Lemma 1. Functional Delta Method (univariate). *Suppose $\mathcal{P}_n$ is the empirical distribution of a random sample $X_1, \dots, X_n$ from a distribution P, and ϕ is a function that maps the distribution of interest into some space. Define the Gateaus derivative*

$$\phi'_P(\delta_x - P) = \frac{d}{dt}\,|_{t=0}\ \phi((1-t)P + t\delta_x) = IF_{\phi,P}(x),$$

which is also the Influence Function, and $\gamma^2 = \int IF_{\phi,P}(x)^2 dP$. If integration and differentiation can be exchanged, then

$$\int \phi'_P(\delta_x - P)dP = 0.$$

Further, if $\sqrt{n}R_n \xrightarrow{P} 0$, where

$$R_n = \phi(\mathcal{P}_n) - \phi(P) - \frac{1}{n}\sum_i \phi'_P(\delta_{x_i} - P),$$

then from the Central Limit Theory that

$$\sqrt{n}(\phi(\mathcal{P}_n) - \phi(P)) \xrightarrow{d} N(0, \gamma^2).$$

Lemma 2. Functional Delta Method (multivariate). *Suppose $\mathcal{P}_n$ is the empirical distribution of a random sample $\mathbf{X}_1, \dots, \mathbf{X}_n$ from a distribution $\mathbf{P}$, and $\phi : \mathcal{R}^d \to \mathcal{R}^k$. Define the Gateaus derivative*

$$\phi'_{\mathbf{P}}(\delta_{\mathbf{x}} - \mathbf{P}) = \frac{d}{dt}\,|_{t=0}\ \phi((1-t)\mathbf{P} + t\delta_{\mathbf{x}}) = IF_{\phi,\mathbf{P}}(\mathbf{x}),$$

which is also the Influence Function, and $V_0 = \int \langle [IF_{\phi,\mathbf{P}}(\mathbf{x})]_{p1}, [IF_{\phi,\mathbf{P}}(\mathbf{x})]_{p2} \rangle d\mathbf{P}$. If integration and differentiation can be exchanged, then

$$\int \phi'_{\mathbf{P}}(\delta_{\mathbf{x}} - \mathbf{P})d\mathbf{P} = 0.$$

Further, if $\sqrt{n}\mathbf{R}_n \xrightarrow{P} 0$, where

$$\mathbf{R}_n = \phi(\mathcal{P}_n) - \phi(\mathbf{P}) - \frac{1}{n}\sum_i \phi'_{\mathbf{P}}(\delta_{\mathbf{x}_i} - \mathbf{P}),$$

then from the Central Limit Theory that

$$\sqrt{n}(\phi(\mathcal{P}_n) - \phi(\mathbf{P})) \xrightarrow{d} MVN(0, V_0).$$

Theorem 1. Asymptotic Distribution of Variable Importance (univariate). *Denote* $D_p : f \to \frac{\partial}{\partial x_p} f$ *the differentiation operator,* $H_p = D_p^\top D_p$ *the inner product of* D_p*, define*

$$\psi_p(f) = \|D_p(f)\|_n^2 = \langle D_p f, D_p f \rangle = f^\top H_p f, \tag{1}$$

then

$$\sqrt{n}(\psi_p(\hat{f}_n) - \psi_p(f_0)) \xrightarrow{d} N(0, 4\sigma^2 \|H_p f_0\|_n^2),$$

where σ^2 *is the variance of the observations.*

Proof. From the definition in (1), we have

$$\psi_p'(f) = \frac{\partial}{\partial f} \psi_p(f) = 2H_p f.$$

Define a mean function $m : F \to E(F)$, where F is the distribution. Then in our case, $f_0 = E(F) = m(F)$. According to **Lemma 1**, we have

$$\psi_p(f_0) = \psi_p(E(F)) = \psi_p(m(F)) = \phi(F),$$

i.e., $\phi(\cdot) = \psi_p(m(\cdot))$. Therefore,

$$\begin{aligned}
\phi_F'(\delta_x - F) &= \psi_p'(m(\delta_x - F)) \\
&= \frac{d}{dt} |_{t=0}\ \psi_p(m((1-t)F + t\delta_x)) \\
&= \frac{d}{dt} |_{t=0}\ \psi_p((1-t)f_0 + tx) \\
&= \frac{d}{dt} |_{t=0}\ [(1-t)f_0 + tx]^\top H_p [(1-t)f_0 + tx] \\
&= 2(x - f_0)^\top H_p f_0 \\
&= IF_{\phi,F}(x).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\gamma^2 &= \int IF_{\phi,F}(x)^2 dF \\
&= 4 \int f_0^\top H_p (x - f_0)(x - f_0)^\top H_p f_0 dF \\
&= 4\sigma^2 \|H_p f_0\|_n^2,
\end{aligned}$$

where $\sigma^2 \mathbf{I}_n = E[(x - f_0)(x - f_0)^\top]$ is the variance of the observations. Moreover, we have

$$\int \phi_F'(\delta_x - F) dF = 2 \int (x - f_0)^\top H_p f_0 dF = 0,$$

since x is an unbiased estimator of f_0 and

$$\begin{aligned}
\sqrt{n}R_n &= \sqrt{n}[\phi(\mathcal{F}_n) - \phi(F) - \frac{1}{n}\sum_i \phi'_F(\delta_{x_i} - F)] \\
&= \sqrt{n}[\psi_p(\hat{f}_n) - \psi_p(f_0) - \frac{2}{n}\sum_i (x_i - f_0)^\top H_p f_0] \\
&= \sqrt{n}[\hat{f}_n^\top H_p \hat{f}_n - f_0^\top H_p f_0 - \frac{2}{n}\sum_i (x_i - f_0)^\top H_p f_0] \\
&= \sqrt{n}[\hat{f}_n^\top H_p \hat{f}_n - \hat{f}_n^\top H_p f_0 + \hat{f}_n^\top H_p f_0 - f_0^\top H_p f_0 - \frac{2}{n}\sum_i (x_i - f_0)^\top H_p f_0] \\
&= \sqrt{n}[\frac{1}{n}\sum_i (\hat{f}_n - f_0)^\top H_p (\hat{f}_n - f_0) - \frac{2}{n}\sum_i (x_i - f_0)^\top H_p f_0] \\
&\xrightarrow{P} 0?
\end{aligned}$$

Therefore, by **Lemma 1**, we have

$$\sqrt{n}(\psi_p(\hat{f}_n) - \psi_p(f_0)) \xrightarrow{d} N(0, 4\sigma^2 \|H_p f_0\|_n^2).$$

□

Theorem 2. Asymptotic Distribution of Variable Importance (multivariate). *Denote* $\boldsymbol{\psi} = [\psi_1, \ldots, \psi_P]$ *for* ψ_p *as defined in* (1)*, then* $\boldsymbol{\psi}(\hat{\mathbf{f}}_n)$ *asymptotically converges toward a multivariate normal distribution surrounding* $\boldsymbol{\psi}(\mathbf{f}_0)$*, i.e.,*

$$\sqrt{n}(\boldsymbol{\psi}(\hat{\mathbf{f}}_n) - \boldsymbol{\psi}(\mathbf{f}_0)) \xrightarrow{d} MVN(\mathbf{0}, V_0),$$

where V_0 *is a* $P \times P$ *matrix such that* $[V_0]_{p1,p2} = 4\sigma^2_{p1,p2}\langle H_{p1}[\mathbf{f}_0]_{p1}, H_{p2}[\mathbf{f}_0]_{p2}\rangle_n$.

Proof. Define a mean function $\mathbf{m} : \mathbf{F} \to E(\mathbf{F})$, where $\mathbf{F} \in \mathcal{R}^P$ is the distribution. Then in our case, $\mathbf{f}_0 = E(\mathbf{F}) = \mathbf{m}(\mathbf{F})$. According to **Lemma 2**, we have

$$[\boldsymbol{\psi}(\mathbf{f}_0)]_p = \psi_p(E(\mathbf{F})) = \psi_p([\mathbf{m}(\mathbf{F})]_p) = [\boldsymbol{\phi}(\mathbf{F})]_p,$$

i.e., $\boldsymbol{\phi}(\cdot) = \boldsymbol{\psi}(\mathbf{m}(\cdot))$ and $[\boldsymbol{\phi}(\cdot)]_p = \psi_p([\mathbf{m}(\cdot)]_p)$, where $\boldsymbol{\phi} : \mathcal{R}^P \to \mathcal{R}^P$. Therefore,

$$\begin{aligned}
[\boldsymbol{\phi}'_{\mathbf{F}}(\delta_{\mathbf{x}} - \mathbf{F})]_p &= \psi'_p([\mathbf{m}(\delta_{\mathbf{x}} - \mathbf{F})]_p) \\
&= \frac{d}{dt}\,|_{t=0}\ \psi_p([\mathbf{m}((1-t)\mathbf{F} + t\delta_{\mathbf{x}})]_p) \\
&= \frac{d}{dt}\,|_{t=0}\ \psi_p([(1-t)\mathbf{f}_0 + t\mathbf{x}]_p) \\
&= \frac{d}{dt}\,|_{t=0}\ [(1-t)[\mathbf{f}_0]_p + t[\mathbf{x}]_p]^\top H_p[(1-t)[\mathbf{f}_0]_p + t[\mathbf{x}]_p] \\
&= 2([\mathbf{x}]_p - [\mathbf{f}_0]_p)^\top H_p [\mathbf{f}_0]_p \\
&= [IF_{\phi,\mathbf{F}}(\mathbf{x})]_p.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
V_0 &= \int \langle [IF_{\phi,\mathbf{F}}(\mathbf{x})]_{p1}, [IF_{\phi,\mathbf{F}}(\mathbf{x})]_{p2}\rangle d\mathbf{F} \\
&= 4\int [\mathbf{f}_0]_{p1}^\top H_{p1}([\mathbf{x}]_{p1} - [\mathbf{f}_0]_{p1})([\mathbf{x}]_{p2} - [\mathbf{f}_0]_{p2})^\top H_{p2}[\mathbf{f}_0]_{p2} d\mathbf{F} \\
&= 4\sigma^2_{p1,p2}\langle H_{p1}[\mathbf{f}_0]_{p1}, H_{p2}[\mathbf{f}_0]_{p2}\rangle_n,
\end{aligned}$$

where $\sigma^2_{p1,p2}\mathbf{I}_n = \mathrm{E}[([\mathbf{x}]_{p1} - [\mathbf{f}_0]_{p1})([\mathbf{x}]_{p2} - [\mathbf{f}_0]_{p2})^\top]$ is the covariance of the $p1^{th}$ and $p2^{th}$ features of the observations. Moreover, we have

$$[\int \boldsymbol{\phi}'_{\mathbf{F}}(\delta_{\mathbf{x}} - \mathbf{F})d\mathbf{F}]_p = 2\int([\mathbf{x}]_p - [\mathbf{f}_0]_p)^\top H_p[\mathbf{f}_0]_p d\mathbf{F} = 0,$$

since $\mathbf{x}$ is an unbiased estimator of $\mathbf{f}_0$ and

$$\begin{aligned}
[\sqrt{n}\mathbf{R}_n]_p &= \sqrt{n}[[\boldsymbol{\phi}(\mathcal{F}_n)]_p - [\boldsymbol{\phi}(\mathbf{F})]_p - \frac{1}{n}\sum_i [\boldsymbol{\phi}'_{\mathbf{F}}(\delta_{\mathbf{x}_i} - \mathbf{F})]_p] \\
&= \sqrt{n}[[\hat{\mathbf{f}}_n]_p^\top H_p[\hat{\mathbf{f}}_n]_p - [\mathbf{f}_0]_p^\top H_p[\mathbf{f}_0]_p - \frac{2}{n}\sum_i([\mathbf{x}_i]_p - [\mathbf{f}_0]_p)^\top H_p[\mathbf{f}_0]_p] \\
&= \sqrt{n}[\frac{1}{n}\sum_i([\hat{\mathbf{f}}_n]_p - [\mathbf{f}_0]_p)^\top H_p([\hat{\mathbf{f}}_n]_p - [\mathbf{f}_0]_p) - \frac{2}{n}\sum_i([\mathbf{x}_i]_p - [\mathbf{f}_0]_p)^\top H_p[\mathbf{f}_0]_p] \\
&\xrightarrow{P} 0?
\end{aligned}$$

Therefore, by **Lemma 2**, we have

$$\sqrt{n}(\boldsymbol{\psi}(\hat{\mathbf{f}}_n) - \boldsymbol{\psi}(\mathbf{f}_0)) \xrightarrow{d} \mathrm{MVN}(\mathbf{0}, V_0),$$

where V_0 is a $P \times P$ matrix such that $[V_0]_{p1,p2} = 4\sigma^2_{p1,p2}\langle H_{p1}[\mathbf{f}_0]_{p1}, H_{p2}[\mathbf{f}_0]_{p2}\rangle_n$. $\square$