Resampling

MC & Bootstrap → Data when CLT doesn't apply

↳ Statistical model

↓

unbiased estimator when $E[\hat{\theta}] = \theta$

↳ so when bias $b(\hat{\theta}) = E[\hat{\theta}] - \theta$

Importance Sampling

- Choosing an adequate sampling dist^n in order to approximate the EV with good control on the estimation variance

MC Q 18-19 Q3

$Y_i = \theta^* x_i + Z_i$ $\quad Z_i \sim \chi^2_1$

what could have a negative impact on the estimation of θ^*

- noise is heavy tailed & skewed positively
  ↳ could make observations larger than the true data.
  ↳ not an issue for slope but more outliers make skew the distribution

2021 Winter Q2

A data analyst is implementing a Monte Carlo simulation of $M = 1,000$ random samples of realisations of the model

$$Y_i = \theta^* X_i + Z_i, \qquad i = 1, \ldots, n \tag{1}$$

with $n = 100$, $\theta^* = 8$ and a sequence of i.i.d. realizations $Z_i \overset{iid}{\sim} t_d$ with $d = 3$ degrees of freedom, using a single sample $\{X_i\}_{i=1}^n$ from $X \sim \mathcal{U}(1, 2)$ to generate all M Monte Carlo samples, and computes and stores the Monte Carlo least squares estimates of θ^* for analysis. Note the analyst is making sure to not include an intercept in the regression model when fitting it to the simulated data.

(a) Quote the theoretic expected value of Y, i.e. the true value of $E(Y)$, showing your calculation.

(b) Quote the theoretic expected value of $\hat{\theta}$, i.e. the true value of $E(\hat{\theta})$, justifying your answer with a brief statement.

$$\begin{aligned} E[Y] &= E[\theta X_i] + E[Z_i] \qquad Z_i \text{ \& } X_i \text{ independent} \\ &= \theta E[X_i] + E[Z_i] \\ &= 8 \cdot (1.5) + 0 \quad = 12 \end{aligned}$$

Mean of a t-dist^n $= 0$ $\forall$ df > 1

$E[\theta]$

$\rightarrow$ 1000 samples

$\rightarrow$ we find the mean at each sample where we have 100 observations.

$$\hat{\theta} = OLS = \frac{\sum X_i Y_i}{\sum X_i^2}$$

Ans

- The OLS being consistent under this model $E[\theta^*] = 8$
- The MC simulation demonstrates this consistency.

c) How would MC estimate differ if $Z \sim N(0,1)$

t dist^n variance $= \frac{df}{df-2}$

So good at modelling low population samples

- So if $Z \sim N(0,1)$, the mean would be closer to the true value.

- There would be less variance in the model.

Actual Answer:

- Mean should remain the same since the OLS is still unbiased
- The variance would decrease, since there is less outliers in our data

d) - Bias decrease $E(\hat{\theta}_M) \xrightarrow[M\to\infty]{LLN} 0$

- Higher proportion of outliers
  ↳ more symmetric distribution as M increases.

Confidence Intervals for BS

→ naive: just take quantiles of the estimated value

→ Appropriate: 2(true model parameter) − quantile of the estimated value.

MC estimation of Standard normal CDF

M → Sample size of each estimate.

G → number of grid pts where the CDF is evaluated

x → values where CDF is evaluated.

CDF → store the value at each x

g → Fnc to be evaluated $\left(e^{-\frac{x^2}{2}}\right)$
↳ take the mean of it

u →