Exercise 8.4

- consider the bagging estimator $\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}(x_b^*)$
- let $\hat{f}(x)$ be the B-spline estimator $\hat{f}(x) = Hy$

where $H = b(x)(B^T B)^{-1} B^T$ idempotent $H^2 = H$

- consider the parametric bootstrap $y_b^* = \hat{f}(x) + \varepsilon_b^*$, $\varepsilon_b^* \sim N(0, \sigma^2)$

→ show that $\hat{f}_{bag}(x) \xrightarrow{B\to\infty} \hat{f}(x)$

$$\hat{f}(x) = Hy$$

$$\hat{f}(x_b^*) = Hy^* = H\left(\hat{f}(x) + \varepsilon_b^*\right) = H\hat{f}(x) + H\varepsilon_b^*$$

$$= H^2 y + H\varepsilon_b^* \quad \overset{H^2=H}{=} \quad Hy + H\varepsilon_b^*$$

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{f}(x_b^*) = \frac{1}{B}\sum_{b=1}^{B} Hy + \frac{1}{B}\sum_{b=1}^{B} H\varepsilon_b^*$$

$$= Hy + H\left(\frac{1}{B}\sum_{b=1}^{B} \varepsilon_b^*\right)$$

$B\to\infty$ → $\hat{f}(x)$

$\left(\frac{1}{B}\sum_{b=1}^{B} \varepsilon_b^*\right) \to 0$ for $B\to\infty$ for the LLN

Exercise 15.1

If X_b are identically distributed (~~NOT~~ necessarily independent) with positive correlation ρ

then $\mathrm{Var}\left(\frac{1}{B}\sum_{b=1}^{B} X_b\right) = \rho\sigma^2 + \frac{1-\rho}{B}\sigma^2$

where σ^2 is the variance of X_b, μ is the mean of X_b

- why this formula fails with $\rho < 0$

In general $\mathrm{Var}[A] = \underline{E[A^2]} - (E[A])^2 \qquad A = \frac{1}{B}\sum_{b=1}^{B} X_b$

$$E[A^2] = E\left[\left(\frac{1}{B}\sum_{b=1}^{B} X_b\right)^2\right] = \frac{1}{B^2} E\left[\sum_{b=1}^{B} X_b^2 + 2\sum_{b=1}^{B}\sum_{c>b} X_b X_c\right]$$

$$= \frac{1}{B^2} E\left[\sum_{b=1}^{B} X_b^2\right] + \frac{2}{B^2} E\left[\sum_{b=1}^{B}\sum_{c>b} X_b X_c\right]$$

$$= \frac{B}{B^2} E[X_b^2] + \frac{2}{B^2}\binom{B}{2} E[X_b X_c]$$

$$= \frac{1}{B}\left(\sigma^2 + \mu^2\right) + \frac{2}{B^2}\frac{B(B-1)}{2} E[X_b X_c]$$

to compute the last term, we start from the definition of correlation

$$\rho = \frac{E[(X_b - \mu)(X_c - \mu)]}{\sqrt{\sigma^2}\sqrt{\sigma^2}} = \frac{E[X_b X_c] - \mu^2}{\sigma^2}$$

therefore $\sigma^2\rho = E[X_b X_c] - \mu^2 \iff E[X_b X_c] = \sigma^2\rho + \mu^2$

$$= \frac{1}{B}\left(\sigma^2 + \mu^2\right) + \frac{(B-1)}{B}\left(\sigma^2\rho + \mu^2\right)$$

The other term

$$\left(E[A]\right)^2 = \left(E\left[\frac{1}{B}\sum_{b=1}^{B} X_b\right]\right)^2 = \left(\frac{1}{B} B\mu\right)^2 = \mu^2$$

Therefore

$$\mathrm{Var}\left(\frac{1}{B}\sum_{b=1}^{B} X_b\right) = \frac{1}{B}\left(\sigma^2 + \mu^2\right) + \frac{B-1}{B}\left(\sigma^2\rho + \mu^2\right) - \frac{B}{B}\mu^2 \qquad (15.1)$$

$$= \frac{\sigma^2 + \mu^2 + (B-1)\sigma^2\rho + (B-1)\mu^2 - B\mu^2}{B} = \frac{\sigma^2 + B\sigma^2\rho - \sigma^2\rho}{B} = \rho\sigma^2 + \frac{(1-\rho)\sigma^2}{B}$$

Since $\mathrm{Var}\left(\frac{1}{B}\sum_{b=1}^{B} X_b\right) \geq 0$, it means that $\rho\sigma^2 + \frac{(1-\rho)\sigma^2}{B} \geq 0$

$\sigma^2 \geq 0$, so $\rho + \frac{1-\rho}{B} \geq 0 \iff B\rho + 1 - \rho \geq 0$

$\iff (B-1)\rho + 1 \geq 0$

$\iff \rho \geq -\frac{1}{B-1} \xrightarrow{B\to\infty} 0 \Rightarrow \rho$ must be non negative

Exercise 15.4

X_i, $i=1,\dots,N$ are iid with mean μ and variance σ^2

$\bar{X}_1^* = \frac{1}{N}\sum_{i=1}^{N} x_{i1}^*$, where $x_1^* = (x_{11},\dots,x_{N1})$ is a bootstrap sample

$\bar{X}_2^* = \frac{1}{N}\sum_{j=1}^{N} x_{j2}$, where $\bar{x}_2^* = (x_{12},\dots,x_{N2})$ is another bootstrap sample

- show that $corr(\bar{X}_1^*, \bar{X}_2^*) = \frac{N}{2N-1} \approx 50\%$
- derive $var(\bar{X}_1^*)$
- derive $var(\bar{X}_{bag})$

→ we start with $Var(\bar{X}_1^*)$

$$Var(\bar{X}_1^*) = E[\bar{X}_1^{*2}] - (E[\bar{X}_1^*])^2 \qquad E[\bar{X}] = E[X] = \mu$$

$$= E\left[\left(\frac{1}{N}\sum_{i=1}^{N} \bar{X}_{i1}^*\right)^2\right] - \mu^2$$

from Ex 15.1

$$= \frac{1}{N}(\sigma^2+\mu^2) + \frac{N-1}{N} E[X_{i1}X_{j1}] - \mu^2$$

Here $E[X_{i1}X_{j1}] = E[X_{i1}^2]\Pr(X_{i1}=X_{j1}) + E[X_{i1}]E[X_{j1}]\Pr(X_{i1}\neq X_{j1})$

$$= (\sigma^2+\mu^2)\frac{1}{N} + \mu^2\frac{N-1}{N}$$

$$= \frac{\sigma^2 + \mu^2 + N\mu^2 - \mu^2}{N} = \frac{\sigma^2}{N} + \mu^2$$

$$= \frac{1}{N}(\sigma^2+\mu^2) + \frac{N-1}{N}\left(\frac{\sigma^2}{N}+\mu^2\right) - \mu^2$$

$$= \frac{\sigma^2}{N} + \frac{\mu^2}{N} + \frac{\sigma^2}{N} - \frac{\sigma^2}{N^2} + \mu^2 - \frac{\mu^2}{N} - \mu^2$$

$$= \frac{2N-1}{N^2}\sigma^2 = Var(\bar{X}_1^*) = Var(\bar{X}_2^*)$$

$$\rightarrow Corr\left(\bar{x}_1^*, \bar{x}_2^*\right) = \frac{Cov\left(\bar{x}_1^*, \bar{x}_2^*\right)}{\sqrt{Var\left(\bar{x}_1^*\right)} \cdot \sqrt{Var\left(\bar{x}_2^*\right)}}$$

from point above

$$Cov\left(\bar{x}_1^*, \bar{x}_2^*\right) = E\left[\frac{1}{N}\sum_{i=1}^{N} x_{i1} \cdot \frac{1}{N}\sum_{j=1}^{N} x_{j2}\right] - \mu^2$$

$$= \frac{1}{N^2} E\left[\sum_{i=1}^{N} x_{i1} \sum_{j=1}^{N} x_{j2}\right] - \mu^2$$

$$= \frac{1}{N^2} \sum_{i=1}^{N}\sum_{j=1}^{N} E\left[x_{i1} x_{j2}\right] - \mu^2$$

$$= \frac{N^2}{N^2}\left(\frac{\sigma^2}{N} + \mu^2\right) - \mu^2 = \frac{\sigma^2}{N}$$

$$= \frac{\sigma^2}{N} \cdot \frac{N^2}{(2N-1)\sigma^2} = \frac{N}{2N-1} \approx 50\%$$

$$\rightarrow Var\left(\bar{x}_{bag}\right) = Var\left(\frac{1}{B}\sum_{b=1}^{B} \bar{x}_b\right)$$

$$= \frac{1}{B^2}\left(\sum_{b=1}^{B} Var\left(\bar{x}_b\right) + \sum_{i \neq j} Cov\left(\bar{x}_b \bar{x}_c\right)\right)$$

$$= \frac{1}{B^2}\left(B\,\frac{2N-1}{N^2}\sigma^2 + B(B-1)\frac{\sigma^2}{N}\right)$$

$$= \frac{2N-1+NB-N}{BN^2}\sigma^2$$

$$= \frac{BN+N-1}{BN^2}\sigma^2 = \frac{\sigma^2}{N} + \frac{N-1}{BN^2}\sigma^2$$

$\rightarrow 0$ as $B \rightarrow \infty$

Some variance of $\bar{x}$

Exercise 15.5

Show that the sampling correlation between two trees (random forest) of a point x is

$$\rho = \frac{Var_Z\left(E_{\Theta|Z}\left[T(x;\Theta(Z))\right]\right)}{Var_Z\left(E_{\Theta|Z}\left[T(x;\Theta(Z))\right]\right) + E_Z\left[Var_{\Theta|Z}\left(T(x;\Theta(Z))\right)\right]}$$

where

$T(x;\Theta(Z))$ is a random forest tree

$\frac{1}{B}\sum_{b=1}^{B} T(x_b^*;\Theta_b(Z))$ is the random forest

x_b^* is the b-th bootstrap sample

Θ_b characterises the b-th tree (split variables, cutpoints, values at the terminal nodes)

Z is the training dataset (sample from a population)

$$\rho = \frac{Cov_{\Theta,Z}\left(T_1(x;\Theta(Z)),\ T_2(x;\Theta(Z))\right)}{\sqrt{Var_{\Theta,Z}\left(T_1(x;\Theta(Z))\right)}\sqrt{Var_{\Theta,Z}\left(T_2(x;\Theta(Z))\right)}} = \frac{Cov_{\Theta,Z}\left(T_1, T_2\right)}{Var_{\Theta,Z}\left(T_1\right)}$$

$$\begin{aligned}
Cov_{\Theta,Z}(T_1,T_2) &= E_{\Theta,Z}[T_1T_2] - E_{\Theta,Z}[T_1]E_{\Theta,Z}[T_2] \\
&= E_Z\left[E_{\Theta|Z}[T_1T_2]\right] - E_Z\left[E_{\Theta|Z}[T_1]E_{\Theta|Z}[T_2]\right] \\
&= E_Z\left[E_{\Theta|Z}[T^2]\right] - E_Z\left[E_{\Theta|Z}[T_1]\right]E_Z\left[E_{\Theta|Z}[T_2]\right] \\
&= E_Z\left[E_{\Theta|Z}[T^2]\right] - \left(E_Z\left[E_{\Theta|Z}[T]\right]\right)^2 = Var_Z\left(E_{\Theta|Z}[T]\right)
\end{aligned}$$

$$\begin{aligned}
Var_{\Theta,Z}[T] &= E_{\Theta,Z}[T^2] - \left(E_{\Theta,Z}[T]\right)^2 \\
&= E_Z\left[E_{\Theta|Z}[T^2]\right] - \left(E_Z\left[E_{\Theta|Z}[T]\right]\right)^2 \\
&= E_Z\left[Var_{\Theta|Z}[T] + \left(E_{\Theta|Z}[T]\right)^2\right] - \left(E_Z\left[E_{\Theta|Z}[T]\right]\right)^2 \\
&= E_Z\left[Var_{\Theta|Z}[T]\right] + \underbrace{E_Z\left[E_{\Theta|Z}[T]^2\right] - \left(E_Z\left[E_{\Theta|Z}[T]\right]\right)^2}_{Var_Z\left[E_{\Theta|Z}[T]\right]} \\
&= E_Z\left[Var_{\Theta|Z}[T]\right] + Var_Z\left[E_{\Theta|Z}[T]\right]
\end{aligned}$$

$$\rho = \frac{Var_Z\left(E_{\Theta|Z}[T]\right)}{Var_Z\left(E_{\Theta|Z}[T]\right) + E_Z\left[Var_{\Theta|Z}[T]\right]} \qquad E_{\Theta|Z}[T] = \hat{f}_{RF}(x)$$

$$= \frac{Var_Z\left(\hat{f}_{RF}(x)\right)}{Var_Z\left(\hat{f}_{RF}(x)\right) + E_Z\left[Var_{\Theta|Z}[T]\right]}$$

(2) within-Z variance increases with m decreasing

Consider m the number of variables used at each node to split

↳ variance of a RF with Z stochastic decreases as m decreases