# Robust Kernel Hypothesis Testing under Data Corruption

**Antonin Schrab*** AI Centre, Gatsby Unit, Inria London, University College London, UK
**Ilmun Kim*** Department of Statistics & Data Science, Yonsei University, South Korea
*equal contributions

## Robust Testing Framework

**AISTATS ORAL**

**Space of distributions:** $\mathcal{P}$ partitioned into disjoint $\mathcal{P}_0$ and $\mathcal{P}_1$

**Testing:** given data related to some fixed $P \in \mathcal{P}$, determine whether

$$\mathcal{H}_0 \colon P \in \mathcal{P}_0 \qquad \text{or} \qquad \mathcal{H}_1 \colon P \in \mathcal{P}_1$$

**Data:** **- Standard framework:** i.i.d. samples $X_1, \ldots, X_N$ from $P \in \mathcal{P}$

**- Robust framework:** samples $X_1, \ldots, X_N$ where

- $N - r$ samples are i.i.d. from $P \in \mathcal{P}$
- $r$ samples have potentially been corrupted

Robustness parameter $r$ is specific by the user depending on the application.

**Intuition:** the "robust null hypothesis" is enlarged to include all distributions

$$\left\{ P(1-\epsilon) + Q\epsilon \ : P \in \mathcal{P}_0, \ \ Q \in \mathcal{P}_1, \ \ \epsilon \in [0, r/N] \right\}$$

Our robust setting is actually more general: it allows for adversarial corruption.

**Algorithm 1** Robust DC procedure

**Inputs:** Data $\mathcal{X}_n$, robustness $r$, level $\alpha$, statistic $T$, permutation number $B$.

Generate i.i.d. permutations $\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_B$ of $[n]$.
Set $\boldsymbol{\pi}_0 = \mathrm{Id}$ and compute global sensitivity $\Delta_T$.
Compute $T_i = T(\mathcal{X}_n^{\boldsymbol{\pi}_i})$, $i \in [B]_0$.
Compute $(1-\alpha)$-quantile $q$ of $T_0, \ldots, T_B$.

**Output**: Reject $\mathcal{H}_0$ if $T_0 > q + 2r\Delta_T$.

**Global sensitivity:** $\Delta_T$ maximum difference in the statistic $T$ output when evaluated on permuted datasets whose entries differ by at most one

**Level:** the DC test is well-calibrated non-asymptotically: for any $P_0 \in \mathcal{P}_0$

$$\mathbb{P}_{P_0}\big(\text{DC rejects } \mathcal{H}_0 \mid r \text{ corrupted data}\big) \ \leq \ \alpha$$

**Consistency:** the DC test is consistent in the sense that, for any fixed $P_1 \in \mathcal{P}_1$

$$\lim_{n\to\infty} \mathbb{P}_{P_1}\big(\text{DC rejects } \mathcal{H}_0 \mid r \text{ corrupted data}\big) = 1 \qquad \text{if}$$

$$\lim_{n\to\infty} \mathbb{P}_{P_1}\big(T(\mathbb{X}_n) > T(\mathbb{X}_n^{\pi}) + 4r\Delta_T\big) = 1$$

## Two-sample Testing

**Two-sample problem:** Given independent

- i.i.d. samples $X_1, \ldots, X_m$ from a distribution $P$,
- i.i.d. samples $Y_1, \ldots, Y_n$ from a distribution $Q$,

test whether $\mathcal{H}_0 \colon P = Q$ or $\mathcal{H}_1 \colon P \neq Q$. Let $N = \min(m, n)$.

**Robust testing:** Up to $r$ samples from either $P$ or $Q$ can be corrupted

**Maximum Mean Discrepancy:**

$$\mathrm{MMD} = \sqrt{\mathbb{E}_{P,P}[k(X,X')] - 2\mathbb{E}_{P,Q}[k(X,Y)] + \mathbb{E}_{Q,Q}[k(Y,Y')]}$$

**Statistic:**

$$\widehat{\mathrm{MMD}} = \sqrt{\frac{1}{m^2}\sum_{1\leq i,i'\leq m} k(X_i, X_{i'}) - \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} k(X_i, Y_j) + \frac{1}{n^2}\sum_{1\leq j,j'\leq n} k(Y_j, Y_{j'})}$$

**Global sensitivity:** $\Delta_{\widehat{\mathrm{MMD}}} = \sqrt{2K}/N$

**Permutations:** $Z_i = X_i$, $Z_{m+i} = Y_i$, $(\mathbb{X}_m^{\pi}, \mathbb{Y}_n^{\pi}) = \big(Z_{\pi(1)}, \ldots, Z_{\pi(m+n)}\big)$

**dcMMD:** Apply DC procedure with $\widehat{\mathrm{MMD}}$ and $\Delta_{\widehat{\mathrm{MMD}}}$

## dcMMD Guarantees

**Level:** for any distribution $P$ and any sample size

$$\mathbb{P}_{P,P}\big(\text{dcMMD rejects } \mathcal{H}_0 \mid r \text{ corrupted data}\big) \ \leq \ \alpha$$

**Pointwise Power / Consistency:** for any fixed $P \neq Q$ and $r/N \to 0$

$$\lim_{m,n\to\infty} \mathbb{P}_{P,Q}\big(\text{dcMMD rejects } \mathcal{H}_0 \mid r \text{ corrupted data}\big) \ = \ 1$$

**Uniform Power:** for any distributions $P$ and $Q$ separated as

$$\mathrm{MMD}(P,Q) \ \gtrsim \ \max\left\{\sqrt{\frac{\max\{\log(1/\alpha), \log(1/\beta)\}}{N}}, \frac{r}{N}\right\}$$

dcMMD achieves high test power

$$\mathbb{P}_{P,Q}\big(\text{dcMMD rejects } \mathcal{H}_0 \mid r \text{ corrupted data}\big) \ \geq \ 1-\beta$$

This rate is minimax optimal with respect to $N, r, \alpha, \beta$.

## dcMMD Experiments

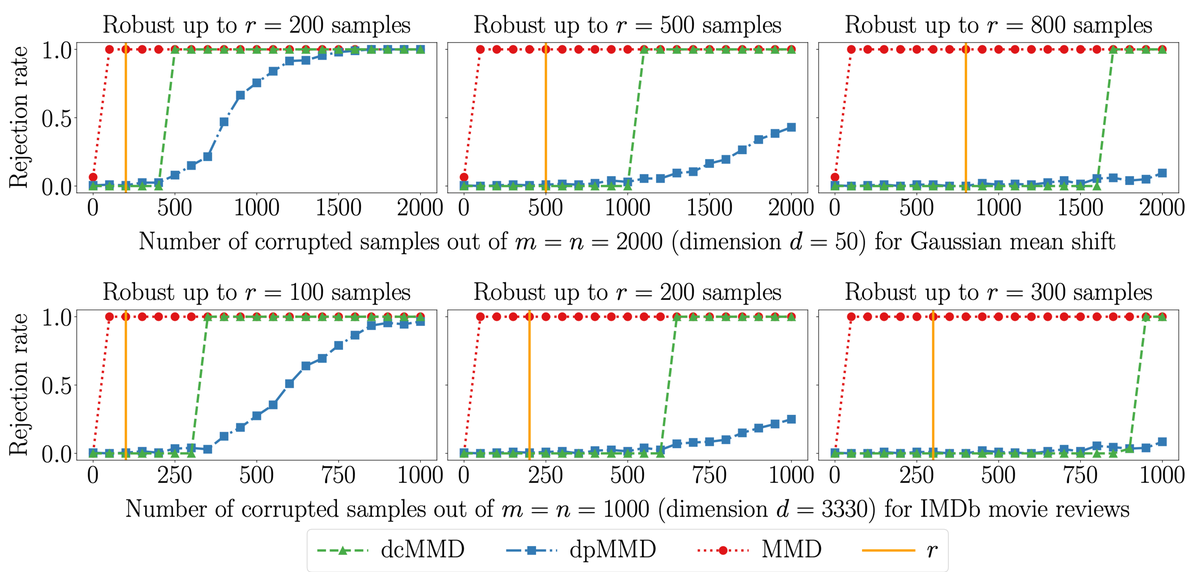


**Figure 1:** Two-sample experiments robust up to $r$ corrupted samples. To have valid level, a robust test needs to control the rejection rate by $\alpha = 0.05$ when fewer than $r$ samples are corrupted. To be powerful, the robust test needs to have a high rejection rate when more than $r$ samples are corrupted. *(Top row: Gaussian mean shift)* Both samples are originally i.i.d. drawn from $\mathsf{Gaussian}(0, 1/10, 50)$, entries of one sample are corrupted being replaced by samples from $\mathsf{Gaussian}(1000, 1/10, 50)$. *(Bottom row: IMDb movie reviews)* Both samples originally consist of movie reviews (using a bag of 3330 words representation). Corrupted entries for one sample are replaced by samples from $\mathsf{Geometric}(0.05, 3330)$.

## Independence testing

**Independence problem:** Given

- paired samples $(X_1, Y_1), \ldots, (X_N, Y_N)$ drawn i.i.d. from a joint $P_{XY}$,

test whether the two components of the pairs are independent

$$\mathcal{H}_0 \colon P_{XY} = P_X \otimes P_Y \qquad \text{vs} \qquad \mathcal{H}_1 \colon P_{XY} \neq P_X \otimes P_Y$$

**Robust testing:** Up to $r$ paired samples can be corrupted

**Hilbert-Schmidt Independence Criterion:**

$$\mathrm{HSIC} = \sqrt{\mathbb{E}_{P_{XY},P_{XY}}\Big[k(X,X')\ell(Y,Y')\Big] - 2\mathbb{E}_{P_{XY}}\Big[\mathbb{E}_{P_X}[k(X,X')]\mathbb{E}_{P_Y}[\ell(Y,Y')]\Big] + \mathbb{E}_{P_X,P_X}\Big[k(X,X')\Big]\mathbb{E}_{P_Y,P_Y}\Big[\ell(Y,Y')\Big]}$$

**Statistic:**

$$\widehat{\mathrm{HSIC}} = \sqrt{\frac{1}{N^2}\sum_{1\leq i,j\leq N} K_{ij}L_{ij} - \frac{2}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j=1}^{N} K_{ij}\right)\left(\frac{1}{N}\sum_{r=1}^{N} L_{ir}\right) + \left(\frac{1}{N^2}\sum_{1\leq i,j\leq N} K_{ij}\right)\left(\frac{1}{N^2}\sum_{1\leq r,s\leq N} L_{rs}\right)}$$

**Global sensitivity:** $\Delta_{\widehat{\mathrm{HSIC}}} \leq 4\sqrt{KL}(N-1)/N^2$ asymptotically tight

**Permutations:** $\big(X_1, Y_{\pi(1)}\big), \ldots, \big(X_N, Y_{\pi(N)}\big)$

**dcHSIC:** Apply DC procedure with $\widehat{\mathrm{HSIC}}$ and $\Delta_{\widehat{\mathrm{HSIC}}}$

## dcHSIC Guarantees

**Level:** for any marginals $P_X, P_Y$ and any sample size

$$\mathbb{P}_{P_X \times P_Y}\big(\text{dcHSIC rejects } \mathcal{H}_0 \mid r \text{ corrupted data}\big) \ \leq \ \alpha$$

**Pointwise Power / Consistency:** for any fixed joint $P_{XY}$ and $r/N \to 0$

$$\lim_{N\to\infty} \mathbb{P}_{P_{XY}}\big(\text{dcHSIC rejects } \mathcal{H}_0 \mid r \text{ corrupted data}\big) \ = \ 1$$

**Uniform Power:** for any joint $P_{XY}$ separated from the null as

$$\mathrm{HSIC}(P_{XY}) \ \gtrsim \ \max\left\{\sqrt{\frac{\max\{\log(1/\alpha), \log(1/\beta)\}}{N}}, \frac{r}{N}\right\}$$

dcHSIC achieves high test power

$$\mathbb{P}_{P_{XY}}\big(\text{dcHSIC rejects } \mathcal{H}_0 \mid r \text{ corrupted data}\big) \ \geq \ 1-\beta$$

This rate is minimax optimal with respect to $N, r, \alpha, \beta$.

## dcHSIC Experiments

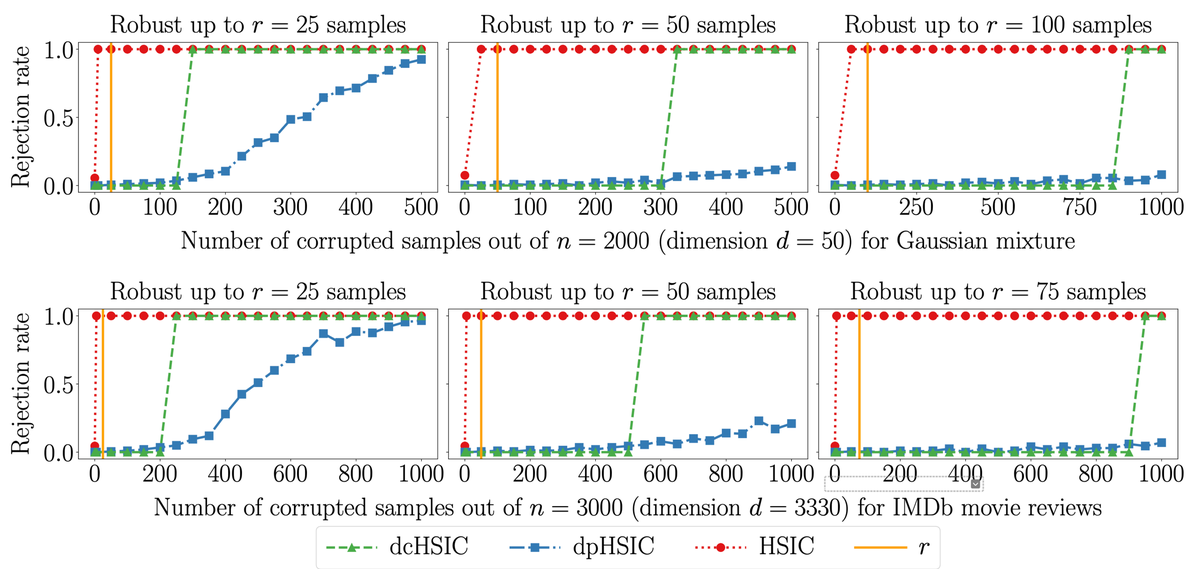


**Figure 2:** Independence experiments robust up to $r$ corrupted samples. To have valid level, a robust test needs to control the rejection rate by $\alpha = 0.05$ when fewer than $r$ samples are corrupted. To be powerful, the robust test needs to have a high rejection rate when more than $r$ samples are corrupted. *(Top row: Gaussian mixture)* Paired samples $(X, Y)$ are originally i.i.d. drawn from two $\mathsf{Gaussian}(0, 1/10, 50)$. Corrupted samples are replaced by $(X, X+\epsilon)$ where $\epsilon \sim \mathsf{Gaussian}(0, 1/10, 50)$ and where $X \sim \mathsf{Gaussian}(s1000, 1/10, 50)$ with $s = 1$ for half of the corrupted samples and $s = -1$ for the other half. *(Bottom row: IMDb movie reviews)* Paired samples $(X, Y)$ originally consist of two independent reviews (represented using a bag of 3330 words). Corrupted samples are replaced by $(X+s, X+s+\epsilon)$ where $X \sim \mathsf{Geometric}(0.05, 3330)$, $\epsilon \sim \mathsf{Gaussian}(0, 1/10, 3330)$ and with $s = 0$ for half of the corrupted samples and $s = 5$ for the other half.

## Summary

- Introduce **DC procedure**: a general approach for constructing robust permutation tests under data corruption.
- Guarantee non-asymptotic **validity** under $r$ data corruption.
- Prove DC **consistency** in power under minimal conditions.
- Construct **dcMMD** and **dcHSIC** for the two-sample and independence robust testing frameworks (valid and consistent tests).
- dcMMD/dcHSIC are **minimax optimal**: they are non-asymptotically powerful against alternatives uniformly separated from the null in the kernel MMD and HSIC metrics at some optimal rate (tight with **matching lower bound**).
- Provide public **implementations** and illustrate the **practicality** of our DC robust tests with strong empirical evidence in terms of **power** and **level** results.

**Paper:**

**Code:**