r(x_0) = \int_{B_r} q_r f_{B_r} \, d\lambda \ , \tag{61}$$

where λ is the Lebesgue measure in $\mathbb{R}^n$. We also let $V_r \triangleq \int_{B_r} d\lambda$ be the volume of B_r, which we will explicitly compute later. Finally, define

$$a_r(y) \triangleq p(X \in B_r, Y). \tag{62}$$

Lemma 4. *For all $x \in B_r$,*

$$q_r(x) = \frac{1 + O(r)}{V_r} \ . \tag{63}$$

Proof. First, Taylor's theorem implies that there exists a linear map T such that

$$p(x) = p(x_0) + T(x - x_0) + O(\|x - x_0\|^2) \tag{64}$$

$$= p(x_0) + T(x - x_0) + O(r^2)\ . \tag{65}$$

Then,

$$p(X \in B_r) = \int_{B_r} p\ d\lambda \tag{66}$$

$$= \int_{B_r} p(x_0) + T(x - x_0) + O(r^2) d^n x\ . \tag{67}$$

The middle term vanishes by spherical symmetry, leaving

$$p(X \in B_r) = p(x_0)V_r + O(r^{n+2})\ . \tag{68}$$

For $x \in B_r$, we therefore have

$$p(x|X \in B_r) = \frac{p(x, X \in B_r)}{p(X \in B_r)} \tag{69}$$

$$= \frac{p(x)}{p(X \in B_r)} \tag{70}$$

$$= \frac{p(x_0) + T(x - x_0) + O(r^2)}{p(x_0)V_r + O(r^{n+2})} \tag{71}$$

$$= \frac{p(x_0) + O(r) + O(r^2)}{p(x_0)V_r} \tag{72}$$

$$= \frac{1 + O(r)}{V_r}\ , \tag{73}$$

as was to be shown. □

Lemma 5. *For all $y \in Y$,*

$$a_r(y) = p(x_0, y)V_r + O(r^{n+2})\ . \tag{74}$$

Proof. Taylor's Theorem again implies that, for any $(x, y) \in B_r \times \mathcal{Y}$, there is a linear map T such that

$$p(x, y) = p(x_0, y) + T(x - x_0) + O(\|x - x_0\|^2) \tag{75}$$

$$= p(x_0, y) + T(x - x_0) + O(r^2). \tag{76}$$

We then have

$$p(X \in B_r, y) = \int_{B_r} p(x, y)\ d^n x \tag{77}$$

$$= \int_{B_r} p(x_0, y) + T(x - x_0) + O(r^2)\ d^n x \tag{78}$$

$$= p(x_0, y)V_r + O(r^{n+2})\ d^n x\ , \tag{79}$$

where the middle term has again vanished due to spherical symmetry. □

Lemma 6. *For any $y \in \mathcal{Y}$, we have $p(y|X \in B_r) = p(y|x_0) + e_y$, where the error terms e_y satisfy $e_y \in O(r^2)$ and $\sum_{y \in \mathcal{Y}} e_y = 0$.*

Proof. That the errors must satisfy $\sum_{y \in \mathcal{Y}} e_y = 0$ follows from the fact that $p(\cdot|X \in B_r)$ must be a valid probability distribution on $\mathcal{Y}$. We'll now show that $e_y \in O(r^2)$. We then have

$$
\begin{aligned}
p(y|X \in B_r) &= \frac{p(X \in B_r, Y)}{p(X \in B_r)} && (80) \\
&= \frac{p(x_0, y)V_r + O(r^{n+2})}{p(x_0)V_r + O(r^{n+2})} && (81) \\
&= p(y|x_0)O(r^2)\ , && (V_r \propto r^n)
\end{aligned}
$$

as needed. □

0.1 *Specification of Spatially-Constrained Information-Theoretic Clustering*