Nonparametric Indepedence Tests: Space Partitioning and Kernel Approaches

Arthur Gretton[1] and László Györfi[2]

1. Gatsby Computational Neuroscience Unit
London, UK
2. Budapest University of Technology and Economics
Budapest, Hungary

September, 2010

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

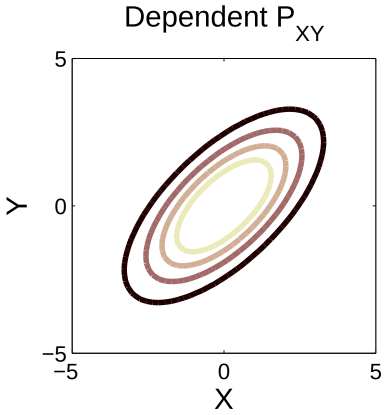
- Statistical tests of independence
  - Multivariate
  - Nonparametric
- Two kinds of tests:
  - Strongly consistent, distribution-free
  - Asymptotically α-level
- Three test statistics:
  - L_1 distance
  - Log-likelihood
  - Kernel dependence measure (HSIC)

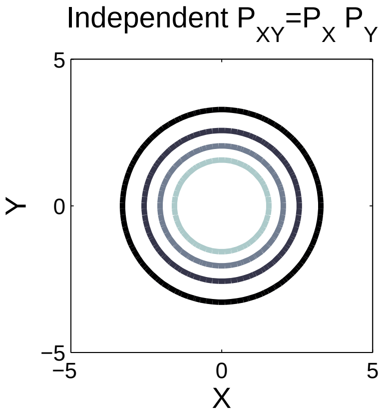
Introduction

- 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)

Problem overview

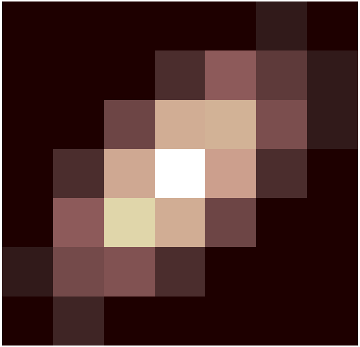
- Given distribution Pr, test $\mathcal{H}_0 \;:\; \Pr = \Pr_x \Pr_y$
- Continuous valued, multivariate: $\mathcal{X} := \mathbb{R}^d$ and $\mathcal{Y} := \mathbb{R}^{d'}$



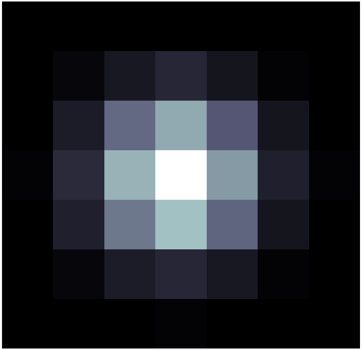


- Finite i.i.d. sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from Pr
- Partition space $\mathcal{X}$ into m_n bins, space $\mathcal{Y}$ into m'_n bins

Discretized empirical P_{XY}



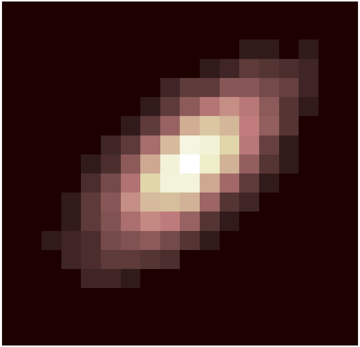
Discretized empirical $P_X \, P_Y$



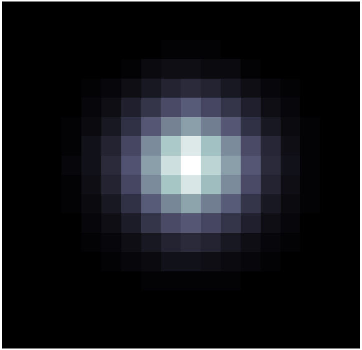
Problem overview

- Finite i.i.d. sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from Pr
- Refine partition m_n, m'_n for increasing n

Discretized empirical P_{XY}



Discretized empirical $P_X P_Y$



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 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

- 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

- Conditions
  - $\lim_{n\to\infty} m_n m'_n / n = 0$,
  - $\lim_{n\to\infty} m_n / \ln n = \infty, \qquad \lim_{n\to\infty} m'_n / \ln n = \infty$,
  - $\mathcal{P}_n$ and $\mathcal{Q}_n$ cells shrink:

$$\lim_{n\to\infty} \max_{A \in \mathcal{P}_n,\, A \cap S \neq 0} \operatorname{diam}(A) = 0$$

S any sphere centred at origin

L_1: Distribution-free strong consistent test

- 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
- Conditions
  - $\lim_{n\to\infty} m_n m_n'/n = 0$,
  - $\lim_{n\to\infty} m_n/\ln n = \infty$, $\qquad \lim_{n\to\infty} m_n'/\ln n = \infty$,
  - $\mathcal{P}_n$ and $\mathcal{Q}_n$ cells shrink:

$$\lim_{n\to\infty} \max_{A\in\mathcal{P}_n,\, A\cap S\neq 0} \operatorname{diam}(A) = 0$$

S any sphere centred at origin

L_1: Distribution-free strong consistent test

Type I error:

- Start with goodness of fit

$$L_n(\mu_{n,1}, \mu_1) = \sum_{A \in \mathcal{P}_n} |\mu_{n,1}(A) - \mu_1(A)|.$$

- Associated bound Beirlant et al. (2001), Biau and Györfi (2005)

$$\Pr\{L_n(\mu_{n,1}, \mu_1) > \varepsilon\} \leq 2^{m_n} e^{-n\varepsilon^2/2} \qquad \text{for all } 0 < \varepsilon$$

- Then decompose:

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

$$+ \underbrace{\sum_{A \in \mathcal{P}_n} \sum_{B \in Q_n} |\nu(A \times B) - \mu_1(A) \cdot \mu_2(B)|}_{=0 \text{ under } \mathcal{H}_0} + \underbrace{\sum_{A \in \mathcal{P}_n} \sum_{B \in Q_n} |\mu_1(A) \cdot \mu_2(B) - \mu_{n,1}(A) \cdot \mu_{n,2}(B)|}_{\leq L_n(\mu_{n,1}, \mu_1) + L_n(\mu_{n,2}, \mu_2)}.$$

L_1: Distribution-free strong consistent test

Type I error:

- Start with goodness of fit

$$L_n(\mu_{n,1}, \mu_1) = \sum_{A \in \mathcal{P}_n} |\mu_{n,1}(A) - \mu_1(A)|.$$

- Associated bound Beirlant et al. (2001), Biau and Györfi (2005)

$$\Pr\{L_n(\mu_{n,1}, \mu_1) > \varepsilon\} \le 2^{m_n} e^{-n\varepsilon^2/2} \qquad \text{for all } 0 < \varepsilon$$

- Then decompose:

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

$$+ \underbrace{\sum_{A \in \mathcal{P}_n} \sum_{B \in Q_n} |\nu(A \times B) - \mu_1(A) \cdot \mu_2(B)|}_{=0 \text{ under } \mathcal{H}_0} + \underbrace{\sum_{A \in \mathcal{P}_n} \sum_{B \in Q_n} |\mu_1(A) \cdot \mu_2(B) - \mu_{n,1}(A) \cdot \mu_{n,2}(B)|}_{\le L_n(\mu_{n,1}, \mu_1) + L_n(\mu_{n,2}, \mu_2)}.$$

L_1: Distribution-free strong consistent test

Type II error:

- Decompose

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) \geq \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\nu(A \times B) - \mu_1(A) \cdot \mu_2(B)|$$
$$\underbrace{- \sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\nu_n(A \times B) - \nu(A \times B)| - \sum_{B \in \mathcal{Q}_n} |\mu_2(B) - \mu_{n,2}(B)| - \sum_{A \in \mathcal{P}_n} |\mu_1(A) - \mu_{n,1}(A)|}_{\to 0 \text{ a.s.}}.$$

- Given cells shrink in $\mathcal{P}_n$ and $\mathcal{Q}_n$,

$$\sum_{A \in \mathcal{P}_n} \sum_{B \in \mathcal{Q}_n} |\nu(A \times B) - \mu_1(A) \cdot \mu_2(B)| \to 2 \sup_C |\nu(C) - \mu_1 \times \mu_2(C)| > 0$$

  - C Borel subset of $\mathbb{R}^d \times \mathbb{R}^{d'}$; see Barron et al. (1992)

L_1: Distribution-free strong consistent test

Type II error:

- Decompose

$$L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) \geq \sum_{A \in \mathcal{P}_n} \sum_{B \in Q_n} |\nu(A \times B) - \mu_1(A) \cdot \mu_2(B)|$$
$$\underbrace{- \sum_{A \in \mathcal{P}_n} \sum_{B \in Q_n} |\nu_n(A \times B) - \nu(A \times B)| - \sum_{B \in Q_n} |\mu_2(B) - \mu_{n,2}(B)| - \sum_{A \in \mathcal{P}_n} |\mu_1(A) - \mu_{n,1}(A)|.}_{\to 0 \text{ a.s.}}$$

- Given cells shrink in $\mathcal{P}_n$ and $\mathcal{Q}_n$,

$$\sum_{A \in \mathcal{P}_n} \sum_{B \in Q_n} |\nu(A \times B) - \mu_1(A) \cdot \mu_2(B)| \to 2 \sup_C |\nu(C) - \mu_1 \times \mu_2(C)| > 0$$

  - C Borel subset of $\mathbb{R}^d \times \mathbb{R}^{d'}$; see Barron et al. (1992)

L_1: Asymptotic α-level test

- Additional conditions:

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

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

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

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

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

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

  - 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 $\le \alpha$

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

Proof: sketch only

- Problem: samples in each bin not independent
- Define the poisson random variable $N_n \sim \mathrm{Poisson}(n)$.
- Define the poisson variables:

$$n\nu_{N_n}(A \times B) = \#\{i : (X_i, Y_i) \in A \times B, i = 1, \ldots, N_n\},$$

$$n\mu_{N_n,1}(A) = \#\{i : X_i \in A, i = 1, \ldots, N_n\},$$

and

$$n\mu_{N_n,2}(B) = \#\{i : Y_i \in B, i = 1, \ldots, N_n\}$$

- New Poissonized statistic:

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

L_1: Asymptotic α-level test

Proof: sketch only

- Problem: samples in each bin not independent
- Define the poisson random variable $N_n \sim \mathrm{Poisson}(n)$.
- Define the poisson variables:

$$n\nu_{N_n}(A \times B) = \#\{i : (X_i, Y_i) \in A \times B, i = 1, \ldots, N_n\},$$

$$n\mu_{N_n,1}(A) = \#\{i : X_i \in A, i = 1, \ldots, N_n\},$$

and

$$n\mu_{N_n,2}(B) = \#\{i : Y_i \in B, i = 1, \ldots, N_n\}$$

- New Poissonized statistic:

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

Proof: sketch only

- Problem: samples in each bin not independent
- Define the poisson random variable $N_n \sim \mathrm{Poisson}(n)$.
- Define the poisson variables:

$$n\nu_{N_n}(A \times B) = \#\{i : (X_i, Y_i) \in A \times B, i = 1, \dots, N_n\},$$

$$n\mu_{N_n,1}(A) = \#\{i : X_i \in A, i = 1, \dots, N_n\},$$

and

$$n\mu_{N_n,2}(B) = \#\{i : Y_i \in B, i = 1, \dots, N_n\}$$

- New Poissonized statistic:

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

L_1: Asymptotic α-level test

Proof sketch (continued, part 1)

- If poissonized statistic converges...

$$\left(\sqrt{n}\left[\tilde{L}_n - \mathbf{E}(\tilde{L}_n)\right], \frac{N_n - n}{\sqrt{n}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\left[\begin{array}{c} 0 \\ 0 \end{array}\right], \left[\begin{array}{cc} \sigma^2 & 0 \\ 0 & 1 \end{array}\right]\right)$$

- ...then test statistic converges:

$$\frac{\sqrt{n}}{\sigma}\left(L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) - \mathbf{E}(\tilde{L}_n)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

- (proof by comparing characteristic functions) Beirlant et al. (1994)

L_1: Asymptotic α-level test

Proof sketch (continued, part 1)

- If poissonized statistic converges...

$$\left(\sqrt{n} \left[\tilde{L}_n - \mathbf{E}(\tilde{L}_n) \right], \frac{N_n - n}{\sqrt{n}} \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

- ...then test statistic converges:

$$\frac{\sqrt{n}}{\sigma} \left(L_n(\nu_n, \mu_{n,1} \times \mu_{n,2}) - \mathbf{E}(\tilde{L}_n) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

- (proof by comparing characteristic functions) Beirlant et al. (1994)

L_1: Asymptotic α-level test

Proof sketch (continued, part 2)

- Convergence for Poisson e.g. Jacod and Protter(2004, p. 170)

$$\frac{1}{\sqrt{n}}(Z_n - n) \xrightarrow{\mathcal{D}} Z \qquad \text{as } n \to \infty$$

- where $Z \sim \mathcal{N}(0,1)$. Likewise

$$\nu_{N_n}(A \times B) - \mu_{N_n,1}(A)\mu_{N_n,2}(B) \overset{\mathcal{D}}{\approx} Z\sqrt{\frac{\mu_1(A)\mu_2(B)}{n}}$$

Log likelihood: test statistic

- Test statistic:

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

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

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

Log likelihood: test statistic

- Test statistic:

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

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

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

Log likelihood: test statistic

- Test statistic:

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

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

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

Log likelihood: test statistic

- Test statistic:

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

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

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

Log likelihood: tests

Proof sketch (strong consistent)

- 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)

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

- Under $\mathcal{H}_1$, Pinsker:

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

Kernel Statistic

- Assume $k(x, \cdot) = k(\|x - \cdot\|)$
- Kernel test statistic:

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

  - $K_{i,j} = k(x_i, x_j)$, centering $H = I - \frac{1}{n} 1_n 1_n^\top$
  - $K'_{i,j} = k(y_i, y_j)$
- Decreasing bandwidth: L^2 distance between densities
- Fixed bandwidth: Smoothed distance between characteristic functions Feuerverger (1993) OR difference between distribution embeddings to RKHS

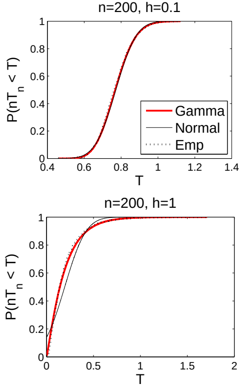
  Berlinet and Thomas-Agnan (2003), Gretton et al. (2008)

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) Feuerverger (1993), Gretton (2008)

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

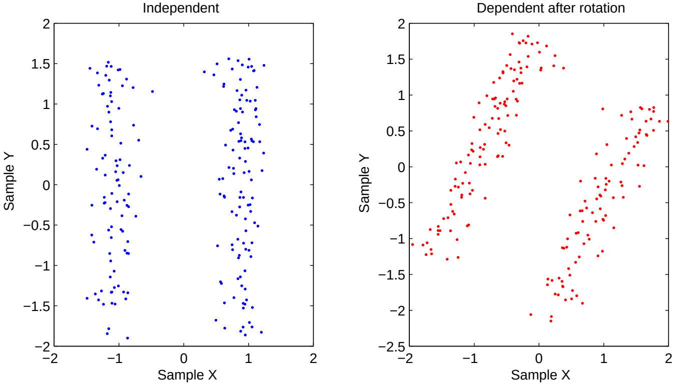
$$\alpha = \frac{(\mathbf{E}(T_n))^2}{\mathrm{var}(T_n)}, \quad \beta = \frac{n\mathrm{var}(T_n)}{\mathbf{E}(T_n)}.$$



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

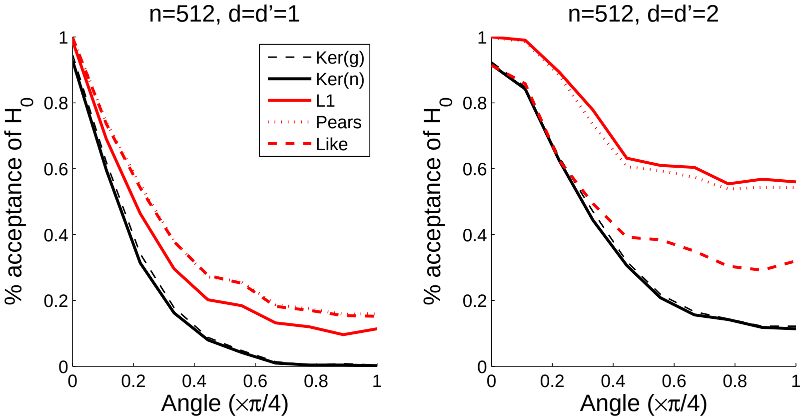
Experimental results

- 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



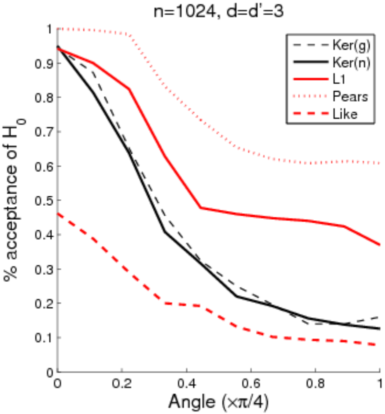
Experimental results

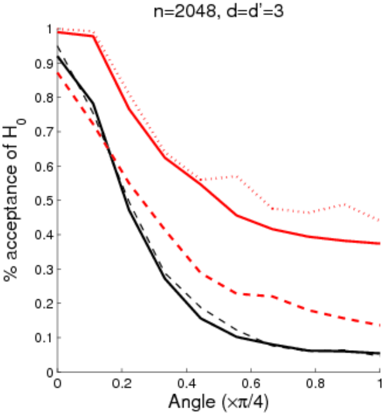
- Comparison of asymptotically α-level tests
- Given distribution Pr, test $\mathcal{H}_0 \ : \ \Pr = \Pr_x \Pr_y$
- Continuous valued, multivariate: $\mathcal{X} := \mathbb{R}^d$ and $\mathcal{Y} := \mathbb{R}^{d'}$



Experimental results

- Higher dimensions: $d = d' = 3$





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