Applications - Lecture 2

Pointwise Estimation in Parametric Models

Ingénierie Mathématique & Informatique

02/10/2026

Section	Notion	Used in
2.1	Bias, quadratic risk (MSE), consistency	App. 1 (bias-variance)
2.2	Parametric models & moment estimators	App. 2 (R_0)
2.3.2	Asymptotic normality & the Delta method	App. 2 (R_0)

1. **Decompose** the prediction error of a machine learning model using the bias-variance decomposition of the MSE.
2. **Build** a moment estimator of the epidemic reproduction number R_0, then get its asymptotic variance with the Delta method.

App. 1 - Bias-Variance Tradeoff in Machine Learning

From an abstract estimator to a prediction

Problem. A learner fits $\hat{f}$ on a training set of size n and wants to predict a fresh noisy label $y = f^*(x) + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ independent of the training set, at a fixed test point x.

Here $\mathrm{QoI} = f^*(x) \in \mathbb{R}$ and $Z_n = \hat{f}(x)$ is an estimator of it, with bias $b(Z_n) = \mathbb{E}[\hat{f}(x)] - f^*(x)$. Proposition 2.1.11 decomposes its MSE (Def. 2.1.10) as

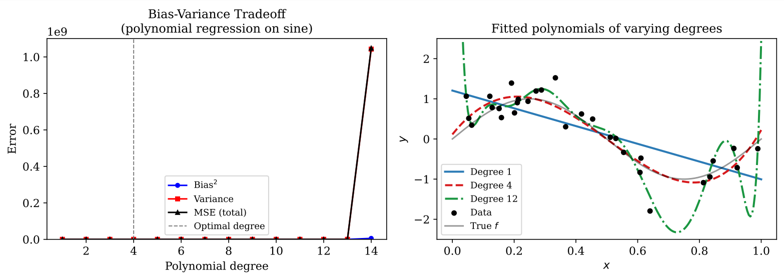
$$\mathrm{MSE}(Z_n) = b(Z_n)^2 + \mathrm{Var}(\hat{f}(x)).$$

Since ε is independent of $\hat{f}(x)$ with mean 0, the total prediction error adds the irreducible noise term:

$$\mathbb{E}[(y - \hat{f}(x))^2] = \underbrace{b(Z_n)^2}_{\mathrm{Bias}^2} + \underbrace{\mathrm{Var}(\hat{f}(x))}_{\text{Variance}} + \sigma^2.$$

For polynomial regression of degree d, raising d shrinks Bias^2 but inflates the Variance

Simulation and take-aways



Left: Bias^2, Variance, and total MSE vs. degree d ($n = 30$, $f^*(x) = \sin(2\pi x)$, $\varepsilon \sim \mathcal{N}(0, 0.09)$); minimum at $d^* = 4$. Right: fitted polynomials of degree 1, 4, 12 against the true f^* (black).

- Bias and variance cannot be minimised simultaneously: $d = 1$ underfits (high bias), $d = 12$ overfits (high variance).
- Regularisation (ridge, lasso, early stopping) trades a little bias for a large drop in variance, lowering the total MSE.

App. 2 - Estimating the Reproduction Number R_0

The SIR model and a moment estimator of R_0

Problem. From early case counts, estimate the basic reproduction number R_0 (average secondary infections per infectious individual): $R_0 > 1$ means the epidemic spreads.

Model. In the SIR model with transmission rate β and recovery rate γ, $R_0 = \beta/\gamma$. While $S \approx N$, infections grow as $I(t) \approx I_0 e^{rt}$ with $r = \gamma(R_0 - 1)$, so $R_0 = g(r) := 1 + r/\gamma$ (γ known).

Step 1. Let $W_1, \ldots, W_n$ be the daily log-increments of new cases during the exponential phase, with $\mathbb{E}_r[W_1] = r$: taking $\varphi(x) = x$, $m(r) = r$, the moment estimator is the trivial one, $\hat{r}_n = \overline{W}_n$, strongly consistent by the LLN.

Step 2. The plug-in estimator of $R_0 = g(r)$ is

$$\hat{R}_{0,n} = g(\hat{r}_n) = 1 + \frac{\overline{W}_n}{\gamma}.$$

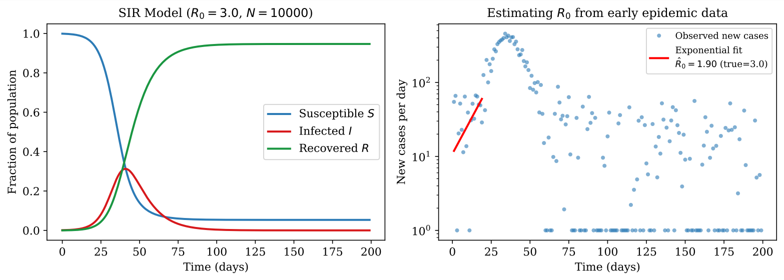
Step 3: the Delta method gives the asymptotic variance

By the CLT, $\sqrt{n}(\hat{r}_n - r) \to Y \sim \mathcal{N}(0, \sigma^2)$ in distribution, where $\sigma^2 = \mathrm{Var}_r(W_1)$. Since g is C^1 with $g'(r) = 1/\gamma$, the Delta method applies with $\zeta_n = \hat{r}_n$, $a = r$, $\varphi = g$:

$$\sqrt{n}\big(\hat{R}_{0,n} - R_0\big) = \sqrt{n}\big(g(\hat{r}_n) - g(r)\big) \longrightarrow g'(r)\, Y \sim \mathcal{N}\Big(0,\ \frac{\sigma^2}{\gamma^2}\Big).$$

So $\hat{R}_{0,n}$ is asymptotically normal with asymptotic variance $K(R_0) = \sigma^2/\gamma^2$

Fitting on simulated data and take-aways



Left: SIR trajectories ($N = 10\,000$, $R_0 = 3$, $\beta = 0.3\,\mathrm{day}^{-1}$, $\gamma = 0.1\,\mathrm{day}^{-1}$). Right: new cases (semi-log) with the log-linear fit on the first 20 days used to recover $\hat{R}_0$.

- The Delta method turns a CLT for the growth rate r into a CLT for the (nonlinear) quantity of interest $R_0 = g(r)$
- Valid only in the exponential phase ($S \approx N$); a wrong γ also biases $\hat{R}_{0,n}$ directly, since it enters g.