Bugni and Horowitz (2021) Permutation Tests for the Equality of Distributions of Functional Data

Master Thesis presented to the
Department of Economics at the
Rheinische Friedrich-Wilhelms-Universität Bonn

In Partial Fulfillment of the Requirements for the Degree of
Master of Science (M.Sc.)

Supervisor: Prof. Dr. Dominik Liebl

Submitted in June 2022 by:
Jakob R. Juergens
Matriculation Number: 2996491

Contents

1 Introduction

- Introduce general idea and possible hypothesis to test
- Maybe focus on two sample setting

2 Functional Data Analysis

- Ramsay and Silverman 2005
- Kokoszka and Reimherr 2021
- Hsing and Eubank 2015

2.1 Hilbert Space of Square Integrable Functions

2.2 Random Functions

2.3 Probability Measures on $\mathbb{L}^2$

3 Cramér-von Mises Tests

- Darling 1957
- Anderson and Darling 1952
- Büning and Trenkler 2013

3.1 Empirical Distribution Functions

Gibbons and Chakraborti 2021

Definition 1. Order Statistic
Let $\{x_i \mid i = 1, \dots, n\}$ be a random sample from a population with continuous cumulative distribution function F_X. Then there almost surely exists a unique ordered arrangement within the sample.

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}$$

$X_{(r)} \quad r \in \{1, \dots, n\}$ is called the rth-order statistic.

Definition 2. Empirical Distribution Function

$$F_n(x) = \begin{cases} 0 & \text{if} \quad x < x_{(1)} \\ \frac{r}{n} & \text{if} \quad x_{(r)} \leq x < x_{(r+1)} \\ 1 & \text{if} \quad x \geq x_{(n)} \end{cases} \tag{1}$$

3.2 Cramér-von Mises Statistics

Büning and Trenkler 2013

$$C_{m,n} = \left(\frac{nm}{n+m}\right) \int_{-\infty}^{\infty} (F_m(x) - G_n(x))^2 \,\mathrm{d}\left(\frac{mF_m(x) + nG_n(x)}{m+n}\right) \tag{2}$$

3.3 Asymptotic Distributions

4 Permutation Tests

- Lehmann and Romano 2005
- Vaart and Wellner 1996

4.1 Functional Principle of Permutation Tests

4.2 Size and Power

4.3 Permutation Test for Equality of scalar-valued Distributions

5 Multiple Testing

- Dunn 1961

5.1 Bonferroni Correction

6 Test by Bugni and Horowitz (2021) - Two Samples

- Bugni and Horowitz 2021
- Bugni, Hall, et al. 2009

6.1 Nullhypothesis

6.2 Assumptions

Assumption 1. Contains two assumptions

1. $X(t)$ and $Y(t)$ are separable, μ-measurable stochastic processes.
2. $\{X_i(t) \mid i = 1, \dots, n\}$ is an independent random sample of the process $X(t)$. $\{Y_i(t) \mid i = 1, \dots, m\}$ is an independent random sample of $Y(t)$ and is independent of $\{X_i(t) \mid i = 1, \dots, n\}$.

Assumption 2. $\mathbb{E}X(t)$ and $\mathbb{E}Y(t)$ exist and are finite for all $t \in [0, T]$.

Assumption 3. $X_i(t)$ and $Y_i(t)$ are observed for all $t \in \mathcal{I}$.

6.3 Cramér-von Mises type Test

Distribution Functions

$$\begin{aligned} F_X(z) &= \mathbb{P}\left[X(t) \le z(t) \quad \forall t \in \mathcal{I}\right] \\ F_Y(z) &= \mathbb{P}\left[Y(t) \le z(t) \quad \forall t \in \mathcal{I}\right] \end{aligned} \tag{3}$$

Empirical Distribution Functions

$$\begin{aligned}\hat{F}_X(z) &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left[X_i(t) \leq z(t) \quad \forall t \in \mathcal{I}\right] \\ \hat{F}_Y(z) &= \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}\left[Y_i(t) \leq z(t) \quad \forall t \in \mathcal{I}\right]\end{aligned} \tag{4}$$

Test statistic

$$\tau = \int_{\mathbb{L}^2(\mathcal{I})} \left[F_X(z) - F_Y(z)\right]^2 \mathrm{d}\mu(z) \tag{5}$$

Sample analog:

$$\tau_{n,m} = (n+m)\int_{\mathbb{L}^2(\mathcal{I})} \left[\hat{F}_X(z) - \hat{F}_Y(z)\right]^2 \mathrm{d}\mu(z) \tag{6}$$

6.4 Mean focused Test

Test statistic

$$\nu = \int_0^T \left[\mathbb{E}X(t) - \mathbb{E}Y(t)\right]^2 \mathrm{d}t \tag{7}$$

Sample Analog

$$\nu_{n,m} = (n+m)\int_0^T \left[\hat{\mathbb{E}}X(t) - \hat{\mathbb{E}}Y(t)\right]^2 \mathrm{d}t \tag{8}$$

6.5 Combined Permutation Test

6.6 Properties

7 Simulation Study

7.1 Use of High-Performance Computing

- bonna - HPC/A-Cluster der Universität Bonn
- `https://www.dice.uni-bonn.de/de/hpc/hpc-a-bonn/infrastruktur`

7.2 Simulation Setup

7.3 Results

8 Application

9 Bibliography

Anderson, T. W. and D. A. Darling (1952). "Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes". In: *The Annals of Mathematical Statistics* 23.2, pp. 193–212. DOI: 10.1214/aoms/1177729437.

Bugni, Federico A., Peter Hall, et al. (2009). "Goodness-of-fit tests for functional data". In: *The Econometrics Journal* 12.S1, S1–S18. ISSN: 1368-4221. URL: https://www.jstor.org/stable/23116593.

Bugni, Federico A. and Joel L. Horowitz (2021). "Permutation tests for equality of distributions of functional data". en. In: *Journal of Applied Econometrics* 36.7, pp. 861–877. DOI: 10.1002/jae.2846.

Büning, Herbert and Götz Trenkler (2013). *Nichtparametrische statistische Methoden*. De Gruyter. ISBN: 978-3-11-090299-0. DOI: 10.1515/9783110902990. URL: https://www.degruyter.com/document/doi/10.1515/9783110902990/html?lang=en.

Darling, D. A. (1957). "The Kolmogorov-Smirnov, Cramer-von Mises Tests". In: *The Annals of Mathematical Statistics* 28.4, pp. 823–838. DOI: 10.1214/aoms/1177706788.

Dunn, Olive Jean (1961). "Multiple Comparisons among Means". In: *Journal of the American Statistical Association* 56.293. Publisher: Taylor & Francis, pp. 52–64. ISSN: 0162-1459. DOI: 10.1080/01621459.1961.10482090.

Gibbons, Jean Dickinson and Subhabrata Chakraborti (2021). *Nonparametric statistical inference*. 6th edition. Boca Raton: CRC Press. ISBN: 978-1-138-08744-6.

Hsing, Tailen and Randall L. Eubank (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. Wiley series in probability and statistics. John Wiley and Sons, Inc. ISBN: 978-0-470-01691-6.

Kokoszka, Piotr and Matthew Reimherr (2021). *Introduction to functional data analysis*. First issued in paperback. Texts in statistical science series. CRC Press. ISBN: 978-1-03-209659-9 978-1-4987-4634-2.

Lehmann, E. L. and J. P. Romano (2005). *Testing Statistical Hypotheses*. en. Springer Texts in Statistics. Springer New York. ISBN: 978-0-387-98864-1 978-0-387-27605-2. DOI: 10.1007/0-387-27605-X.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer New York. ISBN: 978-0-387-40080-8 978-0-387-22751-1. DOI: 10.1007/b98888.

Vaart, Aad W. van der and Jon A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer New York. ISBN: 978-1-4757-2547-6 978-1-4757-2545-2. DOI: 10.1007/978-1-4757-2545-2.

10 Appendix

11 Versicherung an Eides statt

Ich versichere hiermit, dass ich die vorstehende Masterarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, dass die vorgelegte Arbeit noch an keiner anderen Hochschule zur Prüfung vorgelegt wurde und dass sie weder ganz noch in Teilen bereits veröffentlicht wurde. Wörtliche Zitate und Stellen, die anderen Werken dem Sinn nach entnommen sind, habe ich in jedem einzelnen Fall kenntlich gemacht.

Bonn, XX.XX.2021 ____
Jakob R. Juergens