- (Portfolio example): We have two assets with returns X and Y. We want to allocate our money so that we minimize variance

$$Var(\alpha X + (1-\alpha) Y) \quad . \text{ Solution: } \alpha = \frac{Var(Y) - Cov(X,Y)}{Var(X) + Var(Y) - 2Cov(X,Y)}$$

(from R code): res.boot\$t0 = $\hat{\alpha}$ (estimate of original data)
res.boot\$t = $\hat{\alpha}^{*(1)}, \ldots, \hat{\alpha}^{*(R)}$

- Bootstrapping consistency. The bootstrap is consistent for $\hat{\theta}_n$ if, for an increasing sequence a_n (often $a_n = \sqrt{n}$) and for all x,

$$\underbrace{P(a_n(\hat{\theta}_n - \theta_0) \le x)}_{\text{true rescaled distribution centered @ } \theta_0} - P^*(\underbrace{a_n(\hat{\theta}_n^* - \hat{\theta}_n)}_{\text{bootstrap rescaled distribution centered @ } \hat{\theta}_n} \le x) \xrightarrow{P} 0 \text{ as } n \to \infty$$

P^*: Prob. wrt. bootstrap distribution conditional on the data (P^* is a r.v.).

- We then have (i) $\dfrac{Var^*(\hat{\theta}_n^*)}{Var(\hat{\theta}_n)} \xrightarrow{P} 1$

and (ii) $\dfrac{E^*(\hat{\theta}_n^*) - \hat{\theta}_n}{E(\hat{\theta}_n) - \theta} \xrightarrow{P} 1$

We are interested in these!
① Variance $Var(\hat{\theta}_n)$ and
② Bias $E(\hat{\theta}_n) - \theta_0$ (of $\hat{\theta}_n$). the estimated θ.

We can estimate the variance① and bias② by the corresponding numerators (ie those of the bootstrapped data).

- Bias We estimate the bias of $\hat{\theta}_n$ ② by.

(② $Bias(\hat{\theta}_n) = E(\hat{\theta}_n) - \theta_0$) $\rightsquigarrow E^*(\hat{\theta}_n^*) - \hat{\theta}_n$ where

relationship is equality when $B = \infty$

$$E^*(\hat{\theta}_n^*) = \int \hat{\theta}_n^* dP^* \approx \frac{1}{B} \sum_{i=1}^{n} \hat{\theta}_n^{*(i)} = \overline{\theta}_n^{*n}$$

- Variance We estimate the variance by

(① $Var(\hat{\theta}_n)$) $\rightsquigarrow Var^*(\hat{\theta}_n^*) = \int (\hat{\theta}_n^* - E^*(\hat{\theta}_n))^2 dP^*$, so

$$= \frac{1}{B-1} \sum_{i=1}^{n} (\hat{\theta}_n^{*(i)} - \overline{\theta}_n^*)^2 .$$

- The sampling distribution $\hat{\theta}_n = \text{median}(Z_1, \ldots, Z_n)$ is approximately

$$\mathcal{N}\Big(\underset{\uparrow \text{ true median}}{\theta}, \underbrace{\frac{1}{4\sigma^2(\theta) n}}_{\text{density}}\Big).$$

 Computational Statistics Section 23 March 18

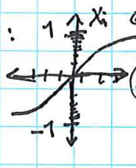
1) $X : \Omega \to \mathbb{R}^p$ $\quad$ (X, Y) is a vector of r.v.s X and Y, so
$Y : \Omega \to \mathbb{R}$
$\varepsilon : \Omega \to \mathbb{R}$ $\quad$ $(X, Y) : \Omega \to \mathbb{R}^p \times \mathbb{R}$ and

$Y = f(X) + \varepsilon$. $\quad$ • $(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)}) \in \mathbb{R}^p \times \mathbb{R}$

We want to estimate f as $\hat{f} : \mathbb{R}^p \to \mathbb{R}$.

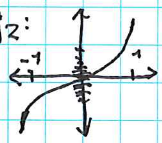
(a) $X = (x_1, \dots, x_p)$ where $x_i = g(z_i)$ and g_i is a distribution.

ex. g_1 :

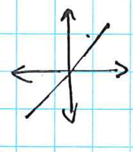


concentrated around 1 and -1, (z_i uniform) where g is steep.

g_2 :



g_3 :



uniform.

(c) $f : x \to f_{1dim}(x_1)$, i.e. only first dimension matters

$$\mathbb{E}\left[\left(Y^* - \hat{f}_{(X^{(1)}, Y^{(1)}), \dots, (X^{(n)}, Y^{(n)})}(X^*)\right)^2\right]$$

$$= \mathbb{E}\Big[\underbrace{\mathbb{E}\left[\left(Y^* - \hat{f}_{(X^{(1)}, Y^{(1)}), \dots, (X^{(n)}, Y^{(n)})}(X^*)\right)^2 \,\middle|\, (X^{(1)}, Y^{(1)}), \dots, (X^{(n)}, Y^{(n)})\right]}\Big]$$

$$= \mathbb{E}[(*)]$$ so, by law of large numbers,

$$(*) \approx \frac{1}{1000} \sum_{m=1}^{1000} \frac{1}{2000} \sum_{l=1}^{2000} \left(Y_l^* - \hat{f}_{(X_m^{(1)}, Y_m^{(1)}), \dots, (X_m^{(n)}, Y_m^{(n)})}(X_l^*)\right)^2$$

? ($\forall m$, we simulate and obtain X^* data to train on, ($|X^*| = 2000$) and avg over all 1000 bootstraps)