APPENDICES

A. Proof on proper metric d_{ext}

PROOF. We prove that d_{ext} is a proper metric by verifying three properties:

1. (Positive Definiteness) It is apparent that $d_{ext}(f_1, f_2) \geq 0$. By Theorem 1, there exists γ_{12}, such that $f_1 = (f_2; \gamma_{12})$. Therefore, $d_{ext}(f_1, f_2) = 0 \Leftrightarrow \|1 - \sqrt{\dot{\gamma}_{12}}\| = 0 \Leftrightarrow \gamma_{12}(t) = \gamma_{id}$. Hence, $f_1 = f_2$.
2. (Symmetry) $\|1 - \sqrt{\dot{\gamma}_{21}}\|^2 = \|1 - \sqrt{\dot{\gamma}_{12}^{-1}}\|^2 = \int_0^1 \left(1 - \sqrt{\dot{\gamma}_{12}^{-1}(s)}\right)^2 ds = \|1 - \sqrt{\dot{\gamma}_{12}(t)}\|^2$. Therefore, $d_{ext}(f_1, f_2) = d_{ext}(f_2, f_1)$.
3. (Triangle Inequality) Let $f_2 = f_1(\gamma_{12}(t))$, $f_3 = f_2(\gamma_{23}(t))$, $\gamma_{13} = \gamma_{12} \circ \gamma_{23}$. Then, $d_{ext}(f_1, f_3) = \|1 - \sqrt{\dot{\gamma}_{13}}\| = \|1 - \sqrt{(\dot{\gamma}_{12} \circ \gamma_{23})\, \dot{\gamma}_{23}}\| \leq \|1 - \sqrt{\dot{\gamma}_{23}}\| + \|\sqrt{\dot{\gamma}_{23}} - \sqrt{\dot{\gamma}_{13}}\| = \|1 - \sqrt{\dot{\gamma}_{23}}\| + \|1 - \sqrt{\dot{\gamma}_{12}}\|$. Note that $\|\sqrt{\dot{\gamma}_{23}} - \sqrt{\dot{\gamma}_{13}}\| = \|(1, \gamma_{23}) - (1, \gamma_{13})\| = \|(1, \gamma_{23}) - (1, \gamma_{12} \circ \gamma_{23})\| = \|1 - \sqrt{\dot{\gamma}_{12}}\|$ (by isometry) Thus, $d_{ext}(f_1, f_3) \leq d_{ext}(f_1, f_2) + d_{ext}(f_2, f_3)$.

B. Proof on the consistency of $\hat{f}$

LEMMA 1. *Let g be a probability density function on [0, 1]. $\{X_i\}_{i=1}^n$ are a set of i.i.d. random variables with density g. If $\hat{g}_n$ is a modified kernel estimate with optimal bandwidth given in Algorithm 2, then*

$$\int_0^1 |\hat{g}_n(t) - g(t)| dt \xrightarrow{a.s.} 0 \; (\textit{when } n \to \infty)$$

PROOF. Let $\tilde{g}_n(t) = \frac{1}{nh_n} \sum_{i=1}^n K(\frac{t - X_i}{h_n})$ be the classical kernel estimator with kernel function K and optimal bandwidth h_n (i.e. $h_n \to 0$ and $nh_n \to \infty$). Then, we can obtain from Equation 3.84 of that $\int_0^1 |\tilde{g}_n(t) - g(t)| dt \xrightarrow{a.s.} 0$.

As $K(t) = 0$ when $|t| < 1$, we have

$$\begin{aligned}
& \int_0^1 |\hat{g}_n(t) - \tilde{g}_n(t)| dt \\
= \; & \int_0^{h_n} |\hat{g}_n(t) - \tilde{g}_n(t)| dt + \int_{1-h_n}^1 |\hat{g}_n(t) - \tilde{g}_n(t)| dt \\
& + \int_{h_n}^{1-h_n} |\frac{1}{n+1} \tilde{g}_n(t) + \frac{1}{n+1}| dt \\
\leq \; & \int_0^{h_n} \hat{g}_n(t) dt + \int_0^{h_n} \tilde{g}_n(t) dt + \int_{1-h_n}^1 \hat{g}_n(t) dt + \int_{1-h_n}^1 \tilde{g}_n(t) dt + \frac{2}{n+1}. \qquad (1)
\end{aligned}$$

Here we will show that the first term goes to 0 (a.s.). Indeed,

$$\begin{aligned}
\int_0^{h_n} \hat{g}_n(t) dt &= \int_0^{h_n} \frac{1}{nh_n} \sum_{i=1}^n K(\frac{t - X_i}{h_n}) dt = \int_0^{h_n} \frac{1}{nh_n} \sum_{X_i \leq 2h_n} K(\frac{t - X_i}{h_n}) dt \\
&\leq \int_0^1 \frac{1}{nh_n} \sum_{X_i \leq 2h_n} K(\frac{t - X_i}{h_n}) dt = \frac{1}{n} \sum_{X_i \leq 2h_n} 1 = \frac{1}{n} \sum_{i=0}^1 \mathbf{1}_{\{X_i \leq 2h_n\}}
\end{aligned}$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function. By the Strong Law of Large Numbers on triangular arrays,

$$\frac{1}{n}\sum_{i=0}^{n}(\mathbf{1}_{\{X_i\le 2h_n\}} - E\mathbf{1}_{\{X_i\le 2h_n\}}) \to 0.(a.s.)$$

As $E\mathbf{1}_{\{X_i\le 2h_n\}} = \int_0^{2h_n} f(t)dt \to 0$, we have $\int_0^{h_n} \hat{g}_n(t)dt \xrightarrow{a.s.} 0$. The convergence to 0 for the second to fourth terms on the RHS of Eqn. (1) can be similarly proven, and therefore $\int_0^1 |\hat{g}_n(t) - \tilde{g}_n(t)|dt \xrightarrow{a.s.} 0$. Finally, we have

$$\int_0^1 |\hat{g}_n(t) - g(t)|dt \le \int_0^1 |\hat{g}_n(t) - \tilde{g}_n(t)|dt + \int_0^1 |\tilde{g}_n(t) - g(t)|dt \xrightarrow{a.s} 0.$$

LEMMA 2. *Let G and $\hat{G}_n$ denote the cumulative distribution functions of g and $\hat{g}_n$ in Lemma 1, respectively. Assume the density g is continuous and for any $t \in [0,1]$, $0 < m \le g(t) \le M < \infty$ (Condition 2 in Section 2.2). If G and $\hat{G}_n$ are invertible and the inverse functions are differentiable, then*

$$\int_0^1 (\sqrt{\dot{\hat{G}}_n^{-1}(t)} - \sqrt{\dot{G}^{-1}(t)})^2 dt \xrightarrow{a.s.} 0 \ (\textit{when } n \to \infty)$$

PROOF. To simplify notation, we let $F = G^{-1}$, $\hat{F}_n = \hat{G}_n^{-1}$, $f = \dot{F} = \dot{G}^{-1}$, and $\hat{f}_n = \dot{\hat{F}}_n = \dot{\hat{G}}_n^{-1}$. For any $t \in [0,1], |\hat{G}_n(t) - G(t)| \le \int_0^1 |\hat{g}_n(t) - g(t)|dt \xrightarrow{a.s.} 0$ (by Lemma 3). That is, $\hat{G}_n \rightrightarrows G$ (uniform convergence) almost surely. By the theory on convergence of inverse functions, we also got that $\hat{F}_n \rightrightarrows F\ (a.s.)$.

By definition, $G(F(t)) = t$ and $\hat{G}_n(\hat{F}_n(t)) = t$. Using the chain rule, we have $g(F(t))f(t) = 1$ and $\hat{g}_n(\hat{F}_n(t))\hat{f}_n(t) = 1$. Therefore,

$$\begin{aligned}
& \int_0^1 |\hat{f}_n(t) - f(t)|dt \\
= & \int_0^1 |\frac{1}{\hat{g}_n(\hat{F}_n(t))} - \frac{1}{g(F(t))}|dt \\
\le & \int_0^1 |\frac{1}{\hat{g}_n(\hat{F}_n(t))} - \frac{1}{g(\hat{F}_n(t))}|dt + \int_0^1 |\frac{1}{g(\hat{F}_n(t))} - \frac{1}{g(F(t))}|dt
\end{aligned}$$

Here we will show that each integration in the right-hand side indeed converges to 0 (a.s.). By Lemma 1,

$$\begin{aligned}
& \int_0^1 |\frac{1}{\hat{g}_n(\hat{F}_n(t))} - \frac{1}{g(\hat{F}_n(t))}|dt \\
= & \int_0^1 |\frac{1}{\hat{g}_n(s)} - \frac{1}{g(s)}|\hat{g}_n(s)ds \ \text{(by change of variable)} \\
= & \int_0^1 \frac{1}{g(s)}|\hat{g}_n(s) - g(s)|ds \le \frac{1}{m}\int_0^1 |\hat{g}_n(s) - g(s)|ds \xrightarrow{a.s.} 0
\end{aligned}$$

By assumption, g is continuous and positively bounded. Hence, $1/g$ is also continuous. This continuity is uniform because the domain $[0,1]$ is compact. That is, for any $\epsilon > 0$, there exists $\delta > 0$, such that for all $a, b \in [0,1]$ with $|a-b| < \delta, |1/g(a) - 1/g(b)| < \epsilon$. We have shown that $\hat{F}_n \rightrightarrows F\ (a.s.)$. Hence, with probability 1, there exists an integer N such that for any $n > N$ and $t \in [0,1]$, we have $|\hat{F}_n(t) - F(t)| < \delta$. $\int_0^1 |\frac{1}{g(\hat{F}_n(t))} - \frac{1}{g(F(t))}|dt \leq \int_0^1 \epsilon dt = \epsilon$. Therefore, we have shown that

$$\int_0^1 |\frac{1}{g(\hat{F}_n(t))} - \frac{1}{g(F(t))}|dt \xrightarrow{a.s.} 0.$$

Finally, based on the simple inequality $(\sqrt{a} - \sqrt{b})^2 \leq |a-b|$, we have

$$\int_0^1 (\sqrt{\hat{f}_n(t)} - \sqrt{f(t)})^2 dt \leq \int_0^1 |\hat{f}_n(t) - f(t)|dt \xrightarrow{a.s.} 0.$$

LEMMA 3. *Let $\{\gamma_i\}$ be a sequence of warping functions that satisfy Condition 3 in Section 2.2, and $\bar{\gamma}$ be the Karcher mean of $\{\gamma_i^{-1}\}$. Then $\bar{\gamma}$ converges to γ_{id} almost surely. That is,*

$$||(1, \bar{\gamma}) - 1|| \xrightarrow{a.s.} 0 \quad (\textit{when } n \to \infty)$$

PROOF. By assumption, $E(\sqrt{\dot{\gamma}_i^{-1}(t)}) \equiv \beta > 0, i = 1, \cdots, n$. Let $S_n = n\bar{\gamma} = \sum_{i=1}^n \sqrt{\dot{\gamma}_i^{-1}}$. As $\{\sqrt{\dot{\gamma}_i^{-1}}\}$ are i.i.d.,

$$E\left(\|S_n - n\beta\|^4\right) = E\left(\left\|\sum_{i=1}^n \left(\sqrt{\dot{\gamma}_i^{-1}} - \beta\right)\right\|^4\right)$$

$$= nE\left(\left\|\sqrt{\dot{\gamma}_1^{-1}} - \beta\right\|^4\right) + n(n-1)\left(E\left(\left\|\sqrt{\dot{\gamma}_1^{-1}} - \beta\right\|^2\right)\right)^2$$

$$+2n(n-1)\left(E\left(\int_0^1 \left(\sqrt{\dot{\gamma}_1^{-1}(t)} - \beta\right)\left(\sqrt{\dot{\gamma}_2^{-1}(t)} - \beta\right)dt\right)^2\right)$$

As $||\sqrt{\dot{\gamma}_1^{-1}}|| = 1$, there exist positive constants C and N, such that $E(||S_n - n\beta||^4) < Cn$ when $n > N$.

Using the generalized Chebyshev inequality, for any $\epsilon > 0$ and $n > N$,

$$P\left(\left\|\frac{S_n - n\beta}{n}\right\| > \epsilon\right) \leq \frac{1}{(n\epsilon)^4} E(||S_n - n\beta||^4) \leq \frac{C}{\epsilon^4 n^2}.$$

This indicates that $\sum_{n=1}^\infty P(||S_n - n\beta|| \geq n\epsilon) < \infty$. By the Borel-Cantelli lemma, $P(||S_n - n\beta|| \geq n\epsilon\ \ i.o.) = 0$. Therefore, $||\frac{1}{n}\sum_{i=1}^n \sqrt{\dot{\gamma}_i^{-1}} - \beta|| \xrightarrow{a.s.} 0$. Finally, we have

$$||(1, \bar{\gamma}) - 1|| \;=\; ||\sqrt{\dot{\bar{\gamma}}} - 1|| = \left\|\frac{\frac{1}{n}\sum_{i=1}^n \sqrt{\dot{\gamma}_i^{-1}}}{\frac{1}{n}||\sum_{i=1}^n \sqrt{\dot{\gamma}_i^{-1}}||} - 1\right\| \xrightarrow{a.s.} 0.$$

LEMMA 4. *Assume Y_m is a random variable following a Poisson distribution with mean Λ_m. If $\lambda_m \geq \alpha \log(m), \alpha > 1$ for sufficiently large m (Condition 4 in Section 2.2), then $Y_m \to \infty$ $(a.s.)$ when $m \to \infty$.*

PROOF. Based on the Poisson density formula, for any $K = 1, 2, \cdots$, $P(Y_m \leq K) = e^{-\lambda_m} \sum_{k=0}^{K} \frac{\lambda_m^k}{k!}$. By assumption, $\lambda_m \geq \alpha \log(m), \alpha > 1$ for sufficiently large m. It is apparent that when m is sufficiently large, $e^{-\frac{\alpha/2}{1+\alpha}\lambda_m} \sum_{k=0}^{K} \frac{\lambda_m^k}{k!} < 1$. Hence,

$$\begin{aligned} m^{1+\alpha/2} P(Y_m \leq K) &= m^{1+\alpha/2} e^{-\lambda_m} \sum_{k=0}^{K} \frac{\lambda_m^k}{k!} \\ &= e^{-\frac{1+\alpha/2}{1+\alpha}\lambda_m + (1+\alpha/2)\log m} \left(e^{-\frac{\alpha/2}{1+\alpha}\lambda_m} \sum_{k=0}^{K} \frac{\lambda_m^k}{k!} \right) \\ &\leq 1 \cdot 1 = 1. \end{aligned}$$

Consequently, for sufficiently large m, $P(Y_m \leq K) \leq \frac{1}{m^{1+\alpha/2}}$. Hence, $\sum_{m=1}^{\infty} P(Y_m \leq K) < \infty$. By the Borel-Cantelli lemma,

$$P(\limsup\{Y_m \leq K\}) = P(Y_m \leq K \ \ i.o.) = 0.$$

Equivalently, we have $P(Y_m > K \ \text{ eventually}) = 1$, for any $K = 1, 2, ...$ Therefore,

$$\lim_{m \to \infty} Y_m = \infty. \ \ (a.s.)$$