Notations. Soit $Z = \hat{f}(x_0)$, fonction de l'échantillon $(\mathcal{X}, \mathcal{Y})$, avec $\mathcal{X} = (X_1, \dots, X_n)$ et $\mathcal{Y} = (Y_1, \dots, Y_n)$. On pose

$$m(\mathcal{X}) = \mathbb{E}_{\mathcal{Y}|\mathcal{X}}[Z], \qquad v(\mathcal{X}) = \mathrm{Var}_{\mathcal{Y}|\mathcal{X}}(Z).$$

Ce sont des fonctions de $\mathcal{X}$, donc des variables aléatoires.

Loi de la tour. Pour toute variable W intégrable,

$$\mathbb{E}[W] = \mathbb{E}_{\mathcal{X}}\big[\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[W]\big].$$

On calcule en deux temps : on fixe $\mathcal{X}$ et on moyenne sur $\mathcal{Y}$, puis on moyenne sur $\mathcal{X}$.

1. Biais. La loi de la tour appliquée à $W = Z$ donne

$$\mathbb{E}[Z] = \mathbb{E}_{\mathcal{X}}\big[\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[Z]\big] = \mathbb{E}_{\mathcal{X}}[m(\mathcal{X})].$$

Le biais inconditionnel étant $\mathbb{E}[Z] - f(x_0)$,

$$\mathrm{Bias}^2(x_0) = \big(f(x_0) - \mathbb{E}[Z]\big)^2 = \Big(f(x_0) - \mathbb{E}_{\mathcal{X}}\big[\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[\hat{f}(x_0)]\big]\Big)^2.$$

2. Variance. *(a) Ce que vaut Z^2 sachant $\mathcal{X}$.* La variance conditionnelle est la variance usuelle calculée avec la loi de Z sachant $\mathcal{X}$:

$$v(\mathcal{X}) = \mathbb{E}_{\mathcal{Y}|\mathcal{X}}[Z^2] - \big(\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[Z]\big)^2 = \mathbb{E}_{\mathcal{Y}|\mathcal{X}}[Z^2] - m(\mathcal{X})^2,$$

donc

$$\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[Z^2] = v(\mathcal{X}) + m(\mathcal{X})^2.$$

(b) Loi de la tour appliquée à $W = Z^2$.

$$\mathbb{E}[Z^2] = \mathbb{E}_{\mathcal{X}}\big[\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[Z^2]\big] = \mathbb{E}_{\mathcal{X}}\big[v(\mathcal{X}) + m(\mathcal{X})^2\big] = \mathbb{E}_{\mathcal{X}}[v(\mathcal{X})] + \mathbb{E}_{\mathcal{X}}[m(\mathcal{X})^2].$$

(c) Conclusion. Avec $\mathbb{E}[Z] = \mathbb{E}_{\mathcal{X}}[m(\mathcal{X})]$ (étape 1),

$$\begin{aligned}
\mathrm{Var}(Z) &= \mathbb{E}[Z^2] - \big(\mathbb{E}[Z]\big)^2 \\
&= \mathbb{E}_{\mathcal{X}}[v(\mathcal{X})] + \underbrace{\mathbb{E}_{\mathcal{X}}[m(\mathcal{X})^2] - \big(\mathbb{E}_{\mathcal{X}}[m(\mathcal{X})]\big)^2}_{=\mathrm{Var}_{\mathcal{X}}(m(\mathcal{X}))} \\
&= \mathbb{E}_{\mathcal{X}}\big[\,\mathrm{Var}_{\mathcal{Y}|\mathcal{X}}(\hat{f}(x_0))\big] + \mathrm{Var}_{\mathcal{X}}\left(\mathbb{E}_{\mathcal{Y}|\mathcal{X}}[\hat{f}(x_0)]\right).
\end{aligned}$$

En mots : variance totale = moyenne des variances conditionnelles + variance des moyennes conditionnelles.

Exemple chiffré. X vaut 0 ou 1 avec probabilité $\frac{1}{2}$; $Z \mid X = 0$ de moyenne 0 et de variance 1 ; $Z \mid X = 1$ de moyenne 2 et de variance 1. Alors $v = 1$ et $m^2 \in \{0, 4\}$, d'où $\mathbb{E}[Z^2] = \frac{1}{2}(1+0) + \frac{1}{2}(1+4) = 3$, $\mathbb{E}[Z] = 1$ et $\mathrm{Var}\, Z = 2$. Avec la formule : $\mathbb{E}[v] = 1$ et $\mathrm{Var}(m) = \mathrm{Var}(\{0, 2\}) = 1$, soit $1 + 1 = 2$.

3. Cas du k-NN. Avec $Y_i = f(X_i) + \varepsilon_i$, ε_i i.i.d. indépendants de $\mathcal{X}$, $\mathbb{E}\varepsilon_i = 0$, $\mathrm{Var}\,\varepsilon_i = \sigma^2$, et $\hat{f}(x_0) = \frac{1}{k}\sum_{i \in N_k(x_0)} Y_i$ où $N_k(x_0)$ ne dépend que de $\mathcal{X}$ (constant sachant $\mathcal{X}$) :

$$m(\mathcal{X}) = \frac{1}{k}\sum_{i \in N_k(x_0)} f(X_i), \qquad v(\mathcal{X}) = \frac{1}{k^2}\sum_{i \in N_k(x_0)} \mathrm{Var}(Y_i \mid \mathcal{X}) = \frac{\sigma^2}{k}.$$

D'où

$$\mathrm{Var}(x_0) = \frac{\sigma^2}{k} + \mathrm{Var}_{\mathcal{X}}\Big(\frac{1}{k}\sum_{i \in N_k(x_0)} f(X_i)\Big).$$

4. EQM. Comme $\mathbb{E}_{\mathcal{X}}\big[(f(x_0) - m)^2\big] = \text{Bias}^2(x_0) + \text{Var}_{\mathcal{X}}(m)$,

$$\mathbb{E}\big[(\hat{f}(x_0) - f(x_0))^2\big] = \text{Bias}^2(x_0) + \mathbb{E}_{\mathcal{X}}[v] + \text{Var}_{\mathcal{X}}(m).$$