# Remarks on Low-Dimensional Projections of High-Dimensional Distributions

Lutz Dümbgen* and Perla Zerial

December 6, 1996

**Abstract.** Let $P = P^{(q)}$ be a probability distribution on $q$-dimensional space. Necessary and sufficient conditions are derived under which a random $d$-dimensional projection of $P$ converges weakly to a fixed distribution $Q$ on $\mathbf{R}^d$ as $q$ tends to infinity, while $d$ is an arbitrary fixed number. This complements a well-known result of Diaconis and Freedman (1984). Further we investigate $d$-dimenional projections of $\widehat{P}$, where $\widehat{P}$ is the empirical distribution of a random sample from $P$ of size $n$. We prove a conditional Central Limit Theorem for random projections of $n^{1/2}(\widehat{P} - P)$ given the data $\widehat{P}$, as $q$ and $n$ tend to infinity.

*Correspondence to:* Lutz Dümbgen, Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, D-69120 Heidelberg, Germany;
`lutz@statlab.uni-heidelberg.de`

*Research supported in part by European Union Human Capital and Mobility Program ERB CHRX-CT 940693.

# 1 Introduction

A standard method of exploring high-dimensional datasets is to examine various low-dimensional projections thereof. In fact, many statistical procedures are based explicitly or implicitly on a "projection pursuit", cf. Huber (1985). Diaconis and Freedman (1984) showed that under weak regularity conditions on a distribution $P = P^{(q)}$ on $\mathbf{R}^q$, "most" $d$-dimensional orthonormal projections of $P$ are similar (in the weak topology) to a mixture of centered, spherically symmetric Gaussian distribution on $\mathbf{R}^d$ if $q$ tends to infinity while $d$ is fixed. A graphical demonstration of this disconcerting phenomenon is given by Buja *et al.* (1996). It should be pointed out that it is *not* a simple consequence of Poincaré's (1912) Lemma, although the latter is at the heart of the proof. The present paper provides further insight into this phenomenon. We extend Diaconis and Freedman's (1984) results in two directions.

Section 2 gives necessary and sufficient conditions on the sequence $(P^{(q)})_{q \geq d}$ such that "most" $d$-dimensional projections of $P$ are similar to some distribution $Q$ on $\mathbf{R}^d$. It turns out that these conditions are essentially the conditions of Diaconis and Freedman (1984). The novelty here is necessity. The limit distribution $Q$ is automatically a mixture of centered, spherically symmetric Gaussian distributions. The family of such measures arises in Eaton (1981) in another, related context.

More precisely, let $\Gamma = \Gamma^{(q)}$ be uniformly distributed on the set of column-wise orthonormal matrices in $\mathbf{R}^{q \times d}$ (cf. Section 4.2). Defining

$$\gamma^\top P \ := \ \mathcal{L}_{X \sim P}(\gamma^\top X)$$

for $\gamma \in \mathbf{R}^{d \times q}$, we investigate under what conditions the random distribution $\Gamma^\top P$ converges weakly in probability to an arbitrary fixed distribution $Q$ as $q \to \infty$, while $d$ is fixed.

Section 3 studies the difference between $P$ and the empirical distribution $\widehat{P} = \widehat{P}^{(q,n)}$ of $n$ independent random vectors with distribution $P$. Suppose that $(P^{(q)})_{q \geq d}$

satisfies the conditions of Section 2 and $\Gamma$ is independent from $\widehat{P}$. Then, as $n$ and $q$ tend to infinity, the standardized empirical measure $n^{1/2}(\Gamma^\top \widehat{P} - \Gamma^\top P)$ satisfies a conditional Central Limit Theorem *given the data* $\widehat{P}$.

Proofs are deferred to Section 4. The main ingredients are Poincaré's (1912) Lemma and a modification of a method invented by Hoeffding (1952) in order to prove weak convergence of conditional distributions, which is of independent interest. Further we utilize some results from the theory of empirical processes.

# 2 The Diaconis-Freedman Effect

Let us first settle on some terminology. A random distribution $\widehat{Q}$ on a separable metric space $(\mathbf{M}, \rho)$ is a mapping from some probability space into the set of Borel probability measures on $\mathbf{M}$ such that $\int f \, d\widehat{Q}$ is measurable for any function $f \in \mathcal{C}_b(\mathbf{M})$, the space of bounded, continuous functions on $\mathbf{R}^d$. We say that a sequence $(\widehat{Q}_k)_k$ of random distributions on $\mathbf{M}$ converges weakly in probability to some fixed distribution $Q$ if for each $f \in \mathcal{C}_b(\mathbf{M})$,

$$\int f \, d\widehat{Q}_k \ \to_{\mathrm{p}} \ \int f \, dQ \quad \text{as } k \to \infty.$$

In symbols, $\widehat{Q}_k \to_{\mathrm{w,p}} Q$ as $k \to \infty$. We say that the sequence $(\widehat{Q}_k)_k$ converges weakly in distribution to a random distribution $\widehat{Q}$ on $\mathbf{M}$ if for each $f \in \mathcal{C}_b(\mathbf{M})$,

$$\int f \, d\widehat{Q}_k \ \to_{\mathcal{L}} \ \int f \, d\widehat{Q} \quad \text{as } k \to \infty.$$

In symbols, $\widehat{Q}_k \to_{\mathrm{w},\mathcal{L}} \widehat{Q}$ as $k \to \infty$. Standard arguments show that $(\widehat{Q}_k)_k$ converges in probability to $Q$ if, and only if,

$$\sup_{f \in \mathcal{F}_{bL}} \Big| \int f \, d\widehat{Q}_k - \int f \, dQ \Big| \ \to_{\mathrm{p}} \ 0 \quad (k \to \infty),$$

where $\mathcal{F}_{bL}$ stands for the class of functions $f : \mathbf{M} \to [-1, 1]$ such that $|f(x) - f(y)| \leq \rho(x, y)$ for $x, y \in \mathbf{M}$.

Now we can state the first result.

**Theorem 2.1** *The following two assertions on the sequence $(P^{(q)})_{q \geq d}$ are equivalent:*

**(A1)** *There exists a probability measure $Q$ on $\mathbf{R}^d$ such that*

$$\Gamma^\top P \ \to_{\mathrm{w,p}} \ Q \quad \text{as } q \to \infty.$$

**(A2)** *If $X = X^{(q)}, \widetilde{X} = \widetilde{X}^{(q)}$ are independent random vectors with distribution $P$, then*

$$\mathcal{L}(q^{-1}\|X\|^2) \ \to_{\mathrm{w}} \ R \quad \text{and} \quad q^{-1}X^\top\widetilde{X} \ \to_{\mathrm{p}} \ 0 \quad \text{as } q \to \infty$$

*for some probability measure $R$ on $[0, \infty[$.*

*(Throughout, $\|x\|$ denotes Euclidean norm $(x^\top x)^{1/2}$.) The limit distribution $Q$ is equal to the normal mixture*

$$\int \mathcal{N}_d(0, \sigma^2 I)\, R(d\sigma^2).$$

**Corollary 2.2** *The random probability measure $\Gamma^\top P$ converges weakly to the standard Gaussian distribution $\mathcal{N}_d(0, I)$ in probability if, and only if, the following condition is satisfied:*

**(B)** *For independent random vectors $X = X^{(q)}, \widetilde{X} = \widetilde{X}^{(q)}$ with distribution $P$,*

$$q^{-1}\|X\|^2 \ \to_{\mathrm{p}} \ 1 \quad \text{and} \quad q^{-1}X^\top\widetilde{X} \ \to_{\mathrm{p}} \ 0 \quad \text{as } q \to \infty. \qquad \square$$

The implication "(A2) $\Longrightarrow$ (A1)" in Theorem 2.1 as well as sufficiency of condition (B) in Corollary 2.2 are due to Diaconis and Freedman (1984, Theorem 1.1 and Proposition 4.2).

**Example 2.3** Conditions (A1-2) are not very restrictive requirements. For instance, suppose that $P = \mathcal{L}\big((\mu_k + \sigma_k Z_k)_{1 \leq k \leq q}\big)$, where $(Z_k)_{k \geq 1}$ is a sequence of independent, identically distributed random variables with mean zero and variance one, and $\mu = \mu^{(q)} \in \mathbf{R}^q$, $\sigma = \sigma^{(q)} \in [0, \infty[^q$. Then conditions (A1-2) are satisfied if, and only if,

$$\textbf{(A3)} \qquad q^{-1}\|\mu\|^2 \ \to\ 0, \quad q^{-1}\|\sigma\|^2 \ \to\ r \geq 0 \quad \text{and} \quad q^{-1} \max_{1 \leq k \leq q} \sigma_k^2 \ \to\ 0$$

as $q \to \infty$, where $R = \delta_r$.

# 3 Empirical Distributions

In some sense Theorem 2.1 is a negative, though mathematically elegant result. It warns us against hasty conclusions about high-dimensional data sets after examining a couple of low-dimensional projections. In particular, one should not believe in multivariate normality only because several projections of the data "look normal". On the other hand, even small differences between different low-dimensional projections of $\widehat{P}$ may be intriguing. Therefore in the present section we study the relationship between projections of the empirical distribution $\widehat{P}$ and corresponding projections of $P$.

In particular, we are interested in the halfspace norm

$$\|\Gamma^\top \widehat{P} - \Gamma^\top P\|_{\mathrm{KS}} \ :=\ \sup_{\text{closed halfspaces } H \subset \mathbf{R}^d} |\Gamma^\top \widehat{P}(H) - \Gamma^\top P(H)|$$

of $\Gamma^\top \widehat{P} - \Gamma^\top P$. In case of $d = 1$ this is the usual Kolmogorov-Smirnov norm of $\Gamma^\top \widehat{P} - \Gamma^\top P$. In what follows we use several well-known results from empirical process theory. Instead of citing original papers in various places we simply refer to the excellent treatises of Pollard (1984) and van der Vaart and Wellner (1996). It is known that

$$\text{(3.1)} \qquad \mathbb{E} \sup_{\gamma \in \mathbf{R}^{q \times d}} \|\gamma^\top \widehat{P} - \gamma^\top P\|_{\mathrm{KS}} \ \leq\ C(q/n)^{1/2}$$

for some universal constant $C$. For the latter supremum is just the halfspace norm of $\widehat{P} - P$, and generally the set of closed halfspaces in $\mathbf{R}^k$ is a Vapnik-Cervonenkis class with Vapnik-Cervonenkis index $k+1$. Inequality (3.1) does not capture the *typical* deviation between $d$-dimensional projections of $\widehat{P}$ and $P$. In fact,

$$\sup_{\gamma \in \mathbf{R}^{q \times d}} \mathbb{E}\, \|\gamma^\top \widehat{P} - \gamma^\top P\|_{\mathrm{KS}} \;\le\; C(d/n)^{1/2}.$$

This implies that

$$\mathbb{E}\, \|\Gamma^\top \widehat{P} - \Gamma^\top P\|_{\mathrm{KS}} \;\le\; C(d/n)^{1/2}, \tag{3.2}$$

where the random projector $\Gamma$ and $\widehat{P}$ are always assumed to be stochastically independent. The subsequent results imply precise information about the *conditional* distribution of $n^{1/2}\|\Gamma^\top \widehat{P} - \Gamma^\top P\|_{\mathrm{KS}}$ *given the data* $\widehat{P}$. This point of view is natural in connection with exploratory projection pursuit. It turns out that under condition (B) of Corollary 2.2, this conditional distribution converges weakly in probability to a fixed distribution. Under the weaker conditions (A1-2) of Theorem 2.1 it converges weakly in distribution to a specific random distribution on the real line.

More generally, let $\mathcal{H}$ be a countable class of measurable functions from $\mathbf{R}^d$ into $[-1,1]$. Any finite signed measure $\nu$ on $\mathbf{R}^d$ defines an element $h \mapsto \nu(h) := \int h\, d\nu$ of the space $\ell_\infty(\mathcal{H})$ of all bounded functions on $\mathcal{H}$ equipped with supremum norm $\|z\|_{\mathcal{H}} := \sup_{h \in \mathcal{H}} |z(h)|$. We shall impose the following condition on the class $\mathcal{H}$ and some distribution $Q$ on $\mathbf{R}^d$.

**(C1)** There exists a countable subset $\mathcal{H}_o$ of $\mathcal{H}$ auch that each $h \in \mathcal{H}$ can be represented as pointwise limit of some sequence in $\mathcal{H}_o$.

**(C2)** The set $\mathcal{H}$ satisfies the uniform entropy condition

$$\int_0^1 \sqrt{\log(N(u,\mathcal{H})}\, du \;<\; \infty.$$

Here $N(u,\mathcal{H})$ is the supremum of $N(u,\mathcal{H},\widetilde{Q})$ over all probability measures $\widetilde{Q}$ on $\mathbf{R}^d$, and $N(u,\mathcal{H},\widetilde{Q})$ is the smallest number $m$ such that $\mathcal{H}$ can be covered with $m$ balls having radius $u$ with respect to the pseudodistance

$$\rho_{\widetilde{Q}}(g,h) \;:=\; \sqrt{Q((g-h)^2)}.$$

**(C3)** For any sequence $(Q_k)_k$ of probability measures converging weakly to $Q$,

$$\|Q_k - Q\|_{\mathcal{H}} \ \to \ 0 \quad \text{as } k \to \infty.$$

An example for conditions (C1-3) is the set $\mathcal{H}$ of (indicators of) closed halfspaces in $\mathbf{R}^d$ and any distribution $Q$ on $\mathbf{R}^d$ such that $Q(E) = 0$ for any hyperplane $E$ in $\mathbf{R}^d$. Here condition (C3) is a consequence of Billingsley and Topsoe's (1967) results.

Condition (C1) ensures that random elements such as $\|\Gamma^\top \widehat{P} - \Gamma^\top P\|_{\mathcal{H}}$ are measurable. A particular consequence of (C2) is existence of a centered Gaussian process $B_Q$ having uniformly continuous sample paths with respect to $\rho_Q$ and covariances

$$\mathbb{E}\, B_Q(g) B_Q(h) \ = \ Q(gh) - Q(g)Q(h).$$

This is proved via a Chaining argument. In the subsequent theorem we consider a decomposition of $B_Q$ as a sum $B_{Q,1} + B_{Q,2}$ of two independent centered Gaussian processes on $\mathcal{H}$. With the help of Anderson's (1955) Lemma or further application of Chaining one can show that $B_{Q,1}$ and $B_{Q,2}$ admit versions with uniformly continuous sample paths.

**Theorem 3.1** *Suppose that the sequence $(P^{(q)})_{q \geq d}$ satisfies conditions (A1-2) of Theorem 2.1, and suppose that conditions (C1-3) are satisfied with $Q$ being the corresponding limit measure $\int \mathcal{N}_d(0, \sigma^2 I)\, R(d\sigma^2)$. Define*

$$B^{(q,n)} \ := \ \Big(n^{1/2}(\Gamma^\top \widehat{P} - \Gamma^\top P)(h)\Big)_{h \in \mathcal{H}},$$

*and let $F$ be a continuous functional on $\ell_\infty(\mathcal{H})$ such that $F(B^{(q,n)})$ is measurable for all $q \geq d$ and $n \geq 1$. Then, as $n$ and $q$ tend to infinity,*

$$\mathcal{L}\Big(F(B^{(q,n)}) \,\Big|\, \widehat{P}\Big) \ \to_{\mathrm{w},\mathcal{L}} \ \mathcal{L}\Big(F(B_{Q,1} + B_{Q,2}) \,\Big|\, B_{Q,2}\Big),$$

*where $B_{Q,1}$ and $B_{Q,2}$ are two independent centered Gaussian processes having uniformly continuous sample paths with respect to $\rho_Q$ and covariances*

$$\mathbb{E}\, B_{Q,1}(g) B_{Q,1}(h) \ = \ Q(gh) - \int \mathcal{N}_d(0, \sigma^2 I)(g)\, \mathcal{N}_d(0, \sigma^2 I)(h)\, R(d\sigma^2)$$

$$= \int \Big( \mathcal{N}_d(0,\sigma^2 I)(gh) - \mathcal{N}_d(0,\sigma^2 I)(g)\, \mathcal{N}_d(0,\sigma^2 I)(h) \Big)\, R(d\sigma^2)$$

$$\mathrm{I\!E}\, B_{Q,2}(g) B_{Q,2}(h) = \int \mathcal{N}_d(0,\sigma^2 I)(g)\, \mathcal{N}_d(0,\sigma^2 I)(h)\, R(d\sigma^2) - Q(g)Q(h).$$

*(Thus $B_{Q,1} + B_{Q,2}$ defines a version of $B_Q$.)*

**Corollary 3.2** *Suppose that the sequence $(P^{(q)})_{q \geq d}$ satisfies condition (B) of Corollary 2.2, and suppose that conditions (C1-3) are satisfied for $Q = \mathcal{N}_d(0, I)$. Let $F$ be as in Theorem 3.1. Then, as $n$ and $q$ tend to infinity,*

$$\mathcal{L}\Big( F(B^{(q,n)}) \,\Big|\, \widehat{P} \Big) \ \to_{\mathrm{w,p}} \ \mathcal{L}\Big( F(B_Q) \Big). \qquad \Box$$

The measurability of $F(B^{(q,n)})$ can be dropped, provided that our definition of weak convergence of random distributions is suitably extended; see Remark 4.3 in Section 4.1.

# 4 Proofs

## 4.1 Hoeffding's (1952) technique and a modification thereof

In connection with randomization tests, Hoeffding (1952) observed that weak convergence of conditional distributions of test statistics is equivalent to the weak convergence of the *unconditional* distribution of suitable statistics in $\mathbf{R}^2$. His result can be extended straightforwardly as follows.

**Lemma 4.1** (Hoeffding). *For $k \geq 1$ let $X_k, \widetilde{X}_k \in \mathbf{X}_k$ and $T_k \in \mathbf{T}_k$ be independent random variables, where $X_k, \widetilde{X}_k$ are identically distributed. Further let $\gamma_k$ be some measurable mapping from $\mathbf{X}_k \times \mathbf{T}_k$ into the separable metric space $(\mathbf{M}, \rho)$, and let $Q$ be a fixed Borel probability measure on $\mathbf{M}$. Then, as $k \to \infty$, the following two assertions are equivalent:*

**(D1)**
$$\mathcal{L}\Big( \gamma_k(X_k, T_k) \,\Big|\, T_k \Big) \ \to_{\mathrm{w,p}} \ Q.$$

**(D2)**
$$\mathcal{L}\Big( \gamma_k(X_k, T_k), \gamma_k(\widetilde{X}_k, T_k) \Big) \ \to_{\mathrm{w}} \ Q \otimes Q.$$

An application of this equivalence with non-Euclidean spaces $\mathbf{M}$ is given by Romano (1989). We shall utilize Lemma 4.1 in order to prove Theorem 2.1. In connection with empirical measures we use the following modification of Lemma 4.1, which is of independent interest.

**Lemma 4.2** *For $\mathbf{k} \in \{1,2,3,\ldots\} \cup \{\infty\}$ let $X_{\mathbf{k}}, X_{\mathbf{k},1}, X_{\mathbf{k},2}, \ldots \in \mathbf{X_k}$ and $T_{\mathbf{k}} \in \mathbf{T_k}$ be independent random variables, where $X_{\mathbf{k}}, X_{\mathbf{k},1}, X_{\mathbf{k},2}, \ldots$ are identically distributed. Further let $\gamma_{\mathbf{k}}$ be some measurable mapping from $\mathbf{X_k} \times \mathbf{T_k}$ into $(\mathbf{M}, \rho)$. Then, as $k \to \infty$, the following two assertions are equivalent:*

(**E1**) $$\mathcal{L}\Big(\gamma_k(X_k, T_k) \Big| T_k\Big) \ \to_{\mathrm{w},\mathcal{L}} \ \mathcal{L}\Big(\gamma_\infty(X_\infty, T_\infty) \Big| T_\infty\Big).$$

(**E2**) *For any integer $L \geq 1$,* $\Big(\gamma_k(X_{k,\ell}, T_k)\Big)_{1\leq\ell\leq L} \ \to_{\mathcal{L}} \ \Big(\gamma_\infty(X_{\infty,\ell}, T_\infty)\Big)_{1\leq\ell\leq L}.$

**Remark 4.3** (Non-separablity and non-measurability). Suppose that the metric space $(\mathbf{M}, \rho)$ is possibly nonseparable, and that the mappings $\gamma_k$, $1 \leq k < \infty$, are possibly non-measurable. The implications "(D2) $\Longrightarrow$ (D1)" and "(E2) $\Longrightarrow$ (E1)" remain valid, provided that the limit distributions $Q$ in Lemma 4.1 and $\mathcal{L}(\gamma_\infty(X_\infty, T_\infty))$ in Lemma 4.2 have separable support, if one uses Hoffmann-Jorgensens notion of weak convergence (cf. van der Vaart and Wellner 1996, Chapter 1). The conditional distribution $\mathcal{L}\Big(\gamma_k(X_k, T_k) \Big| T_k = t_k\Big)$ stands for the outer measure $\mathbb{P}^*\Big\{\gamma_k(X_k, t_k) \in \cdot\Big\}$ on $\mathbf{M}$, and $\mathcal{L}\Big(\gamma_k(X_k, T_k) \Big| T_k\Big)$ is said to converge weakly to $Q$ in probability if for each fixed $f \in \mathcal{C}_b(\mathbf{M})$, the real-valued random element $\mathbb{E}^*\Big(f(\gamma_k(X_k, T_k)) \Big| T_k\Big)$ converges in outer probability to $Q(f)$. Analogously, $\mathcal{L}\Big(\gamma_k(X_k, T_k) \Big| T_k\Big)$ converges weakly in distribution to $\mathcal{L}\Big(\gamma_\infty(X_\infty, T_\infty) \Big| T_\infty\Big)$ if for any fixed $f \in \mathcal{C}_b(\mathbf{M})$, $\mathbb{E}^*\Big(f(\gamma_k(X_k, T_k)) \Big| T_k\Big)$ converges in distribution (in the sense of Hoffmann-Jorgensen) to the random variable $\mathbb{E}\Big(f(\gamma_\infty(X_\infty, T_\infty)) \Big| T_\infty\Big)$.

In this framework the reverse implications "(D1) $\Longrightarrow$ (D2)" and "(E1) $\Longrightarrow$ (E2)" remain valid under *some* measurability. For instance, these conclusions are correct, provided that for each $k \in \{1,2,3,\ldots\}$ the mapping $\gamma_k(X_k, T_k)$ is measurable with respect to the $\sigma$-field on $\mathbf{M}$ generated by closed balls with respect to $\rho$.

Given some familiarity with these concepts, one can easily adapt the subsequent proofs of Lemmas 4.1 and 4.2.

**Proof of Lemma 4.1.** Define $Y_k := \gamma_k(X_k, T_k)$ and $\widetilde{Y}_k := \gamma_k(\widetilde{X}_k, T_k)$. Suppose first that $\mathcal{L}(Y_k, \widetilde{Y}_k) \to_{\mathrm{w}} Q \otimes Q$. Then for any $f \in \mathcal{C}_b(\mathbf{M})$,

$$\begin{aligned}
&\mathbb{E}\Big(\big(\mathbb{E}(f(Y_k)\,|\,T_k) - Q(f)\big)^2\Big) \\
&\quad= \mathbb{E}\Big(\mathbb{E}(f(Y_k)\,|\,T_k)^2\Big) - 2Q(f)\,\mathbb{E}\,\mathbb{E}(f(Y_k)\,|\,T_k) + Q(f)^2 \\
&\quad= \mathbb{E}\,\mathbb{E}(f(Y_k)f(\widetilde{Y}_k)\,|\,T_k) - 2Q(f)\,\mathbb{E}\,\mathbb{E}(f(Y_k)\,|\,T_k) + Q(f)^2 \\
&\quad= \mathbb{E}(f(Y_k)f(\widetilde{Y}_k)) - 2Q(f)\mathbb{E}\,f(Y_k) + Q(f)^2 \\
&\quad\to \int f(y)f(\tilde{y})\,Q(dy)Q(d\tilde{y}) - Q(f)^2 \\
&\quad= 0.
\end{aligned}$$

Thus $\mathcal{L}(Y_k\,|\,T_k) \to_{\mathrm{w,p}} Q$.

On the other hand, suppose that $\mathcal{L}(Y_k\,|\,T_k) \to_{\mathrm{w,p}} Q$. Then for arbitrary $f, g \in \mathcal{C}_b(\mathbf{M})$,

$$\begin{aligned}
\mathbb{E}\,f(Y_k)g(\widetilde{Y}_k) &= \mathbb{E}\,\mathbb{E}\Big(f(Y_k)g(\widetilde{Y}_k)\,\Big|\,T_k\Big) \\
&= \mathbb{E}\Big(\mathbb{E}(f(Y_k)\,|\,T_k)\,\mathbb{E}(f(\widetilde{Y}_k)\,|\,T_k)\Big) \\
&\to Q(f)Q(g),
\end{aligned}$$

because $\mathbb{E}(h(Y_k)\,|\,T_k) \to_{\mathrm{p}} \int h\,dQ$ and $\Big|\mathbb{E}(h(Y_k)\,|\,T_k)\Big| \le \|h\|_\infty < \infty$ for each $h \in \mathcal{C}_b(\mathbf{M})$. Thus we know that $\mathbb{E}\,F(Y_k, \widetilde{Y}_k) \to \int F\,dQ \otimes Q$ for arbitrary functions $F(y, \tilde{y}) = f(y)g(\tilde{y})$ with $f, g \in \mathcal{C}_b(\mathbf{M})$. But this is known to be equivalent to weak convergence of $\mathcal{L}(Y_k, \widetilde{Y}_k)$ to $Q \otimes Q$; see van der Vaart and Wellner (1996, Chapter 1.4). $\Box$

**Proof of Lemma 4.2.** Define $Y_{\mathbf{k}} := \gamma_{\mathbf{k}}(X_{\mathbf{k}}, T_{\mathbf{k}})$ and $Y_{\mathbf{k},\ell} := \gamma_{\mathbf{k}}(X_{\mathbf{k},\ell}, T_{\mathbf{k}})$. Suppose first that $(Y_{k,\ell})_{1\le\ell\le L} \to_{\mathcal{L}} (Y_{\infty,\ell})_{1\le\ell\le L}$ for any integer $L \ge 1$. For arbitrary fixed

$f \in \mathcal{C}_b(\mathbf{M})$,

$$\begin{aligned}
&\mathbb{E}\Big(\Big(\mathbb{E}(f(Y_{\mathbf{k}}) \,|\, T_{\mathbf{k}}) - L^{-1}\sum_{\ell=1}^{L} f(Y_{\mathbf{k},\ell})\Big)^2\Big) \\
&= \ \mathbb{E}\,\mathbb{E}\Big(\Big(\mathbb{E}(f(Y_{\mathbf{k}}) \,|\, T_{\mathbf{k}}) - L^{-1}\sum_{\ell=1}^{L} f(Y_{\mathbf{k},\ell})\Big)^2 \Big|\, T_{\mathbf{k}}\Big) \\
&= \ \mathbb{E}\,\mathrm{Var}\Big(L^{-1}\sum_{\ell=1}^{L} f(Y_{\mathbf{k},\ell}) \Big|\, T_{\mathbf{k}}\Big) \\
&\leq \ L^{-1}\|f\|_\infty^2.
\end{aligned}$$

Thus the sample mean $L^{-1}\sum_{\ell=1}^{L} f(Y_{\mathbf{k},\ell})$ approximates the conditional expectation $\mathbb{E}(f(Y_{\mathbf{k}}) \,|\, T_{\mathbf{k}})$ arbitrarily well in quadratic mean, provided that $L$ is sufficiently large. However, the variable $L^{-1}\sum_{\ell=1}^{L} f(Y_{k,\ell})$ converges in distribution to $L^{-1}\sum_{\ell=1}^{L} f(Y_{\infty,\ell})$, according to the Continuous Mapping Theorem. Consequently, $\mathbb{E}(f(Y_k) \,|\, T_k)$ converges in distribution to $\mathbb{E}(f(Y_\infty) \,|\, T_\infty)$, whence $\mathcal{L}(Y_k \,|\, T_k) \to_{\mathrm{w},\mathcal{L}} \mathcal{L}(Y_\infty \,|\, T_\infty)$.

On the other hand, suppose that the conditional distribution $\mathcal{L}(Y_k \,|\, T_k)$ converges weakly in distribution to $\mathcal{L}(Y_\infty \,|\, T_\infty)$. In order to show that $(Y_{k,\ell})_{1\leq\ell\leq L}$ converges in distribution to $(Y_{\infty,\ell})_{1\leq\ell\leq L}$ one has to show that

$$\mathbb{E}\prod_{\ell=1}^{L} f_\ell(Y_{k,\ell}) \ \to \ \mathbb{E}\prod_{\ell=1}^{L} f_\ell(Y_{\infty,\ell})$$

for arbitrary functions $f_1, f_2, \ldots, f_\ell \in \mathcal{C}_b(\mathbf{M})$ (cf. van der Vaart and Wellner, 1996, Chapter 1.4). But

$$\mathbb{E}\prod_{\ell=1}^{L} f_\ell(Y_{\mathbf{k},\ell}) \ = \ \mathbb{E}\,\mathbb{E}\Big(\prod_{\ell=1}^{L} f_\ell(Y_{\mathbf{k},\ell}) \Big|\, T_{\mathbf{k}}\Big) \ = \ \mathbb{E}\ \prod_{\ell=1}^{L} \mathbb{E}(f_\ell(Y_{\mathbf{k}}) \,|\, T_{\mathbf{k}}).$$

Thus it suffices to show that $\Big(\mathbb{E}(f_\ell(Y_k) \,|\, T_k)\Big)_{1\leq\ell\leq L}$ converges in distribution to $\Big(\mathbb{E}(f_\ell(Y_\infty) \,|\, T_\infty)\Big)_{1\leq\ell\leq L}$. This follows easily from our assumption on $\mathcal{L}(Y_{\mathbf{k}} \,|\, T_{\mathbf{k}})$ via Fourier transformation, since for arbitrary $\lambda \in \mathbf{R}^L$,

$$\mathbb{E}\,\exp\Big(\sqrt{-1}\sum_{\ell=1}^{L} \lambda_\ell \mathbb{E}(f_\ell(Y_{\mathbf{k}}) \,|\, T_{\mathbf{k}})\Big) \ = \ \mathbb{E}\,\exp\Big(\sqrt{-1}\,\mathbb{E}(F(Y_{\mathbf{k}}) \,|\, T_{\mathbf{k}})\Big)$$

with $F := \sum_{\ell=1}^{L} \lambda_\ell f_\ell \in \mathcal{C}_b(\mathbf{M})$. $\square$

## 4.2 Proofs for Section 2

That $\Gamma = \Gamma^{(q)}$ is "uniformly" distributed on the set of column-wise orthonormal matrices in $\mathbf{R}^{q\times d}$ means that $\mathcal{L}(\tau\Gamma) = \mathcal{L}(\Gamma)$ for any fixed orthonormal matrix $\tau \in \mathbf{R}^{q\times q}$. For existence and uniqueness of the latter distribution we refer to Eaton (1989, Chapters 1 and 2). For the present purposes the following explicit construction described in Eaton (1989, Chapter 7) of $\Gamma$ is sufficient. Let $Z = Z^{(q)} := (Z_1, Z_2, \ldots, Z_d)$ be a random matrix in $\mathbf{R}^{q\times d}$ with independent, standard Gaussian column vectors $Z_j$ in $\mathbf{R}^q$. Then

$$\Gamma \;:=\; Z(Z^\top Z)^{-1/2}$$

has the desired distribution, and

$$\Gamma \;=\; q^{-1/2} Z\,(I + O_{\mathrm{p}}(q^{-1/2})) \quad \text{as } q \to \infty. \tag{4.1}$$

This equality can be viewed as an extension of Poincaré's (1912) Lemma.

**Proof of Theorem 2.1.** Let $\Gamma = \Gamma(Z)$ as above. Suppose that $Z = Z^{(q)}$, $X = X^{(q)}$ and $\widetilde{X} = \widetilde{X}^{(q)}$ are independent with $\mathcal{L}(X) = \mathcal{L}(\widetilde{X}) = P$, and let $Y, \widetilde{Y}$ be two independent random vectors in $\mathbf{R}^d$ with distribution $Q$. According to Lemma 4.1, condition (A1) is equivalent to

$$\textbf{(A1}'\textbf{)} \qquad \begin{pmatrix} \Gamma^\top X \\ \Gamma^\top \widetilde{X} \end{pmatrix} \;\to_{\mathcal{L}}\; \begin{pmatrix} Y \\ \widetilde{Y} \end{pmatrix}.$$

Because of equation (4.1) this can be rephrased as

$$\textbf{(A1}''\textbf{)} \qquad \begin{pmatrix} Y^{(q)} \\ \widetilde{Y}^{(q)} \end{pmatrix} \;:=\; \begin{pmatrix} q^{-1/2} Z^\top X \\ q^{-1/2} Z^\top \widetilde{X} \end{pmatrix} \;\to_{\mathcal{L}}\; \begin{pmatrix} Y \\ \widetilde{Y} \end{pmatrix}.$$

Now we prove equivalence of (A1") and (A2) starting from the observation that

$$\mathcal{L}\left(\begin{pmatrix} Y^{(q)} \\ \widetilde{Y}^{(q)} \end{pmatrix}\right) \;=\; \mathbb{E}\,\mathcal{L}\left(\begin{pmatrix} Y^{(q)} \\ \widetilde{Y}^{(q)} \end{pmatrix} \,\Big|\, X, \widetilde{X}\right) \;=\; \mathbb{E}\,\mathcal{N}_{2d}(0, \Sigma^{(q)}),$$

where

$$\Sigma^{(q)} \;:=\; \begin{pmatrix} q^{-1}\|X\|^2\, I & q^{-1} X^\top \widetilde{X}\, I \\ q^{-1} X^\top \widetilde{X}\, I & q^{-1}\|\widetilde{X}\|^2\, I \end{pmatrix} \;\in\; \mathbf{R}^{2d\times 2d}.$$

Suppose that condition (A2) holds. Then $\Sigma^{(q)}$ converges in distribution to a random diagonal matrix

$$\Sigma \ := \ \begin{pmatrix} S^2 I & 0 \\ 0 & \widetilde{S}^2 I \end{pmatrix}$$

with independent random variables $S^2, \widetilde{S}^2$ having distribution $R$. Clearly this implies that

$$\mathbb{E}\,\mathcal{N}_{2d}(0, \Sigma^{(q)}) \ \to_{\mathrm{w}} \ \mathbb{E}\,\mathcal{N}_{2d}(0, \Sigma) \ = \ \mathcal{L}\left(\begin{pmatrix} Y \\ \widetilde{Y} \end{pmatrix}\right)$$

with $Q = \mathbb{E}\,\mathcal{N}(0, S^2 I)$. Hence (A1") holds.

On the other hand, suppose that (A1") holds. For any $t = (t_1^\top, t_2^\top)^\top \in \mathbf{R}^{2d}$, the Fourier transform of $\mathcal{L}((Y^{(q)\top}, \widetilde{Y}^{(q)\top})^\top)$ at $t$ equals

$$\mathbb{E}\,\exp\Big(\sqrt{-1}\,(t_1^\top Y^{(q)} + t_2^\top \widetilde{Y}^{(q)})\Big) \ = \ \mathbb{E}\,\exp(-t^\top \Sigma^{(q)} t/2) \ = \ H^{(q)}(a(t)),$$

where $a(t) := \Big(-\|t_1\|^2/2, -\|t_2\|^2/2, -t_1^\top t_2\Big)^\top \in \mathbf{R}^3$, and

$$H^{(q)}(a) \ := \ \mathbb{E}\,\exp\Big(a_1 q^{-1}\|X\|^2 + a_2 q^{-1}\|\widetilde{X}\|^2 + a_3 q^{-1} X^\top \widetilde{X}\Big)$$

denotes the Laplace transform of $\mathcal{L}\Big(\Big(q^{-1}\|X\|^2, q^{-1}\|\widetilde{X}\|^2, q^{-1}X^\top\widetilde{X}\Big)^\top\Big)$ at $a \in \mathbf{R}^3$. By assumption, the Fourier transform at $t$ converges to

$$\mathbb{E}\,\exp(\sqrt{-1}\,t_1^\top Y)\mathbb{E}\,\exp(\sqrt{-1}\,t_2^\top Y).$$

Setting $t_2 = 0$ and varying $t_1$ shows that the Laplace transform of $\mathcal{L}(q^{-1}\|X\|^2)$ converges pointwise on $]-\infty, 0]$ to a continuous function. Hence $q^{-1}\|X\|^2$ converges in distribution to some random variable $S^2 \geq 0$, and $Q = \mathbb{E}\,\mathcal{N}_d(0, S^2 I)$. Therefore, if $\widetilde{S}^2$ denotes an independent copy of $S^2$, we know that $H^{(q)}(a(t))$ converges to

$$\mathbb{E}\,\exp(a_1(t)S^2)\mathbb{E}\,\exp(a_2(t)S^2) \ = \ \mathbb{E}\,\exp\Big(a_1(t)S^2 + a_2(t)\widetilde{S}^2 + a_3(t)\cdot 0\Big).$$

A problem at this point is that for dimension $d = 1$ the set $\{a(t) : t \in \mathbf{R}^{2d}\} \subset \mathbf{R}^3$ has empty interior. Thus we cannot apply the standard argument about weak convergence and convergence of Laplace transforms. However, letting $t_2 = \pm t_1$ with $\|t_1\|^2/2 = 1$, one may conclude that for arbitrary $\epsilon, r > 0$,

$$
\begin{aligned}
0 &= \lim_{q\to\infty} \Big( H^{(q)}(-1,-1,-2) + H^{(q)}(-1,-1,2) - 2H^{(q)}(-1,0,0)^2 \Big) \\
&= \lim_{q\to\infty} \Big( H^{(q)}(-1,-1,-2) + H^{(q)}(-1,-1,2) - 2\,\mathbb{E}\,\exp(-q^{-1}\|X\|^2 - q^{-1}\|\widetilde{X}\|^2) \Big) \\
&= 2 \lim_{q\to\infty} \,\mathbb{E}\,\exp\Big(-q^{-1}\|X\|^2 - q^{-1}\|\widetilde{X}\|^2\Big)\Big(\cosh(2q^{-1}X^\top\widetilde{X}) - 1\Big) \\
&\geq 2\exp(-2r)(\cosh(2\epsilon) - 1)\ \cdot \\
&\qquad \cdot \limsup_{q\to\infty}\ \mathbb{P}\Big\{ q^{-1}\|X\|^2 < r, q^{-1}\|\widetilde{X}\|^2 < r, |q^{-1}X^\top\widetilde{X}| \geq \epsilon \Big\} \\
&\geq 2\exp(-2r)(\cosh(2\epsilon) - 1) \limsup_{q\to\infty} \Big( \mathbb{P}\{|q^{-1}X^\top\widetilde{X}| \geq \epsilon\} - 2\,\mathbb{P}\{q^{-1}\|X\|^2 \geq r\} \Big) \\
&\geq 2\exp(-2r)(\cosh(2\epsilon) - 1) \Big( \limsup_{q\to\infty}\ \mathbb{P}\{|q^{-1}X^\top\widetilde{X}| \geq \epsilon\} - 2\,\mathbb{P}\{S^2 \geq r\} \Big),
\end{aligned}
$$

whence

$$\limsup_{q\to\infty}\ \mathbb{P}\{|q^{-1}X^\top\widetilde{X}| \geq \epsilon\} \ \leq\ 2\,\mathbb{P}\{S^2 \geq r\}.$$

Consequently, $q^{-1}X^\top\widetilde{X} \to_{\mathrm{p}} 0$. $\square$

**Proof of equivalence of (A1-2) and (A3).** Proving that (A3) implies (A1-2) is elementary. In order to show that (A1-2) implies (A3) note first that conditions (A1-2) for the distributions $P^{(q)}$ imply the same conditions for the symmetrized distributions

$$P_o = P_o^{(q)} \ := \ \mathcal{L}_{(X,\widetilde{X})\sim P\otimes P}(X - \widetilde{X}) \ = \ \mathcal{L}\Big( (\sigma_k(Z_k - Z_{q+k})_{1\leq k\leq q} \Big).$$

Condition (A2) for these distributions reads as follows.

$$\mathcal{L}\Big( q^{-1}\sum_{k=1}^{q} \sigma_k^2 (Z_k - Z_{q+k})^2 \Big) \ \to_{\mathrm{w}} \ R_o = R\star R \quad \text{and} \tag{4.2}$$

$$q^{-1}\sum_{k=1}^{q} \sigma_k^2 (Z_k - Z_{q+k})(Z_{2q+k} - Z_{3q+k}) \ \to_{\mathrm{p}} \ 0. \tag{4.3}$$

The summands $q^{-1}\sigma_k^2(Z_k - Z_{q+k})(Z_{2q+k} - Z_{3q+k})$, $1 \leq k \leq q$, in (4.3) are independent and symmetrically distributed. Therefore one can easily deduce from (4.3) that

$q^{-1} \max_{1 \le k \le q} \sigma_k^2 \to 0$. But then

$$q^{-1} \sum_{k=1}^{q} \sigma_k^2 (Z_k - Z_{q+k})^2 \ = \ 2q^{-1} \|\sigma\|^2 + o_{\mathrm{p}}(1 + q^{-1}\|\sigma\|^2),$$

and one can deduce from (4.2) that $q^{-1}\|\sigma^{(q)}\|^2$ converges to some fixed number $r$; in particular, $R = \delta_r$. Now we return to the original distributions $P$. Here the second half of (A2) means that

$$\begin{aligned} & q^{-1} \sum_{k=1}^{k} (\mu_k + \sigma_k Z_k)(\mu_k + \sigma_k Z_{q+k}) \\ = \ & q^{-1}\|\mu\|^2 + q^{-1} \sum_{k=1}^{q} \mu_k \sigma_k (Z_k + Z_{q+k}) + q^{-1} \sum_{k=1}^{q} \sigma_k^2 Z_k Z_{q+k} \\ = \ & o_{\mathrm{p}}(1). \end{aligned}$$

Since

$$\begin{aligned} \mathbb{E}\Big(\Big(q^{-1} \sum_{k=1}^{q} \mu_k \sigma_k (Z_k + Z_{q+k})\Big)^2\Big) \ &= \ q^{-2} \sum_{k=1}^{q} \mu_k^2 \sigma_k^2 \ = \ o(q^{-1}\|\mu\|^2), \\ \mathbb{E}\Big(\Big(q^{-1} \sum_{k=1}^{q} \sigma_k^2 Z_k Z_{q+k}\Big)^2\Big) \ &= \ q^{-2} \sum_{k=1}^{q} \sigma_k^4 \ \to \ 0, \end{aligned}$$

it follows that $q^{-1}\|\mu\|^2 \to 0$. □

## 4.3 Proof of Theorem 3.1

Let $(\Gamma^{(q,\ell)})_{\ell \ge 1}$ be a sequence of independent copies of $\Gamma$ which is stochastically independent from $\widehat{P}$. Define

$$B^{(q,n,\ell)} \ := \ \Big(n^{1/2}(\Gamma^{(q,\ell)\top}\widehat{P} - \Gamma^{(q,\ell)\top}P)(h)\Big)_{h \in \mathcal{H}}.$$

The $B^{(q,n,\ell)}$, $\ell \ge 1$, are *dependent* copies of $B^{(q,n)}$. Further consider independent processes $B_{Q,1}^{(1)}, B_{Q,1}^{(2)}, B_{Q,1}^{(3)}, \ldots$ and $B_{Q,2}$ with $\mathcal{L}(B_{Q,1}^{(\ell)}) = \mathcal{L}(B_{Q,1})$ and $\mathcal{L}(B_{Q,2})$ as described in Theorem 3.1. According to Lemma 4.2 it suffices to show that for any fixed integer $L \ge 1$ and $\Lambda := \{1, 2, \ldots, L\}$, the random elements

$$\vec{B}^{(q,n)} \ := \ \Big(B^{(q,n,\ell)}(h)\Big)_{(\ell,h) \in \Lambda \times \mathcal{H}}$$

converge in distribution in $\ell_\infty(\Lambda \times \mathcal{H})$ to

$$\vec{B} \ := \ \Big((B_{Q,1}^{(\ell)} + B_{Q,2})(h)\Big)_{(\ell,h)\in\Lambda\times\mathcal{H}}$$

as $q \to \infty$ and $n \to \infty$. For that purpose it suffices to verify the following two claims.

**(F1)** As $q \to \infty$ and $n \to \infty$, the finite-dimensional marginal distributions of the process $\vec{B}^{(q,n)}$ converge to the corresponding finite-dimensional distributions of $\vec{B}$.

**(F2)** As $q \to \infty$, $n \to \infty$ and $\delta \downarrow 0$,

$$\max_{\ell\in\Lambda} \sup_{g,h\in\mathcal{H}:\rho_Q(g,h)<\delta} \Big|B^{(q,n,\ell)}(g) - B^{(q,n,\ell)}(h)\Big| \ \to_{\mathrm{p}} \ 0.$$

In order to verify assertions (F1-2) we consider the conditional distribution of $\vec{B}^{(q,n)}$ given the random matrix

$$\vec{\Gamma} = \vec{\Gamma}^{(q)} \ := \ (\Gamma^{(q,1)}, \Gamma^{(q,2)}, \ldots, \Gamma^{(q,L)}) \ \in \ \mathbf{R}^{q\times Ld}.$$

In fact, if we define

$$\vec{f}_{\ell,h}(v) \ := \ h(v_\ell) \quad \text{for } v = (v_1^\top, \ldots, v_L^\top)^\top \in \mathbf{R}^{Ld},$$

then

$$B^{(q,n,\ell)}(h) \ = \ n^{1/2}(\vec{\Gamma}^\top\widehat{P} - \vec{\Gamma}^\top P)(\vec{f}_{\ell,h}).$$

Thus $\mathcal{L}(\vec{B}^{(q,n)} \,|\, \vec{\Gamma})$ is essentially the distribution of an empirical process based on $n$ independent random vectors with distribution $\vec{\Gamma}^\top P$ on $\mathbf{R}^{Ld}$ and indexed by the family $\vec{\mathcal{H}} := \{\vec{f}_{\ell,h} : \ell \in \Lambda, h \in \mathcal{H}\}$.

The multivariate version of Lindeberg's Central Limit Theorem entails that for large $q$ and $n$ the finite-dimensional marginal distributions of $\vec{B}^{(q,n)}$, conditional on $\vec{\Gamma}$, can be approximated by the corresponding finite-dimensional distributions of a centered Gaussian process on $\Lambda \times \mathcal{H}$ with the same covariance function, namely

$$\begin{aligned}
\Sigma^{(q)}\Big((\ell,g),(m,h)\Big) \ &:= \ \mathrm{Cov}\Big(B^{(q,n,\ell)}(g), B^{(q,n,m)}(h) \,\Big|\, \vec{\Gamma}\Big) \\
&= \ \vec{\Gamma}^\top P(\vec{f}_{\ell,g}\vec{f}_{m,h}) - \vec{\Gamma}^\top P(\vec{f}_{\ell,g})\vec{\Gamma}^\top P(\vec{f}_{m,h}).
\end{aligned}$$

It follows from equality (4.1) and the proof of Theorem 2.1 that

$$\vec{\Gamma}^\top P \ \to_{\mathrm{w}} \ \vec{Q} \ := \ \int \mathcal{N}_{Ld}(0, \sigma^2 I)\, R(d\sigma^2) \quad \text{as } q \to \infty.$$

This implies that the conditional covariance function $\Sigma^{(q)}$ converges pointwise in probability to the covariance function $\Sigma$, where

$$\begin{aligned}
\Sigma\Big((\ell, g), (m, h)\Big) \ &:= \ \vec{Q}(\vec{f}_{\ell,g}\vec{f}_{m,h}) - \vec{Q}(\vec{f}_{\ell,g})\vec{Q}(\vec{f}_{m,h}) \\
&= \ \int \mathcal{N}_{Ld}(0, \sigma^2)(\vec{f}_{\ell,g}\vec{f}_{m,h})\, R(d\sigma^2) - Q(g)Q(h) \\
&= \ \begin{cases} \displaystyle\int \mathcal{N}_d(0, \sigma^2)(gh)\, R(d\sigma^2) - Q(g)Q(h) & \text{if } \ell = m, \\ \displaystyle\int \mathcal{N}_d(0, \sigma^2)(g)\mathcal{N}_d(0, \sigma^2)(h)\, R(d\sigma^2) - Q(g)Q(h) & \text{if } \ell \neq m, \end{cases} \\
&= \ \mathrm{Cov}\Big((B_{Q,1}^{(\ell)} + B_{Q,2})(g), (B_{Q,1}^{(m)} + B_{Q,2})(h)\Big)
\end{aligned}$$

as $q \to \infty$. If we can show that even the supremum of $|\Sigma^{(q)} - \Sigma|$ over $(\Lambda \times \mathcal{H})^2$ tends to zero in probability, then assertion (F1) follows.

As mentioned in Section 2, weak convergence in probability is metrizable. Therefore there exist events $A^{(q)}$ depending only on $\vec{\Gamma}^{(q)}$ such that for arbitrary $\vec{f} \in \mathcal{C}_b(\mathbf{R}^{Ld})$,

$$\mathrm{I\!P}(A^{(q)}) \ \to \ 1 \quad \text{and} \quad \sup_{A^{(q)}} \Big| \int \vec{f}\, d\vec{\Gamma}^\top P - \int \vec{f}\, d\vec{Q} \Big| \ \to \ 0 \quad \text{as } q \to \infty.$$

It was shown by Billingsley and Topsoe (1967) that condition (C3) is equivalent to

$$(4.4) \quad \lim_{\delta \downarrow 0} \sup_{h \in \mathcal{H}} Q\Big\{y \in \mathbf{R}^d : \sup_{z : \|z - y\| < \delta} |h(z) - h(y)| > \epsilon\Big\} \ = \ 0 \quad \text{for any } \epsilon > 0.$$

Note that the $d$-dimensional marginal distributions of $\vec{Q}$ are just $Q$. Therefore one can easily deduce from (4.4) that

$$\lim_{\delta \downarrow 0} \sup_{\vec{g}, \vec{h} \in \vec{\mathcal{H}}} \vec{Q}\Big\{v \in \mathbf{R}^{Ld} : \sup_{w : \|w - v\| < \delta} |\vec{g}\vec{h}(w) - \vec{g}\vec{h}(v)| > \epsilon\Big\} \ = \ 0 \quad \text{for any } \epsilon > 0.$$

Now a second application of Billingsley and Topsoe (1967) shows that

$$(4.5) \qquad \sup_{A^{(q)}} \sup_{\vec{g}, \vec{h} \in \vec{\mathcal{H}}} |\vec{\Gamma}^\top P(\vec{g}\vec{h}) - \vec{Q}(\vec{g}\vec{h})| \ \to \ 0 \quad \text{as } q \to \infty.$$

We assume without loss of generality that $\mathcal{H}$ (and thus $\vec{\mathcal{H}}$) contains the constant function 1. Then fact (4.5) clearly implies uniform convergence of $\Sigma^{(q)}$ to $\Sigma$ on $\Lambda \times \mathcal{H}$ as $q \to \infty$ in probability.

As for assertion (F2), it is well-known from empirical process theory that conditions (C1-2) imply that for arbitrary fixed $\epsilon > 0$ and $\ell \in \Lambda$,

$$\mathbb{P}\Big( \sup_{g,h \in \mathcal{H} : \rho^{(q,\ell)}(g,h) < \delta} \Big| B^{(q,n,\ell)}(g) - B^{(q,n,\ell)}(h) \Big| \geq \epsilon \,\Big|\, \vec{\Gamma} \Big) \ \to_{\mathrm{p}} \ 0 \tag{4.6}$$

as $q \to \infty$, $n \to \infty$ and $\delta \downarrow 0$. Here

$$\rho^{(q,\ell)}(g,h) \ := \ \sqrt{\vec{\Gamma}^\top P((\vec{f}_{\ell,g} - \vec{f}_{\ell,h})^2)}.$$

But it follows from (4.5) that

$$\sup_{A^{(q)}} \sup_{g,h \in \mathcal{H}} |\rho^{(q,\ell)}(g,h)^2 - \rho_Q(g,h)^2| \ \to \ 0$$

as $q \to \infty$, for any fixed $\ell \in \Lambda$. Hence one may replace $\rho^{(q,\ell)}$ in (4.6) with $\rho_Q$ and obtains assertion (F2). $\square$

# References


Anderson, T.W. (1955). The integral of a symmetric unimodal function over a symmetric convex set and some probability inequalities. *Proc. Amer. Math. Soc.* **6**, 170-176

Billingsley, P. and F. Topsoe (1967). Uniformity in weak convergence. *Z. Wahrschein. verw. Geb.* **7**, 1-16.

Buja, A., D. Cook and D.F. Swayne (1996). Interactive High-Dimensional Data Visualization. *J. Comp. Graph. Statist.* **5**, 78-99

Diaconis, P. and D. Freedman (1984). Asymptotics of graphical projection pursuit. *Ann. Statist.* **12**, 793-815

EATON, M.L. (1981). On the projections of isotropic distributions. *Ann. Statist.* **9**, 391-400

EATON, M.L. (1989). *Group Invariance Applications in Statistics.* Regional Conf. Series Prob. Statist. **1**, IMS

HOEFFDING, W. (1952). The large-sample power of tests based on random permutations. *Ann. Math. Statist.* **23**, 169-192

HUBER, P.J. (1985). Projection pursuit. *Ann. Statist.* **13**, 435-475

POINCARÉ, H. (1912). *Calcul des Probabilités.* Hermann, Paris

POLLARD, D. (1984). *Convergence of Stochastic Processes.* Springer, New York

ROMANO, J.P. (1989). Bootstrap and randomization tests of some nonparametric hypotheses. *Ann. Statist.* **17**, 141-159

VAN DER VAART, A.W. AND J.A. WELLNER (1996). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer, New York