AM- GM inequality

summarizes the sample with single representative which is centre of data. (measure of location)

$\underset{\sim}{X} = (X_1 \quad X_2 \quad --- \quad X_n)$ $\qquad$ $Y = (f(X_1) \quad f(X_2) \; --- \; f(X_n))$

$\sum_{i=1}^{n} L\,(X_i, \theta)$

$\downarrow$ $\qquad\qquad$ $\downarrow$

$\hat{\theta}\,(X_i)$ $\qquad\qquad$ $\hat{\theta}(Y)$

estimator which we get by minim loss fn.

we ~~exa~~ hope

$\hat{\theta}(Y) = f(\,\hat{\theta}(X))$.

Summary statistics that have this property (for monotone transformation) are known as transformation - invariants (or equivariant).

Arithematic Mean → Generally not ~~invo~~ ~~eq~~ equivariants.

Median → ~~is~~ equivariant.

✓Excercise : For affine transformations of the type ~~f~~

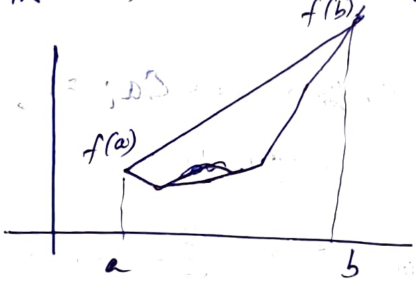
$f(x) = a + bx \quad , \; b \neq 0$, the AM is invariant.

Although AMs is not transformation equivariant, we can say somethings useful for convex and concave transformations.

$f: \mathbb{R} \to \mathbb{R}$ $\qquad$ f is convex, if $tf(a) + (1-t)f(b) > \cancel{ta + (1-t)b}$ $f(ta + (1-t)b)$

(parametric form or linear combination)

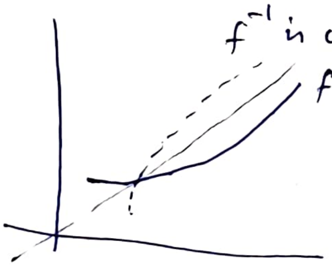
(Power function is concave convex trans)



for any $\lambda_1, \lambda_2 \in (0,1)$ s.t. $\lambda_1 + \lambda_2 = 1$.

$f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2)$.

$\forall \; x_1, x_2 \in dom(f)$.



<u>Jensen's Inequality</u> : For a real convex function ϕ, numbers $x_1, x_2, x_3, \ldots, x_n$ in its domain, and positive weights ~~$\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_n$~~ $a_1, a_2, \ldots, a_n$. such that $\sum_{i=1}^{n} a_i = 1$, we have

$$\phi\left(\sum_{i=1}^{n} a_i x_i\right) \leq \sum a_i \phi(x_i)$$

Similarly, $\phi(\sum a_i x_i) \geq \sum a_i \phi(x_i)$

Let's prove by induction for ϕ convex,

$n=2$; We want to show: for $a_1, a_2 \geq 0$ s.t. $a_1 + a_2 = 1$.

$$\cancel{\phi\left(\sum_{i=1}^{2} a_i x_i\right) \leq \sum_{i=1}^{2}}$$

$$\phi(a_1 x_1 + a_2 x_2) \leq a_1 \phi(x_1) + a_2 \phi(x_2).$$

holds by definition of convexity.

say for $n = m$: this is also true;

for any $a_1, a_2, \ldots, a_n \geq 0$ s.t. $\sum a_i = 1$,

$$\phi(\sum a_i x_i) \leq \sum a_i \phi(x_i)$$

For $n = m+1$: Given, $a_1, a_2, \ldots, a_{m+1} \geq 0$, $\sum_{i=1}^{m+1} a_i = 1$,

To show, $\phi\left(\sum_{i=1}^{m+1} a_i x_i\right) \leq \sum_{i=1}^{m+1} a_i \phi(x_i)$

$$\phi\left(\sum_{i=1}^{m+1} a_i x_i\right) = \phi\left(a_{m+1} x_{m+1} + (1-a_{m+1})\left(\frac{\sum_i^m a_i x_i}{1-a_{m+1}}\right)\right)$$

$$\leq a_{m+1}\,\phi(x_{m+1}) + (1-a_{m+1})\,\phi\left(\sum_{i=1}^{m} \lambda_i x_i\right)$$

$$\leq \cdots + \cancel{(1-a_{m+1})} \sum_{i=1}^{m} \frac{a_i}{\cancel{(1-a_{m+1})}}\phi(x_i)$$

$\boxed{\text{let } \lambda_i = \frac{a_i}{1-a_{m+1}}, \quad i = 1, 2, \ldots, m}$

This inequality allows us to compare any power transformation.

Exercise: Show that for a vector of positive observations, the GM is larger (or ~~equality~~ equal) to the HM.

Q. Can we extend location ~~or~~ estimators or location statistics to higher dimensions?

We can be interested in bivariate data, say $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$.

$(X_i, Y_i) \rightarrow$ measurement from the same observational unit.

Loss function: $L\left(\begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix}\right)$

$\downarrow$ vectors in data in XY plane

$\downarrow$ location in X-Y plane

$$d\left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}\right) = \sqrt{(X_1-Y_1)^2 + (X_2-Y_2)^2}$$

$$= (X-\theta_X)^2 + (Y-\theta_Y)^2$$

We want to minimize: $\lambda\begin{pmatrix} \theta_X \\ \theta_Y \end{pmatrix} = \sum_{i=1}^{n}(x_i - \theta_x)^2 + \sum_{i=1}^{n}(Y_i - \theta_y)^2$

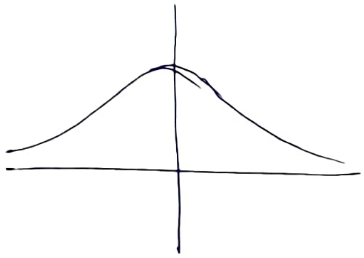
$\therefore \hat{\theta}_X = \bar{X}_i, \hat{\theta}_y = \bar{Y}_i \qquad \begin{pmatrix} \hat{\theta}_X \\ \hat{\theta}_Y \end{pmatrix} = \begin{pmatrix} \bar{X}_i \\ \bar{Y}_i \end{pmatrix}$

If we have absolute errors:

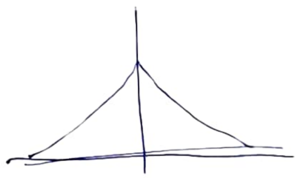
$$L\left(\begin{pmatrix} X \\ Y \end{pmatrix}, \begin{pmatrix} \theta_x \\ \theta_y \end{pmatrix}\right) = \sqrt{(X-\theta_x)^2 + (Y-\theta_y)^2}$$

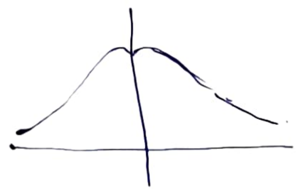
$$\lambda(\theta_x, \theta_y) = \sum_{i=1}^{n} \sqrt{(X_i-\theta_x)^2 + (Y_i-\theta_y)^2}$$ → not always differentiable.

↳ needs to be computed numerically.



We are ~~now~~ from now on think of statistics are estimators.





~~If~~ As we get more and more data ~~we can~~ our estimator will get closer to population ~~θ~~ estimator.

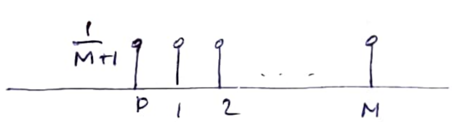
Median is more robust than mean but, not that efficient.

Location parameters → mean, median.

Scaling parameters } For now we consider these two.

→ higher the scale parameter higher the spread.

Model 1:



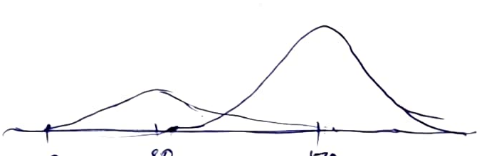
$Y \sim \text{Disc.Unif}(\{\frac{0}{M}, \frac{1}{M}, \frac{2}{M}, \dots, \frac{M}{M}\})$.

pt. of symmetry is 0.5.

There is some parameter θ which we want to estimate.

$$\text{mean}(\underset{\text{Distribution.}}{T}(X_1, X_2, \dots, X_n) - \theta)^2 = MSE(T).$$

rdist <- function(n) 5 + 3*rt(n, df = 5)



$170 \times \frac{99}{100} + 80 \times \frac{1}{100}$ = population mean.