Nonparametric Indepedence Tests: Space Partitioning and Kernel Approaches

Arthur Gretton[1] and László Györfi[2]

1. Max Planck Institute for Biological Cybernetics
Tübingen, Germany
2. Budapest University of Technology and Economics
Budapest, Hungary

October, 2008, ALT

MAX-PLANCK-GESELLSCHAFT

BIOLOGISCHE KYBERNETIK

Introduction

- Statistical tests of independence
  - Multivariate
  - Nonparametric
- Two kinds of tests:
  - Strongly consistent, distribution-free
  - Asymptotically α-level
- Three test statistics:
  - L_1 distance
  - Log-likelihood
  - Kernel dependence measure (HSIC)

Introduction

- Statistical tests of independence
  - Multivariate
  - Nonparametric
- Two kinds of tests:
  - Strongly consistent, distribution-free
  - Asymptotically α-level
- Three test statistics:
  - L_1 distance
  - Log-likelihood
  - Kernel dependence measure (HSIC)

Introduction

- Statistical tests of independence
  - Multivariate
  - Nonparametric
- Two kinds of tests:
  - Strongly consistent, distribution-free
  - Asymptotically α-level
- Three test statistics:
  - L_1 distance
  - Log-likelihood
  - Kernel dependence measure (HSIC)

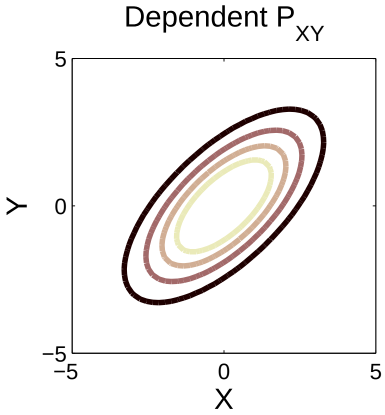
Overview...

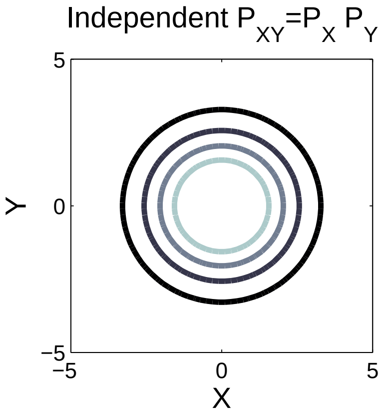
- Problem overview
- L_1 statistic
- Log-likelihood statistic
- Kernel statistic
- Experimental results

- **Problem overview**
- L_1 statistic
- Log-likelihood statistic
- Kernel statistic
- Experimental results

Problem overview

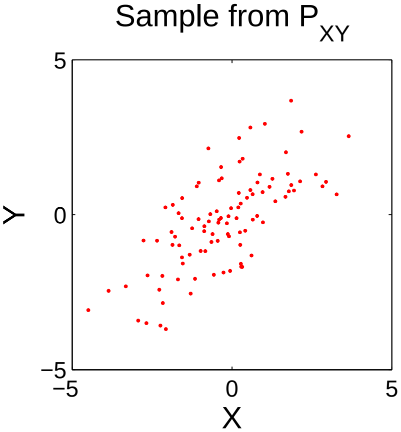
- Given distribution Pr, test $\mathcal{H}_0 \;:\; \mathrm{Pr} = \mathrm{Pr}_x \, \mathrm{Pr}_y$
- Continuous valued, multivariate: $\mathcal{X} := \mathbb{R}^d$ and $\mathcal{Y} := \mathbb{R}^{d'}$





Problem overview

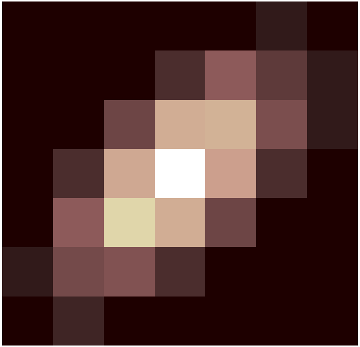
- Given distribution Pr, test $\mathcal{H}_0 \; : \; \mathrm{Pr} = \mathrm{Pr}_x \, \mathrm{Pr}_y$
- Finite sample observed $(X_1, Y_1), \ldots, (X_n, Y_n)$



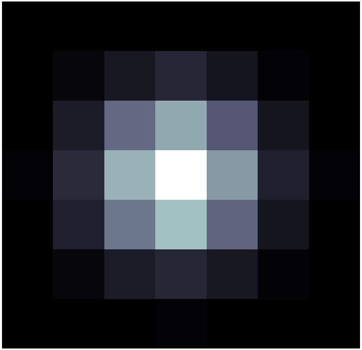
Problem overview

- Given distribution Pr, test $\mathcal{H}_0 \; : \; \Pr = \Pr_x \Pr_y$
- Partition space $\mathcal{X}$ into m_n bins, space $\mathcal{Y}$ into m'_n bins

Discretized empirical P_{XY}



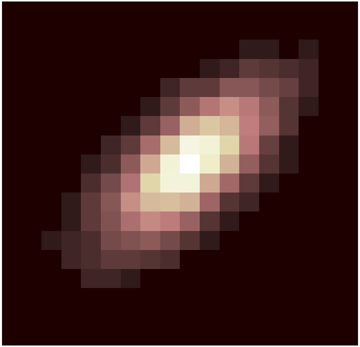
Discretized empirical $P_X \, P_Y$



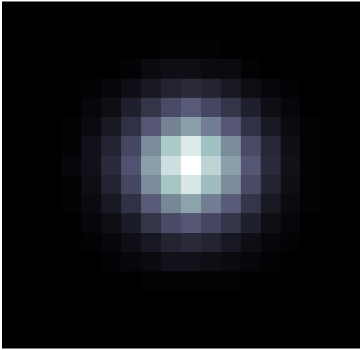
Problem overview

- Given distribution Pr, test $\mathcal{H}_0 \; : \; \mathrm{Pr} = \mathrm{Pr}_x \, \mathrm{Pr}_y$
- Refine partition m_n, m'_n for increasing n

Discretized empirical P_{XY}



Discretized empirical $P_X \, P_Y$



Overview...

- Problem overview
- L_1 **statistic**
- Log-likelihood statistic
- Kernel statistic
- Experimental results

L_1 statistic

- Space partition
  - $\mathcal{P}_n = \{A_{n,1}, \ldots, A_{n,m_n}\}$ of $\mathcal{X} = \mathbb{R}^d$
  - $\mathcal{Q}_n = \{B_{n,1}, \ldots, B_{n,m'_n}\}$ of $\mathcal{Y} = \mathbb{R}^{d'}$,
- Empirical measures
  - $\nu_n(A \times B) = n^{-1}\#\{i : (X_i, Y_i) \in A \times B, i = 1, \ldots, n\}$
  - $\mu_{n,1}(A) = n^{-1}\#\{i : X_i \in A, i = 1, \ldots, n\}$
  - $\mu_{n,2}(B) = n^{-1}\#\{i : Y_i \in B, i = 1, \ldots, n\}$
- Test statistic:

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) = \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\nu_n(A \times B) - \mu_{n,1}(A) \cdot \mu_{n,2}(B)|.$$

L_1 statistic

- Space partition
  - $\mathcal{P}_n = \{A_{n,1}, \ldots, A_{n,m_n}\}$ of $\mathcal{X} = \mathbb{R}^d$
  - $Q_n = \{B_{n,1}, \ldots, B_{n,m'_n}\}$ of $\mathcal{Y} = \mathbb{R}^{d'}$,
- Empirical measures
  - $\nu_n(A \times B) = n^{-1}\#\{i : (X_i, Y_i) \in A \times B, i = 1, \ldots, n\}$
  - $\mu_{n,1}(A) = n^{-1}\#\{i : X_i \in A, i = 1, \ldots, n\}$
  - $\mu_{n,2}(B) = n^{-1}\#\{i : Y_i \in B, i = 1, \ldots, n\}$
- Test statistic:

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) = \sum_{A \in \mathcal{P}_n} \sum_{B \in Q_n} |\nu_n(A \times B) - \mu_{n,1}(A) \cdot \mu_{n,2}(B)|.$$

L_1 statistic

- Space partition
  - $\mathcal{P}_n = \{A_{n,1}, \ldots, A_{n,m_n}\}$ of $\mathcal{X} = \mathbb{R}^d$
  - $Q_n = \{B_{n,1}, \ldots, B_{n,m'_n}\}$ of $\mathcal{Y} = \mathbb{R}^{d'}$,
- Empirical measures
  - $\nu_n(A \times B) = n^{-1}\#\{i : (X_i, Y_i) \in A \times B, i = 1, \ldots, n\}$
  - $\mu_{n,1}(A) = n^{-1}\#\{i : X_i \in A, i = 1, \ldots, n\}$
  - $\mu_{n,2}(B) = n^{-1}\#\{i : Y_i \in B, i = 1, \ldots, n\}$
- Test statistic:

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) = \sum_{A \in \mathcal{P}_n} \sum_{B \in Q_n} |\nu_n(A \times B) - \mu_{n,1}(A) \cdot \mu_{n,2}(B)|.$$

L_1: Distribution-free strong consistent test

- Based on large deviation bound: under $\mathcal{H}_0$, for all $0 < \varepsilon_1$, $0 < \varepsilon_2$ and $0 < \varepsilon_3$,

$$\Pr\{L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > \varepsilon_1 + \varepsilon_2 + \varepsilon_3\} \\ \leq 2^{m_n \cdot m'_n} e^{-n\varepsilon_1^2/2} + 2^{m_n} e^{-n\varepsilon_2^2/2} + 2^{m'_n} e^{-n\varepsilon_3^2/2}.$$

- Require conditions
  - $\lim_{n\to\infty} m_n m'_n / n = 0$,
  - $\lim_{n\to\infty} m_n / \ln n = \infty$, $\lim_{n\to\infty} m'_n / \ln n = \infty$,
- Test: reject $\mathcal{H}_0$ when

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > c_1 \sqrt{\frac{m_n m'_n}{n}}, \qquad \text{where } c_1 > \sqrt{2 \ln 2}.$$

  - Distribution-free
  - Strongly consistent: after random sample size, test makes a.s. no error

L_1: Distribution-free strong consistent test

- Based on large deviation bound: under $\mathcal{H}_0$, for all $0 < \varepsilon_1$, $0 < \varepsilon_2$ and $0 < \varepsilon_3$,

$$\Pr\{L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > \varepsilon_1 + \varepsilon_2 + \varepsilon_3\} \\ \leq 2^{m_n \cdot m'_n} e^{-n\varepsilon_1^2/2} + 2^{m_n} e^{-n\varepsilon_2^2/2} + 2^{m'_n} e^{-n\varepsilon_3^2/2}.$$

- Require conditions
  - $\lim_{n\to\infty} m_n m'_n / n = 0$,
  - $\lim_{n\to\infty} m_n / \ln n = \infty$, $\quad \lim_{n\to\infty} m'_n / \ln n = \infty$,
- Test: reject $\mathcal{H}_0$ when

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > c_1 \sqrt{\frac{m_n m'_n}{n}}, \qquad \text{where } c_1 > \sqrt{2 \ln 2}.$$

  - Distribution-free
  - Strongly consistent: after random sample size, test makes a.s. no error

L_1: Distribution-free strong consistent test

- Based on large deviation bound: under $\mathcal{H}_0$, for all $0 < \varepsilon_1$, $0 < \varepsilon_2$ and $0 < \varepsilon_3$,

$$\Pr\{L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > \varepsilon_1 + \varepsilon_2 + \varepsilon_3\} \\ \leq 2^{m_n \cdot m'_n} e^{-n\varepsilon_1^2/2} + 2^{m_n} e^{-n\varepsilon_2^2/2} + 2^{m'_n} e^{-n\varepsilon_3^2/2}.$$

- Require conditions
  - $\lim_{n\to\infty} m_n m'_n / n = 0$,
  - $\lim_{n\to\infty} m_n / \ln n = \infty$, $\lim_{n\to\infty} m'_n / \ln n = \infty$,
- Test: reject $\mathcal{H}_0$ when

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > c_1 \sqrt{\frac{m_n m'_n}{n}}, \qquad \text{where } c_1 > \sqrt{2 \ln 2}.$$

  - Distribution-free
  - Strongly consistent: after random sample size, test makes a.s. no error

L_1: Asymptotic α-level test

- Additional conditions:

$$\lim_{n\to\infty} \max_{A\in\mathcal{P}_n} \mu_1(A) = 0, \quad \lim_{n\to\infty} \max_{B\in Q_n} \mu_2(B) = 0,$$

- Asymptotic distribution: Under $\mathcal{H}_0$,

$$\sqrt{n}\left(L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) - C_n\right)/\sigma \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad \sigma^2 = 1 - 2/\pi.$$

  - Upper bound: $C_n \leq \sqrt{2m_n m_n'/(\pi n)}$
- Test: reject $\mathcal{H}_0$ when

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > \sqrt{2m_n m_n'/(\pi n)} + \sigma/\sqrt{n}\,\Phi^{-1}(1-\alpha)$$

  - Φ denotes standard normal distribution.
- Properties of test:
  - Not distribution-free (nonatomic)
  - α level: probability of Type I error asymptotically $\leq \alpha$

L_1: Asymptotic α-level test

- Additional conditions:

$$\lim_{n\to\infty} \max_{A\in\mathcal{P}_n} \mu_1(A) = 0, \quad \lim_{n\to\infty} \max_{B\in Q_n} \mu_2(B) = 0,$$

- Asymptotic distribution: Under $\mathcal{H}_0$,

$$\sqrt{n}\left(L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) - C_n\right)/\sigma \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad \sigma^2 = 1 - 2/\pi.$$

  - Upper bound: $C_n \leq \sqrt{2m_n m'_n/(\pi n)}$

- Test: reject $\mathcal{H}_0$ when

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > \sqrt{2m_n m'_n/(\pi n)} + \sigma/\sqrt{n}\,\Phi^{-1}(1-\alpha)$$

  - Φ denotes standard normal distribution.

- Properties of test:
  - Not distribution-free (nonatomic)
  - α level: probability of Type I error asymptotically $\leq \alpha$

L_1: Asymptotic α-level test

- Additional conditions:

$$\lim_{n\to\infty} \max_{A\in\mathcal{P}_n} \mu_1(A) = 0, \quad \lim_{n\to\infty} \max_{B\in Q_n} \mu_2(B) = 0,$$

- Asymptotic distribution: Under $\mathcal{H}_0$,

$$\sqrt{n}\left(L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) - C_n\right)/\sigma \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad \sigma^2 = 1 - 2/\pi.$$

  - Upper bound: $C_n \le \sqrt{2m_n m_n'/(\pi n)}$

- Test: reject $\mathcal{H}_0$ when

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > \sqrt{2m_n m_n'/(\pi n)} + \sigma/\sqrt{n}\,\Phi^{-1}(1-\alpha)$$

  - Φ denotes standard normal distribution.

- Properties of test:
  - Not distribution-free (nonatomic)
  - α level: probability of Type I error asymptotically $\le \alpha$

L_1: Asymptotic α-level test

- Additional conditions:

$$\lim_{n\to\infty} \max_{A\in\mathcal{P}_n} \mu_1(A) = 0, \quad \lim_{n\to\infty} \max_{B\in Q_n} \mu_2(B) = 0,$$

- Asymptotic distribution: Under $\mathcal{H}_0$,

$$\sqrt{n}\left(L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) - C_n\right)/\sigma \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \quad \sigma^2 = 1 - 2/\pi.$$

  - Upper bound: $C_n \le \sqrt{2m_n m'_n/(\pi n)}$

- Test: reject $\mathcal{H}_0$ when

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) > \sqrt{2m_n m'_n/(\pi n)} + \sigma/\sqrt{n}\,\Phi^{-1}(1-\alpha)$$

  - Φ denotes standard normal distribution.

- Properties of test:
  - Not distribution-free (nonatomic)
  - α level: probability of Type I error asymptotically $\le \alpha$

Overview...

- Problem overview
- L_1 statistic
- **Log-likelihood statistic**
- Kernel statistic
- Experimental results

Log likelihood: test statistic

- Test statistic:

$$I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) = 2 \sum_{A \in \mathcal{P}_n} \sum_{B \in Q_n} \nu_n(A \times B) \log \frac{\nu_n(A \times B)}{\mu_{n,1}(A) \cdot \mu_{n,2}(B)}.$$

- Useful bounds:
  - Pinsker:

$$L_n^2(\nu_n, \mu_{n,1} \times \mu_{n,2}) \leq I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}).$$

  - Under $\mathcal{H}_0$

$$I_n(\nu_n, \nu) - I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) = I_n(\mu_{n,1}, \mu_1) + I_n(\mu_{n,2}, \mu_2) \geq 0.$$

- Relevant large deviation bound Kallenberg(1985), Quine and Robinson (1985)

$$\Pr\{I_n(\nu_n, \nu)/2 > \epsilon\} = e^{-n(\epsilon + o(1))}.$$

Log likelihood: test statistic

- Test statistic:

$$I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) = 2 \sum_{A \in \mathcal{P}_n} \sum_{B \in Q_n} \nu_n(A \times B) \log \frac{\nu_n(A \times B)}{\mu_{n,1}(A) \cdot \mu_{n,2}(B)}.$$

- Useful bounds:
  - Pinsker:

$$L_n^2(\nu_n, \mu_{n,1} \times \mu_{n,2}) \leq I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}).$$

  - Under $\mathcal{H}_0$

$$I_n(\nu_n, \nu) - I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) = I_n(\mu_{n,1}, \mu_1) + I_n(\mu_{n,2}, \mu_2) \geq 0.$$

- Relevant large deviation bound Kallenberg(1985), Quine and Robinson (1985)

$$\Pr\{I_n(\nu_n, \nu)/2 > \epsilon\} = e^{-n(\epsilon + o(1))}.$$

Log likelihood: test statistic

- Test statistic:

$$I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) = 2 \sum_{A \in \mathcal{P}_n} \sum_{B \in Q_n} \nu_n(A \times B) \log \frac{\nu_n(A \times B)}{\mu_{n,1}(A) \cdot \mu_{n,2}(B)}.$$

- Useful bounds:
  - Pinsker:

$$L_n^2(\nu_n, \mu_{n,1} \times \mu_{n,2}) \leq I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}).$$

  - Under $\mathcal{H}_0$

$$I_n(\nu_n, \nu) - I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) = I_n(\mu_{n,1}, \mu_1) + I_n(\mu_{n,2}, \mu_2) \geq 0.$$

- Relevant large deviation bound Kallenberg(1985), Quine and Robinson (1985)

$$\Pr\{I_n(\nu_n, \nu)/2 > \epsilon\} = e^{-n(\epsilon + o(1))}.$$

Log likelihood: tests

- Require earlier conditions on m_n, m_n', and n
- Test: reject $\mathcal{H}_0$ when

$$I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) \geq m_n m_n' n^{-1}(2\log(n + m_n m_n') + 1).$$

- Distribution-free, strongly consistent
- Asymptotically α-level test:

$$\Pr\left\{\frac{nI_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) - m_n m_n'}{\sqrt{2m_n m_n'}} \leq x\right\}$$

$$\leq \Pr\left\{\frac{nI_n(\nu_n, \nu) - m_n m_n'}{\sqrt{2m_n m_n'}} \leq x\right\} \to \Phi(x),$$

Log likelihood: tests

- Require earlier conditions on m_n, m'_n, and n
- Test: reject $\mathcal{H}_0$ when

$$I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) \geq m_n m'_n n^{-1}(2\log(n + m_n m'_n) + 1).$$

- Distribution-free, strongly consistent
- Asymptotically α-level test:

$$\Pr\left\{\frac{nI_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) - m_n m'_n}{\sqrt{2m_n m'_n}} \leq x\right\}$$
$$\leq \Pr\left\{\frac{nI_n(\nu_n, \nu) - m_n m'_n}{\sqrt{2m_n m'_n}} \leq x\right\} \to \Phi(x),$$

Log likelihood: tests

- Require earlier conditions on m_n, m'_n, and n
- Test: reject $\mathcal{H}_0$ when

$$I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) \geq m_n m'_n n^{-1}(2 \log(n + m_n m'_n) + 1).$$

- Distribution-free, strongly consistent
- Asymptotically α-level test:

$$\Pr\left\{ \frac{n I_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) - m_n m'_n}{\sqrt{2 m_n m'_n}} \leq x \right\}$$
$$\leq \Pr\left\{ \frac{n I_n(\nu_n, \nu) - m_n m'_n}{\sqrt{2 m_n m'_n}} \leq x \right\} \to \Phi(x),$$

Overview...

- Problem overview
- L_1 statistic
- Log-likelihood statistic
- **Kernel statistic**
- Experimental results

Kernel statistic

- $\mathcal{F}$ RKHS on $\mathcal{X}$ with kernel $k(x_i, x_j)$, $\mathcal{F}'$ RKHS on $\mathcal{Y}$ with kernel $k'(y_i, y_j)$
- $\mathcal{F} = \overline{\mathrm{span}\{k(x, \cdot) | x \in \mathcal{X}\}}$
- Example: $f(\cdot) = \sum_{i=1}^{m} \alpha_i k(x_i, \cdot)$ for arbitrary $m \in \mathbb{N}$, $\alpha_i \in \mathbb{R}$, $x_i \in \mathcal{X}$.
- Covariance operator: $C_{xy} : \mathcal{F}' \to \mathcal{F}$ such that

$$\langle f, C_{xy} g \rangle_{\mathcal{F}} = \mathbf{E}_{\mathrm{x,y}}[f(\mathrm{x})g(\mathrm{y})] - \mathbf{E}_{\mathrm{x}}[f(\mathrm{x})]\mathbf{E}_{\mathrm{y}}[g(\mathrm{y})]$$

- HSIC is the Hilbert-Schmidt norm of C_{xy}:

$$\mathrm{HSIC} := \|C_{xy}\|_{\mathrm{HS}}^2$$

- $\mathrm{HSIC} = 0$ iff $\Pr = \Pr_X \Pr_Y$ when kernels $k(\cdot, \cdot)$ and $k'(\cdot, \cdot)$ characteristic Fukumizu et al. (2008), Sriperumbudur et al. (2008)
  - Examples on $\mathbb{R}^d$: Gaussian, Laplace, B-spline,...

Kernel statistic

- $\mathcal{F}$ RKHS on $\mathcal{X}$ with kernel $k(x_i, x_j)$, $\mathcal{F}'$ RKHS on $\mathcal{Y}$ with kernel $k'(y_i, y_j)$
- $\mathcal{F} = \overline{\mathrm{span}\{k(x,\cdot) | x \in \mathcal{X}\}}$
- Example: $f(\cdot) = \sum_{i=1}^{m} \alpha_i k(x_i, \cdot)$ for arbitrary $m \in \mathbb{N}$, $\alpha_i \in \mathbb{R}$, $x_i \in \mathcal{X}$.
- Covariance operator: $C_{xy} : \mathcal{F}' \to \mathcal{F}$ such that

$$\langle f, C_{xy} g \rangle_{\mathcal{F}} = \mathbf{E}_{\mathrm{x,y}}[f(\mathrm{x})g(\mathrm{y})] - \mathbf{E}_{\mathrm{x}}[f(\mathrm{x})]\mathbf{E}_{\mathrm{y}}[g(\mathrm{y})]$$

- HSIC is the Hilbert-Schmidt norm of C_{xy}:

$$\mathrm{HSIC} := \| C_{xy} \|^2_{\mathrm{HS}}$$

- $\mathrm{HSIC} = 0$ iff $\Pr = \Pr_X \Pr_Y$ when kernels $k(\cdot,\cdot)$ and $k'(\cdot,\cdot)$ characteristic Fukumizu et al. (2008), Sriperumbudur et al. (2008)
  - Examples on $\mathbb{R}^d$: Gaussian, Laplace, B-spline,...

Kernel statistic

- $\mathcal{F}$ RKHS on $\mathcal{X}$ with kernel $k(x_i, x_j)$, $\mathcal{F}'$ RKHS on $\mathcal{Y}$ with kernel $k'(y_i, y_j)$
- $\mathcal{F} = \overline{\mathrm{span}\{k(x,\cdot)|x \in \mathcal{X}\}}$
- Example: $f(\cdot) = \sum_{i=1}^{m} \alpha_i k(x_i, \cdot)$ for arbitrary $m \in \mathbb{N}$, $\alpha_i \in \mathbb{R}$, $x_i \in \mathcal{X}$.
- Covariance operator: $C_{xy} : \mathcal{F}' \to \mathcal{F}$ such that

$$\langle f, C_{xy} g \rangle_{\mathcal{F}} = \mathbf{E}_{\mathrm{x,y}}[f(\mathrm{x})g(\mathrm{y})] - \mathbf{E}_{\mathrm{x}}[f(\mathrm{x})]\mathbf{E}_{\mathrm{y}}[g(\mathrm{y})]$$

- HSIC is the Hilbert-Schmidt norm of C_{xy}:

$$\mathrm{HSIC} := \|C_{xy}\|_{\mathrm{HS}}^2$$

- $\mathrm{HSIC} = 0$ iff $\Pr = \Pr_X \Pr_Y$ when kernels $k(\cdot,\cdot)$ and $k'(\cdot,\cdot)$ characteristic Fukumizu et al. (2008), Sriperumbudur et al. (2008)
  - Examples on $\mathbb{R}^d$: Gaussian, Laplace, B-spline,...

Kernel statistic

- $\mathcal{F}$ RKHS on $\mathcal{X}$ with kernel $k(x_i, x_j)$, $\mathcal{F}'$ RKHS on $\mathcal{Y}$ with kernel $k'(y_i, y_j)$
- $\mathcal{F} = \overline{\mathrm{span}\{k(x, \cdot) | x \in \mathcal{X}\}}$
- Example: $f(\cdot) = \sum_{i=1}^{m} \alpha_i k(x_i, \cdot)$ for arbitrary $m \in \mathbb{N}$, $\alpha_i \in \mathbb{R}$, $x_i \in \mathcal{X}$.
- Covariance operator: $C_{xy} : \mathcal{F}' \to \mathcal{F}$ such that

$$\langle f, C_{xy} g \rangle_{\mathcal{F}} = \mathbf{E}_{\mathrm{x,y}}[f(\mathrm{x})g(\mathrm{y})] - \mathbf{E}_{\mathrm{x}}[f(\mathrm{x})]\mathbf{E}_{\mathrm{y}}[g(\mathrm{y})]$$

- HSIC is the Hilbert-Schmidt norm of C_{xy}:

$$\mathrm{HSIC} := \|C_{xy}\|^2_{\mathrm{HS}}$$

- $\mathrm{HSIC} = 0$ iff $\Pr = \Pr_X \Pr_Y$ when kernels $k(\cdot, \cdot)$ and $k'(\cdot, \cdot)$ characteristic Fukumizu et al. (2008), Sriperumbudur et al. (2008)
  - Examples on $\mathbb{R}^d$: Gaussian, Laplace, B-spline,...

Distribution-free strong consistent test

- Assume $k(x, \cdot) = k(|x - \cdot|)$
- (Biased) empirical HSIC:

$$T_n := \frac{1}{n^2}\mathrm{tr}(KHK'H)$$

  - $K_{i,j} = k(x_i, x_j)$, centering $H = I - \frac{1}{n}1_n 1_n^\top$
- Distribution-free, strong consistent test: reject $\mathcal{H}_0$ when

$$T_n > k(0)k'(0)(\sqrt{4\ln n} + 1)^2(nh^d h^{d'})^{-1}.$$

  - Scaling parameter $k_h(x) = h^{-d}k(x/h)$

Distribution-free strong consistent test

- Assume $k(x,\cdot) = k(|x - \cdot|)$
- (Biased) empirical HSIC:

$$T_n := \frac{1}{n^2}\mathrm{tr}(KHK'H)$$

  - $K_{i,j} = k(x_i, x_j)$, centering $H = I - \frac{1}{n}1_n 1_n^\top$
- Distribution-free, strong consistent test: reject $\mathcal{H}_0$ when

$$T_n > k(0)k'(0)(\sqrt{4\ln n} + 1)^2 (nh^d h^{d'})^{-1}.$$

  - Scaling parameter $k_h(x) = h^{-d} k(x/h)$

Asymptotic α-level tests

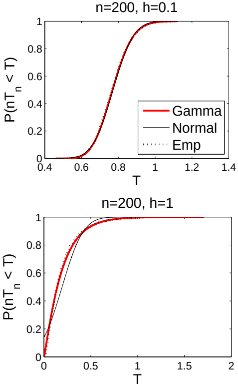
- Asymptotic distribution under $\mathcal{H}_0$:

- Gaussian using first two moments (decreasing bandwidth) Hall (1984), Cotterill and Csörgő (1985)

- Two parameter Gamma using first two moments (fixed bandwidth) Kankainen (1995), Gretton (2008)

$$nT_n \sim \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}$$

$$\alpha = \frac{(\mathbf{E}(T_n))^2}{\mathrm{var}(T_n)}, \quad \beta = \frac{n\mathrm{var}(T_n)}{\mathbf{E}(T_n)}.$$



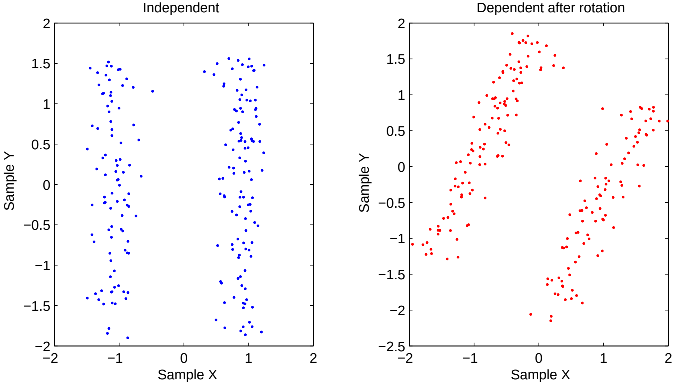
- Threshold NOT distribution-free: moments under $\mathcal{H}_0$ computed using sample

Overview...

- Problem overview
- L_1 statistic
- Log-likelihood statistic
- Kernel statistic
- **Experimental results**

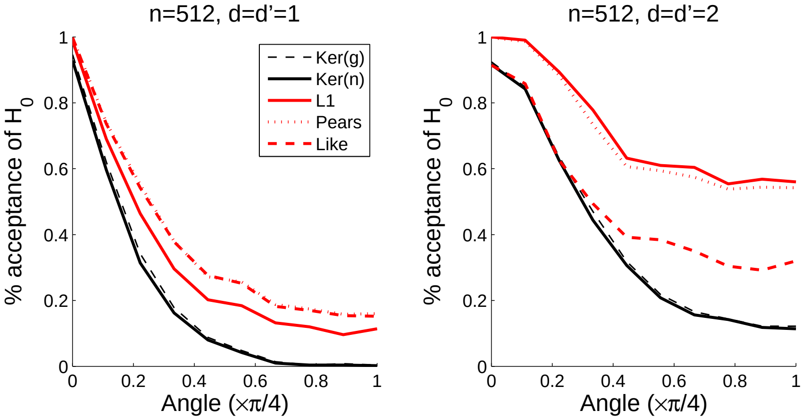
Experimental results

- Experimental setup:
  - Independent sample $(X_1, Y_1), \ldots, (X_n, Y_n)$
  - Rotate sample to create dependence
  - Embed randomly in $\mathbb{R}^d$, remaining subspace Gaussian noise



Experimental results

- Comparison of asymptotically α-level tests
- Given distribution Pr, test $\mathcal{H}_0 : \Pr = \Pr_x \Pr_y$
- Continuous valued, multivariate: $\mathcal{X} := \mathbb{R}^d$ and $\mathcal{Y} := \mathbb{R}^{d'}$



Conclusions

- Multi-dimensional nonparametric independence tests
  - Strongly consistent, distribution-free
  - Asymptotically α-level
- Test statistics
  - L_1 distance
  - Log likelihood
  - Kernel dependence measure (HSIC)
- In experiments:
- Kernel tests have better performance, but...
- ...threshold estimated from sample
- Further work:
  - Distribution-free threshold for kernel α-level test
  - Consistent null distribution estimate for kernel α-level test
  - Proof of χ^2 test conjecture