Ex 1.1

LHS RHS

Prove $|corr(X,Z)| = 1 \iff \exists\, a, b \in \mathbb{R}, b \neq 0$ s.t. $Z = a + bX$

considering $g(t) = E\left[\left(t(X - E(X)) - (Z - E(Z))\right)^2\right]$

$$= E\left[t^2 (X - E(X))^2 + (Z - E(Z))^2 - 2t(X - E(X))(Z - E(Z))\right]$$

using linearity of E

$$\Rightarrow g(t) = E\left(t^2 (X - E(X))^2\right) + E\left((Z - E(Z))^2\right) - E\left(2t(X - E(X))(Z - E(Z))\right)$$

$$= t^2 E\left((X - E(X))^2\right) + E\left((Z - E(Z))^2\right) - 2t\, E\left((X - E(X))(Z - E(Z))\right)$$

Now $var(X) = E(X - E(X))^2$, $cov(X,Z) = E((X - E(X))(Z - E(Z)))$

$$\Rightarrow g(t) = t^2 var(X) + var(Z) - 2t\, cov(X,Z)$$

this is a polynomial of the form $at^2 + bt + c = 0$

$$\text{roots } t_1, t_2 = \frac{2\, cov(X,Z) \pm \sqrt{4\, cov(X,Z)^2 - 4\, var(X)\, var(Z)}}{2\, var(X)}$$

$$= \frac{cov(X,Z) \pm \sqrt{cov(X,Z)^2 - var(X)\, var(Z)}}{var(X)} \quad (1)$$

Now $corr(X,Z) = \dfrac{cov(X,Z)}{\sigma(X)\,\sigma(Z)} = \dfrac{cov(X,Z)}{\sqrt{var(X)}\sqrt{var(Z)}}$

$$\Rightarrow \frac{cov(X,Z)^2}{corr(X,Z)^2} = var(X)\, var(Z)$$

and assuming the LHS of the problem statement to be true

$$\Rightarrow \frac{cov(X,Z)^2}{corr(X,Z)^2} = \frac{cov(X,Z)^2}{\underbrace{|corr(X,Z)|^2}_{1}} = cov(X,Z)^2$$

$$\Rightarrow cov(X,Z)^2 = var(X)\, var(Z)$$

plugging into (1)

$$\Rightarrow \text{roots } t_1, t_2 = \frac{cov(X,Z)}{var(X)} + \cancel{\frac{\sqrt{0}}{var(X)}}^{\,0}$$

$$\cancel{\mathrm{corr}(X,Z) = \frac{\mathrm{cov}(X,Z)}{\sigma(X)\,\sigma(Z)} = \frac{\mathrm{cov}(X,Z)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Z)}}}$$

$$\cancel{\frac{\mathrm{cov}(X,Z)^2}{\mathrm{corr}(X,Z)^2} = 2\,\mathrm{var}(X)\,\mathrm{var}(Z)}$$

$$\Rightarrow \left(t - \frac{\mathrm{cov}(X,Z)}{\mathrm{var}(X)}\right)\left(t - \frac{\mathrm{cov}(X,Z)}{\mathrm{var}(X)}\right) = 0$$

$$\Rightarrow \quad t - \frac{\mathrm{cov}(X,Z)}{\mathrm{var}(X)} = 0$$

$$\Rightarrow \quad \mathrm{cov}(X,Z) = t\,\mathrm{var}(X) \quad (2)$$

again using LHS of the problem statement

$$\left|\frac{\mathrm{cov}(X,Z)}{\sqrt{\mathrm{var}(X)}\sqrt{\mathrm{var}(Z)}}\right|^2 = 1$$

$$\Rightarrow \mathrm{cov}(X,Z)^2 = \mathrm{var}(X)\,\mathrm{var}(Z)$$

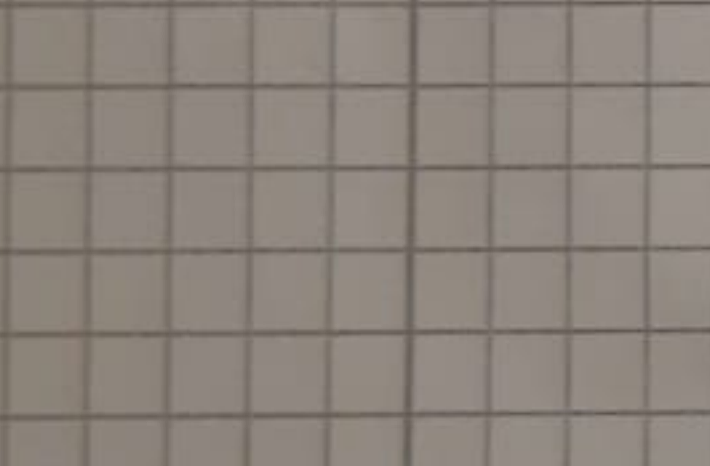
$$\Rightarrow (t\,\mathrm{var}(X))^2 = \mathrm{var}(X)\,\mathrm{var}(Z)$$

(using (2))

$$\Rightarrow t^2\,\mathrm{var}(X) = \mathrm{var}(Z)$$

$$\Rightarrow \mathrm{var}(tX) = \mathrm{var}(Z)$$

Now if the 2 distributions tX and Z have the same variance, their mean can be made same by adding another factor 'b' to move the distribution tX left or right. Also adding this constant factor won't change the variance



$$\text{so, } \mathrm{var}(tX + b) = \mathrm{var}(Z)$$

$$\Rightarrow tX + b = Z \quad \text{for some } t, b$$

$$\Rightarrow aX + b = Z \quad \text{for some } b \neq 0 \text{ and } a = t$$

ex 1.2

$T_1 = \bar{X}_N$

$\mathbb{E}(\bar{X}_N) = \frac{1}{N} \cdot \sum_{j=1}^{N} \mathbb{E}(X_j) = \frac{1}{N} \cdot N\mu = \mu \Rightarrow \text{bias}(\bar{X}_N) = 0$

$\text{mse}(\bar{X}_N) \overset{1.2.2.}{=} \text{var}(\bar{X}_N) \underset{\uparrow \text{Bienayme}}{=} \frac{1}{N^2} \cdot \sum_{j=1}^{N} \text{var}(X_j) = \frac{1}{N} \cdot \sigma^2$

$T_2 = X_1$

$\mathbb{E}(X_1) = \mu \Rightarrow \text{bias}(T_2) = 0$

$\text{mse}(T_2) \overset{1.2.2}{=} \text{var}(X_1) = \sigma^2$

$T_3 = \frac{1}{N+1} \sum_{i=1}^{N} X_i$

$\text{bias}(T_3) = \mathbb{E}(T_3) - \mu = \frac{1}{N+1} \cdot \left(\sum_{i=1}^{N} \mathbb{E}(X_i)\right) - \mu$

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

$\text{mse}(T_3) = \text{var}(T_3) + \text{bias}(T_3)$

$\underset{\uparrow \text{Bienayme}}{=} \frac{1}{(N+1)^2} \cdot \sum_{i=1}^{N} \text{var}(X_i) + \left(\frac{N}{N+1} - 1\right)\mu$

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

If we check the estimators for consistency T_2 fails which makes it my least prefered choice.
T_1 and T_3 are both consistent but T_1 is also unbiased which makes it the best choice in most cases. If we have a data set it is possible that we have measurement errors. If we suspect that these errors are not symmetrical distributed but tend to ~~overshoot the a~~ be a little bit higher we could use estimator T_3 to correct these measurement errors.

Ex 1.3

To prove, $mse(\hat{\sigma}_N^2) < mse(\hat{S}_N^2) \quad \forall N \geq 2$

Expanding RHS using Lemma 1.2.2

$$mse_\theta(\hat{S}_N^2) = var_\theta(\hat{S}_N^2) + (\cancel{bias_\theta(\hat{S}_N^2)}^{\,0})^2$$

from Example 1.2.3 we know $bias_\theta(\hat{S}_N^2) = 0$

from Thm 1.2.5, $\mathcal{L}\left(\frac{(N-1)\hat{S}_N^2}{\sigma^2}\right) = \chi^2_{N-1}$

and $var(X) = 2k$

$$\Rightarrow var\left(\frac{(N-1)\hat{S}_N^2}{\sigma^2}\right) = 2(N-1)$$

$$\Rightarrow \left(\frac{N-1}{\sigma^2}\right)^2 var(\hat{S}_N^2) = 2(N-1)$$

$$\Rightarrow var(\hat{S}_N^2) = \frac{\sigma^4}{(N-1)^2} \cdot 2(N-1) = \frac{2\sigma^4}{N-1}$$

$$\Rightarrow RHS = \frac{2\sigma^4}{N-1}$$

Expanding LHS using Lemma 1.2.2

$$mse_\theta(\hat{\sigma}_N^2) = var_\theta(\hat{\sigma}_N^2) + (bias_\theta(\hat{\sigma}_N^2))^2$$

from Example 1.2.3 we know $bias_\theta(\hat{\sigma}_N^2) = -\frac{\sigma^2}{N}$

$$\Rightarrow LHS = var_\theta(\hat{\sigma}_N^2) + \frac{\sigma^4}{N^2}$$

we know $\hat{\sigma}_N^2 = \frac{1}{N}\sum_{j=1}^{N}(z_j - \bar{z}_N)^2$ & $\hat{S}_N^2 = \frac{1}{N-1}\sum_{j=1}^{N}(z_j - \bar{z}_N)^2$

$$\Rightarrow \hat{S}_N^2 = \frac{N}{N-1} \cdot \hat{\sigma}_N^2$$

$$\Rightarrow \mathcal{L}\left(\frac{(N-1)}{\sigma^2} \cdot \hat{S}_N^2\right) = \chi^2_{N-1} \Rightarrow \mathcal{L}\left(\frac{\cancel{N-1}}{\sigma^2} \cdot \frac{N}{\cancel{N-1}} \cdot \hat{\sigma}_N^2\right) = \chi^2_{N-1}$$

$$\Rightarrow \mathcal{L}\left(\frac{N}{\sigma^2} \cdot \hat{\sigma}_N^2\right) = \chi^2_{N-1}$$

$$\Rightarrow var\left(\frac{N}{\sigma^2} \cdot \hat{\sigma}_N^2\right) = 2(N-1)$$

$$\Rightarrow \left(\frac{N^2}{\sigma^4}\right) var(\hat{\sigma}_N^2) = 2(N-1)$$

$$\Rightarrow var(\hat{\sigma}_N^2) = \frac{\sigma^4}{N^2} \cdot 2(N-1)$$

$$\Rightarrow LHS = \frac{\sigma^4}{N^2} \cdot 2(N-1) + \frac{\sigma^4}{N^2}$$

Now we need to prove $LHS < RHS \quad \forall \; N \geq 2$

i.e. $\frac{\sigma^4}{N^2} \cdot 2(N-1) + \frac{\sigma^4}{N^2} < \frac{2\sigma^4}{N-1}$

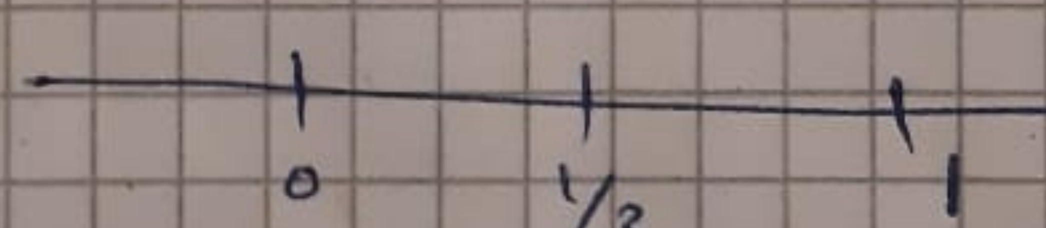
i.e. $\frac{2N-1}{N^2} < \frac{2}{N-1}$

we have the following inflection pts for this inequality

a) $N=0$ (LHS is undefined)

b) $N=1$ (RHS is undefined)

c) $N=1/3$ (LHS = RHS)



plugging in values of N in a), b), c)
we find that the inequality holds true for

$N<0$

$0<N<1/3$

$N>1$

Since N is sample size and must be a positive non fraction we only consider $N>1$

Hence the inequality holds true for $N \geq 2$

$\Rightarrow LHS < RHS \quad \forall \; N \geq 2$