MA8701 Advanced methods in statistical inference and learning

L5: Solutions to Exercise random forest variance

Mette Langaas IMF/NTNU

21 February, 2021

Contents

The variance of the average of B observations of i.i.d random variables X, each with variance σ^2 is $\frac{\sigma^2}{B}$. Now, suppose we have B observations of a random variable X which are identically distributed, each with mean μ and variance σ^2, but not independent.

That is, suppose the variables have a positive correlation ρ

$$\text{Cov}(X_i, X_j) = \rho\sigma^2, \quad i \neq j.$$

then the variance of the average is

$$\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2$$

Proof:

$$\begin{aligned}
\text{Var}(\bar{X}) &= \text{Var}\Big(\frac{1}{B}\sum_{i=1}^{B} X_i\Big) \\
&= \sum_{i=1}^{B} \frac{1}{B^2}\text{Var}(X_i) + 2\sum_{i=2}^{B}\sum_{j=1}^{i-1} \frac{1}{B}\frac{1}{B}\text{Cov}(X_i, X_j) \\
&= \frac{1}{B}\sigma^2 + 2\frac{B(B-1)}{2}\frac{1}{B^2}\rho\sigma^2 \\
&= \frac{1}{B}\sigma^2 + \rho\sigma^2 - \frac{1}{B}\rho\sigma^2 \\
&= \rho\sigma^2 + \frac{1-\rho}{B}\sigma^2 \\
&= \frac{1-(1-B)\rho}{B}\sigma^2 \\
\\
&= \rho\sigma^2 + \frac{1-\rho}{B}\sigma^2
\end{aligned}$$