Permutation Test in Dependent Data

Yu Xinyang

October 9, 2025

Outline

Permutation Test

Block Permutation

Studentization

Basic Idea

- Permutation tests are nonparametric hypothesis tests.
- Key assumption: under H_0, data labels are **exchangeable**.
- Construct the null distribution of a test statistic by permuting labels.
- Distribution-free: does not rely on normality or equal variance assumptions, only on exchangeability.

Framework

Two-sample problem:

$$X_1, \ldots, X_m \sim F, \quad Y_1, \ldots, Y_n \sim G,$$

Test

$$H_0 : F = G \quad \text{vs.} \quad H_1 : F \neq G.$$

1. Compute observed statistic T_{obs}.
2. Pool all samples $\{Z_1, \ldots, Z_{m+n}\}$.
3. Permute group labels, compute $T^{(b)}$.
4. Construct permutation distribution:

$$\hat{F}_T(t) = \frac{1}{B} \sum_{b=1}^{B} I(T^{(b)} \leq t).$$

5. Compute p-value:

$$p = \frac{1}{B} \sum_{b=1}^{B} I(T^{(b)} \geq T_{\text{obs}}).$$

Property

Finite-sample Exactness

If samples are exchangeable under H_0,

$$\mathbb{P}_{H_0}(p \leq \alpha) \leq \alpha,$$

i.e., permutation test is exact in finite samples.

Consistency

If the test statistic T is sensitive to differences, then as $m, n \to \infty$,

$$\mathbb{P}_{H_1}(p \to 0) = 1.$$

Extensions: Permutation-based FDR Control

Step 1: Observed rejections at threshold t:

$$R(t) = \#\{j : T_{j,\text{obs}} \geq t\}.$$

Step 2: Permutation null distribution.

- Permute Y (or residuals from reduced model).
- Recompute $T_j^{(b)}$ for $b = 1, \ldots, B$.

Step 3: Estimate false rejections:

$$\hat{V}(t) = \frac{1}{B} \sum_{b=1}^{B} \#\{j : T_j^{(b)} \geq t\}.$$

Step 4: Estimated FDR:

$$\widehat{\text{FDR}}(t) = \frac{\hat{V}(t)}{R(t) \vee 1}.$$

Extensions: Permutation-based FDR Control

Choose largest threshold t^* such that

$$\widehat{\mathrm{FDR}}(t^*) \leq \alpha.$$

Reject all H_j with $T_{j,\mathrm{obs}} \geq t^*$.

Advantages:

- No reliance on uniform null p-value assumption.
- Accounts for dependence among tests.
- Data-driven, robust in high-dimensional regression.

Extensions: Machine Learning Applications

- Permutation tests used for **feature importance**.
- Shuffle feature X_j, break dependency with Y.
- Compute prediction loss increase:

$$I_j = \mathbb{E}[\ell(\hat{f}(X_{-j}, \pi(X_j)), Y)] - \mathbb{E}[\ell(\hat{f}(X), Y)].$$

- If I_j significantly > 0, feature X_j is important.

Block Permutation(Kirch, 2007)

Goal: Testing for change points in dependent data.

AMOC Model (At Most One Change)

$$X_t = \mu + d1_{\{t>m\}} + e_t, \quad t = 1, \ldots, n$$

- μ: baseline mean
- $1_{\{t>m\}} = 0$ for $t \leq m$, $1_{\{t>m\}} = 1$ for $t > m$
- $e_i := \sum_{j>0} w_j \epsilon_{i-j}$, innovations ϵ_i are i.i.d. such that for some $\nu > 2$,

$$\mathrm{E}(\epsilon_i) = 0, 0 < \mathrm{E}(\epsilon_i^2) = \sigma^2 < \infty, \mathrm{E}|\epsilon_i|^\nu < \infty.$$

- $H_0 : m = n$ against $H_1 : m < n$

Challenge: Observations are not exchangeable due to dependence and change point.

Block Permutation(Kirch, 2007)

Idea:

- Divide data into L blocks of length K (KL=n),

$$B_l = (X_{(l-1)K+1}, \ldots, X_{lK}), \quad j = 1, \ldots, L.$$

- Apply a random permutation $\pi \in S_K$ to the blocks:

$$X^{(\pi)} = \left(B_{\pi(1)}, \ldots, B_{\pi(K)}\right).$$

- Preserves within-block dependence while breaking global order.

Permutation Statistic:

$$T_n^{(L,K)} = f(B_{\pi(1)}, \ldots, B_{\pi(K)}).$$

Block Permutation(Kirch, 2007)

Let $X_1, \ldots, X_n$ be a (weakly dependent) series with sample mean $\bar{X}_n$

$$T_n^{(1)} = \max_{1<m<n} \sqrt{\frac{n}{m(n-m)}} \left|S_m\right|, \quad S_m = \sum_{t=1}^{m} (X_t - \bar{X}_n), \; m = 1, \ldots, n-1.$$

Define the permuted partial-sum process (for $l = 1, \ldots, L, \; k = 1, \ldots, K$ exclude $(l,k) = (L,K)$):

$$S_\pi^{L,K}(l,k) = \sum_{i=1}^{l-1} \sum_{j=1}^{K} (X_{K(\pi(i)-1)+j} - \bar{X}_n) + \sum_{j=1}^{k} (X_{K(\pi(l)-1)+j} - \bar{X}_n),$$

and the block-permuted $T_{L,K}^{(1)}$ is

$$T_{L,K}^{(1)}(\pi, X) = \max_{\substack{l,k \\ (l,k) \neq (L,K)}} \sqrt{\frac{LK}{(K(l-1)+k)\left(LK - K(l-1) - k\right)}} \left|S_\pi^{L,K}(l,k)\right|.$$

Block Permutation(Kirch, 2007)

Long-range variance:

$$\tau^2 = \sigma^2 \left(\sum_{j \geq 0} w_j^2 \right).$$

Block-based estimator:

$$\hat{\tau}^2 = \frac{1}{n} \sum_{j=1}^{K} \left(\sum_{t \in B_j} (X_t - \bar{X}_n) \right)^2.$$

Under H_0, not rely to permutation π,

$$\hat{\tau}^2 = \tau^2 + O_P\left(\frac{1}{\sqrt{L}} + \frac{1}{\sqrt{K}} + \frac{\log\log n}{L} \right) + o(n^{-1}).$$

Consistency requires $K, L \to \infty$.

Block Permutation(Kirch, 2007)

Define $\alpha(x) = \sqrt{2\log x}$ and $\beta(x) = 2\log x + \frac{1}{2}\log\log x - \frac{1}{2}\log\pi$.
Under suitable mixing and moment conditions, under H_0,

$$P\left(\frac{\alpha(\log n)}{\tau}\,T_n^{(1)} - \beta(\log n)\right) \implies \exp(-2e^{-x}).$$

Assumptions.

- Weak dependence: strong mixing with suitable decay; finite moment $\nu > 4$.
- Block growth: $K = K(n) \to \infty$, $L = L(n) \to \infty$, $n = KL$, and $K = O\big((\log n)^{\gamma}\big)$ for some $\gamma > 0$.

Conditioning on the data $X_{1:n}$, under H_0,

$$P\left(\frac{\alpha(\log n)}{\hat{\tau}_{L,K}}\,T_{L,K}^{(1)}(\pi, X) - \beta(\log n) \le x \;\middle|\; X_{1:n}\right) \implies \exp(-2e^{-x}) \quad \text{a.s..}$$

Roadmap of proof(Kirch, 2007)

- Define block sums and within-block residuals:

$$U_\ell = \sum_{j=1}^{K} (X_{K(\ell-1)+j} - \bar{X}_n), \quad R_\ell(k) = \sum_{j=1}^{k} (X_{K(\ell-1)+j} - \bar{X}_n)$$

- For permutation π:

$$S_\pi^{L,K}(l,k) = \underbrace{\sum_{i=1}^{l-1} U_{\pi(i)}}_{\text{block-level main term}} + \underbrace{R_{\pi(l)}(k)}_{\text{within-block residual}}$$

- Normalized CUSUM:

$$T_{l,k}^{(1)}(\pi, X) = \max_{l,k} \sqrt{\frac{LK}{(K(l-1)+k)\left(LK - K(l-1) - k\right)}} \left| S_\pi^{L,K}(l,k) \right|.$$

Roadmap of proof(Kirch, 2007)

- Under some condition, $\max_{l,k} |R_l(k)| \leq O_p(\sqrt{K})$ and

$$\sqrt{\frac{LK}{(K(l-1)+k)\left(LK - K(l-1) - k\right)}} R_{\pi(l)} = O\left(\frac{1}{\sqrt{L}}\right).$$

- Conditioned on the observed block sums $\{U_\ell\}$, permutation is a without-replacement random shuffle of a finite population.
- Define the permutation bridge process:

$$W_L(t) = \frac{\sum_{i=1}^{\lfloor Lt \rfloor} U_{\pi(i)} - t \sum_{i=1}^{L} U_i}{\sqrt{\sum_{i=1}^{L} U_i^2}}, \quad t \in [0,1]$$

- Under finite-population Lindeberg condition: $\max_i U_i^2 / \sum_{i=1}^{L} U_i^2 \to 0$, $W_L \Rightarrow B$, a Brownian bridge.

Roadmap of proof(Kirch, 2007)

With $\hat{\tau}_{L,K} \xrightarrow{P} \tau$, the normalized block permutation statistic:

$$\begin{aligned}
\frac{T^{(1)}_{L,K}(\pi, X)}{\hat{\tau}_{L,K}} &= \frac{\sqrt{\sum_i U_i}}{\hat{\tau}_{L,K}\sqrt{L}} \sup_{t\in(0,1)} \frac{|W_L(t)|}{\sqrt{t(1-t)}} + o_p(1) \\
&\Rightarrow \sup_{t\in(0,1)} \frac{|B(t)|}{\sqrt{t(1-t)}} \\
&\Rightarrow \text{Gumbel}(x).
\end{aligned}$$

Roadmap of proof(Kirch, 2007)

Summary

- Original time correlation: τ, captured by $\hat{\tau}_{L,K}$.
- Conditional on the observed block sums $\{U_\ell\}$, the permutation removes the dependence structure:
  - Within-block: captured by U_ℓ.
  - Across-block: controlled after permutation and asymptotic scaling.
- Block-level maximum residuals vanish as $L \to \infty$, $K \to \infty$ but $K/L \to 0$.
- Conclusion: block permutation statistic (after normalization) has the same limiting distribution as the original CUSUM statistic under H_0.

Problem and Motivation(Romano and Tirlea, 2022)

- $H_0 : \rho(1) = \rho(2) = \cdots = \rho(r) = 0$ for some fixed r.
- Permutation tests may not control Type 1 error asymptotically; also Type 3 (directional) errors.
- Test statistic: Sample autocorrelation

$$\hat{\rho}_n(k) = \frac{\frac{1}{n-k}\sum_{i=1}^{n-k}(X_i - \bar{X}_n)(X_{i+k} - \bar{X}_n)}{\hat{\sigma}_n^2},$$

  where $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$.
- Permutation distribution:

$$\hat{R}_{T_n}^n(t) := \frac{1}{n!}\sum_{\pi_n \in S_n} I\{T_n(X_{\pi_n(1)}, \ldots, X_{\pi_n(n)}) \leq t\},$$

  with S_n be the permutation group of order n.

Preliminaries(Romano and Tirlea, 2022)

- Under α-mixing or m-dependence with moment conditions,

$$\sqrt{n}(\hat{\rho}_n(1) - \rho_1) \xrightarrow{d} \mathcal{N}(0, \gamma_1^2),$$

where

$$\gamma_1^2 = \frac{1}{\sigma^4}(\tau_1^2 - 2\rho_1\nu_1 + \rho_1^2\kappa^2),$$

with

$$\kappa^2 = \mathrm{Var}(X_1^2) + 2\sum_{k\geq 2}\mathrm{Cov}(X_1^2, X_k^2),$$

$$\tau_1^2 = \mathrm{Var}(X_1X_2) + 2\sum_{k\geq 2}\mathrm{Cov}(X_1X_2, X_kX_{k+1}),$$

$$\nu_1 = \mathrm{Cov}(X_1X_2, X_1^2) + \sum_{k\geq 2}\mathrm{Cov}(X_1^2, X_kX_{k+1}) + \sum_{k\geq 2}\mathrm{Cov}(X_1X_2, X_k^2).$$

- Variance components: $\kappa^2, \tau_1^2, \nu_1$ capture long-run covariances.

Key Challenges(Romano and Tirlea, 2022)

- Zero autocorrelation $\neq$ independence: Permutation invariance fails under dependence.
- Asymptotic mismatch:
  - Sample distribution under null is $\sqrt{n}(\hat{\rho}_n(1) - \rho_1) \xrightarrow{d} N(0, \gamma_1^2)$.
  - Let $\mathrm{Var}(X_i) = 1$, under α-mixing and moments, permutation distribution of $\sqrt{n}\hat{\rho}_n(1)$ converges in probability to Φ (standard normal CDF):
  $$\sup_{t \in \mathbb{R}} |\hat{R}_n(t) - \Phi(t)| \xrightarrow{p} 0.$$

  But sample variance is $\gamma_1^2 \neq 1$ in general.
- Studentize the statistic using consistent variance estimator $\hat{\gamma}_n^2$ ($\to \gamma_1$) to match distributions: $\sqrt{n}(\hat{\rho}_n(1) - \rho_1)/\hat{\gamma}_n^2$ ($\to N(0,1)$).

Studentization(Romano and Tirlea, 2022)

- Let $Y_i = (X_i - \bar{X}_n)(X_{i+1} - \bar{X}_n)$, $Z_i = (X_i - \bar{X}_n)^2$.
- Estimator $\hat{\gamma}_n^2 = \frac{1}{\hat{\sigma}_n^4}[\hat{T}_n^2 - 2\hat{\rho}_n\hat{\nu}_n + \hat{\rho}_n^2\hat{K}_n^2]$, with bandwidth $b_n = o(\sqrt{n})$.
- $$\begin{aligned}
\hat{K}_n^2 =& \frac{1}{n}\sum_{i=1}^{n}(Z_i - \bar{Z}_n)^2 + \frac{2}{n}\sum_{j=1}^{b_n}\sum_{i=1}^{n-j}(Z_i - \bar{Z}_n)(Z_{i+j} - \bar{Z}_n), \\
\hat{T}_n^2 =& \frac{1}{n}\sum_{i=1}^{n-1}(Y_i - \bar{Y}_n)^2 + \frac{2}{n}\sum_{j=1}^{b_n}\sum_{i=1}^{n-j-1}(Y_i - \bar{Y}_n)(Y_{i+j} - \bar{Y}_n), \\
\hat{\nu}_n =& \frac{1}{n}\sum_{i=1}^{n-1}(Y_i - \bar{Y}_n)(Z_i - \bar{Z}_n) + \frac{1}{n}\sum_{j=1}^{b_n}\sum_{i=1}^{n-j-1}(Z_i - \bar{Z}_n)(Y_{i+j} - \bar{Y}_n) \\
&+ \frac{1}{n}\sum_{j=1}^{b_n}\sum_{i=1}^{n-j}(Y_i - \bar{Y}_n)(Z_{i+j} - \bar{Z}_n).
\end{aligned}$$
- Uses truncated sums (b_n) to estimate long-run variances under dependence.

Main Results(Romano and Tirlea, 2022)

- Without permutation: we have that, as $n \to \infty$,

$$\sqrt{n}(\hat{\rho}_n - \rho_1)/\hat{\gamma}_n \xrightarrow{d} \mathcal{N}(0, 1).$$

- Under permutation, $\hat{\gamma}_n^2 \xrightarrow{p} \mathrm{Var}(X) = 1$.
- Let $\hat{R}_n$ be the permutation distribution, based on the test statistic $\sqrt{n}\hat{\rho}_n/\hat{\gamma}_n$. Then as $n \to \infty$,

$$\sup_{t \in \mathbb{R}} |\hat{R}_n(t) - \Phi(t)| \xrightarrow{p} 0.$$

Multiple Testing Framework(Romano and Tirlea, 2022)

- For $H_m : \rho(1) = \cdots = \rho(m) = 0$, combine individual permutation tests using multiple testing procedures (e.g., Bonferroni: reject if $\min_i \hat{p}_i \leq \alpha/r$).
- Let $\Sigma = (\sigma_{i,j})_{i,j=0}^{r}$ with

$$\sigma_{i,j} = \begin{cases} \mathrm{Var}(X_1 X_{1+i}) + 2\sum_{l>1} \mathrm{Cov}(X_1 X_{1+i}, X_l X_{l+i}), i = j \\ \mathrm{Cov}(X_1 X_{1+i}, X_1 X_{1+j}) + \sum_{l>1} \left(\mathrm{Cov}(X_1 X_{1+i}, X_l X_{l+j}) + \mathrm{Cov}(X_1 X_{1+j}, X_l X_{l+i})\right), i \neq j. \end{cases}$$

Let $A \in R^{(r+1)*r}$ with $A_{1,i} = -\rho_1/\sigma^4$, $A_{i+1,i} = 1/\sigma^2$ for $i = 1, \cdots, r$ and other elements are 0. Then, as $n \to \infty$,

$$\sqrt{n}\left((\hat{\rho}_1, \cdots, \hat{\rho}_r)^\top - (\rho_1, \cdots, \rho_r)^\top\right) \to \mathcal{N}(0, A^\top \Sigma A).$$

Definition of Monotone Trend(Romano and Tirlea, 2024)

- Data: time series $(X_1, \ldots, X_n)$ from a weakly dependent process.
- Distribution at time t: $F_t(x) = P(X_t \leq x)$.
- **Null hypothesis (strictly stationary):**

$$H_0 : \quad F_1 = F_2 = \cdots = F_n.$$

- **Alternative hypothesis (monotone trend):**

$$H_1 : \quad F_1(x) \geq F_2(x) \geq \cdots \geq F_n(x), \quad \forall x$$

  or the reverse ordering.

- Interpretation: distributions evolve monotonically over time in stochastic order.

Main Results(Romano and Tirlea, 2024)

- Mann-Kendall statistic:

$$U_n = U_n(X_1, \cdots, X_n) = \binom{n}{2}^{-1} \sum_{1 \le i \le j \le n} I(X_j > X_i) - I(X_i > X_j).$$

- The $\hat{R}(t)$ based on $\sqrt{n}U_n$ stisfied, as $n \to \infty$,

$$\sup_{t \in \mathbb{R}} |\hat{R}_n(t) - \Phi(3t/2)| \xrightarrow{p} 0.$$

- Suppose that the β-mixing coefficients of X_n satisfy $\sum_t \beta_X t < \infty$. Let $\sigma^2 = 4/9 + 8/3 \sum_{k \ge 1} \mathrm{Cov}(V_1, V_{1+k})$ with $V_i := 1 - 2F(X_i)$, as $n \to \infty$,

$$\frac{\sqrt{n}U_n}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1).$$

Studentization(Romano and Tirlea, 2024)

- Let $b_n = o(\sqrt{n})$, and, as $n \to \infty$, $b_n \to \infty$. Difine

$$\hat{\sigma}_n^2 := \frac{4}{9} + \frac{8}{3} \sum_{k=1}^{b_n} \sum_{j=1}^{n-k} (1 - 2\hat{F}_n(X_j))(1 - 2\hat{F}_n(X_{j+k})).$$

- The $\hat{R}(t)$ based on $\sqrt{n} U_n / \hat{\sigma}^2$ stisfied, as $n \to \infty$,

$$\sup_{t \in \mathbb{R}} |\hat{R}_n(t) - \Phi(t)| \xrightarrow{p} 0.$$

- Suppose that the β-mixing coefficients of X_n satisfy $\sum_t \beta_X t < \infty$. As $n \to \infty$,

$$\frac{\sqrt{n} U_n}{\hat{\sigma}_n^2} \xrightarrow{d} \mathcal{N}(0, 1).$$

Key point

- Permutation may lead to $F(T_n) \neq F(T_n^\pi)$.
- Want to choose some static T_n such that $T_n/\hat{\sigma}_n \approx T_n^\pi/\hat{\sigma}_n^\pi$.
- Linear rank statistics:

$$T_n = \sum_{i=1}^{n} w_{i,n}\phi(\hat{F}_n(X_i)),$$

typically $\phi(x) = 1 - 2x$.

Reference

Kirch, C. (2007). Block permutation principles for the change analysis of dependent data. *Journal of Statistical Planning and Inference*, 137(7):2453–2474.

Romano, J. P. and Tirlea, M. A. (2022). Permutation testing for dependence in time series. *Journal of time series analysis*, 43(5):781–807.

Romano, J. P. and Tirlea, M. A. (2024). Permutation testing for monotone "trend". *Journal of Time Series Analysis*.