MATH 710, INSTRUCTOR: JACOB KOGAN

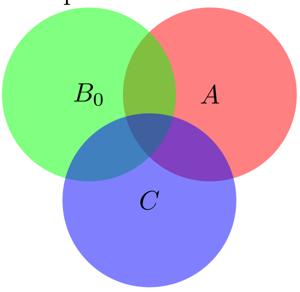
XIAOWEI SONG

November 22, 2024

R embeding test:

```
1+1
## [1] 2
rnorm(5)
## [1]  0.8267035 -0.4852683  1.1466972 -0.3953367  0.3739446
```

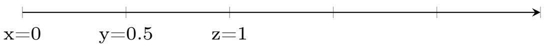
Tikz plot test:



1. Sep 3, 2015.

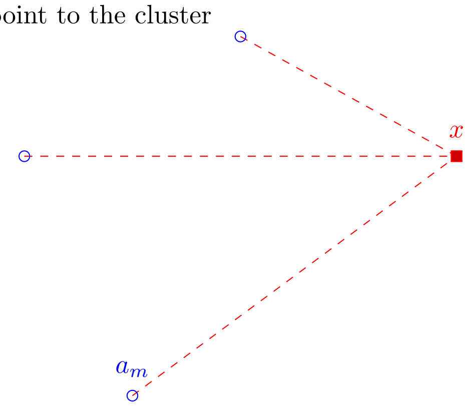
1.1. **Distance.** need to be: $\begin{cases} d(x,y) & = d(y,x) \\ d(x,y) & \geq 0,\ d(x,y) = 0\, iff\, x = y \\ d(x,z) & \leq d(x,y) + d(y,z)\ \text{(Triangular inequality)} \end{cases}$

for $a_1, a_2, \cdots, a_m \in R^n$,$d(x,y) = |x-y|^2$, where the Norm-2 not true for Triangular rule
For example: $x = 0, y = 0.5, z = 1$



$d(x,y) = 0.25, d(y,z) = 0.25, d(x,z) = 1$
which makes $d(x,y) + d(y,z) < d(x,z)$, thus not triangular inequality

1.2. **Centroid is the closest point to all data points in a cluster.** K-means, 1st step:
check each cluster's center, find one center closest to the data point, then assign the data point to the cluster



for the data points above $a_i, 1 \leq i \leq m$, we want to find one point closest to all data points,
define cost function: $f(x) = \sum_{i=1}^{m} |x - a_i|^2$, we want to get $\min_{x \in R^n} f(x)$, then use the found x to represent $a_i, 1 \leq i \leq m$

Algorithm 1

let $0 = \frac{df(x)}{dx} = \frac{d}{dx}\left[\sum (x-a_i)^2\right] = 2\sum(x-a_i) = 2\sum x - 2\sum a_i$
$\Rightarrow mx = \sum_{i=1}^m a_i \Rightarrow x = \frac{1}{m}\sum_{i=1}^m a_i$ which is the mean of all data points.
Note: Set $A = \{a_1, a_2, \ldots, a_m\} \in R^n$, its centroid $c(A) = \frac{1}{m}\sum_{i=1}^m a_i$ depends on distance function

1.3. **K-means.** Set $A = \pi_1 \cup \pi_2 \cup \cdots \cup \pi_k$, considering hard clustering, $\pi_i \cap \pi_j = \Phi$, and $\{\pi_1, \ldots, \pi_k\} = \Pi$ a partition

A partition is associated with a quality number, such as $\sum_{a\in\pi_1} |c(\pi_1) - a|^2 = q(\pi_1)$
All partition's Quality $Q(\Pi) = \sum_{i=1}^k q(\pi_i)$
Batch K-means:

Definition 1. Initialize partition $\Pi = \{\pi_1, \pi_2\}$,$\pi_1 = \{a_1, a_2, a_3\}, \pi_2 = \{a_4, \ldots, a_8\}$
For each given partition, compute its centroid,
Step 1: For each point, compute its distance to centroids, assign it to closest partition
Step 2: Vector Assignment, update centroids

$$Q(\Pi_1) = \sum_{i=1}^{3} |c_1 - a_i|^2 + \sum_{i=4}^{8} |c_2 - a_i|^2$$
$$Q(\Pi_2) = \sum_{i=1}^{4} |c_1^* - a_i|^2 + \sum_{i=5}^{8} |c_2^* - a_i|^2$$

where

$$|c_2 - a_4|^2_{\Pi_1} > |c_1^* - a_4|^2_{\Pi_2}$$
$$\Rightarrow Q(\Pi_1) > Q(\Pi_2) > \cdots > Q(\Pi_n) \geq 0$$

The mid-plane between any 2 centroids = Convex Hull$\{\Pi_1\} \cap Convex\,\{\Pi_2\}$

1.3.1. *Batch K-means local minimum stuck.* Given N points in 1-dimension, find k=2, trying to cut N-1 lines and calculate Quality number and compare the numbers

Suppose 3 points: 0, 1 , 3
Init $\Pi_1 = \{0\} \cup \{1, 3\}$, then $Q(\Pi_1) = 0 + (1 + 1) = 2$, Batch K-means will stop here.
While$\Pi_2 = \{0, 1\} \cup \{3\}$,then $Q(\Pi_2) = \left(\frac{1}{4} + \frac{1}{4}\right) + 0 = 0.5 < Q(\Pi_1)$.
which means batch K-means failed here to find a global optimization
One way to solve:

1.4. **Incremental K-means.** At each step, only move one point to all clusters and compute quality numbers.

Set $a_1, \ldots, a_m \in R^n$, c as the centroid
$f(x) = \sum_{i=1}^m |x - a_i|^2 = \sum_{i=1}^m |x - c + c - a_i|^2 = \sum_{i=1}^m \left[|x-c|^2 + 2(x-c)^T (x-c) + |c-a_i|^2\right]$
$= \sum_{i=1}^m |x-c|^2 + 2(x-c)^T \sum_{i=1}^m (c - a_i) + \sum_{i=1}^m (c-a_i)^2$
$= m|x-c|^2 + 2(x-c)^T \cdot 0 + \sum_{i=1}^m (c-a_i)^2$
Thus $f(x) = \sum_{i=1}^m |x-a_i|^2 = m|x-c|^2 + \underbrace{\sum_{i=1}^{m} (c-a_i)^2}_{\text{Constant}}$

Set $A = \{a_1, a_2, \ldots, a_p\}, B = \{b_1, \ldots, b_q\}$
then $Q(A \cup B) = Q(A) + Q(B) + p|c - c(A)|^2 + q|c - c(B)|^2$
where $c = \frac{pc(A)+qc(B)}{p+q} = \frac{p}{p+q}c(A) + \frac{q}{p+q}c(B)$
Proof:

$$\begin{aligned} Q(A\cup B) &= \sum_{i=1}^{p} |c - a_i|^2 + \sum_{j=1}^{q} |c - b_j|^2 \\ &= \sum_{i=1}^{p} |c(A) - a_i|^2 + p|c(A) - c|^2 + \sum_{i=1}^{q} |c(B) - b_i|^2 + q|c(B) - c|^2 \\ &= Q(A) + Q(B) + p|c - c(A)|^2 + q|c - c(B)|^2 \end{aligned}$$

proved.

Example: $A = \{a_1, a_2\}, B = \{b\}$
$Q(A\cup B) = Q(A \cup \{b\}) = Q(A) + \underbrace{Q(B)}_{=0} + p|c - c(A)|^2 + 1 \cdot |c - b|^2$

where $c=\frac{p}{p+1}c(A)+\frac{1}{p+1}b\Rightarrow c-b=\frac{p}{p+1}\left[c(A)-b\right]$
$c-c(A)=\frac{1}{p+1}\left[-c(A)+b\right]=\frac{1}{p+1}b-\frac{1}{p+1}c(A)$
Thus $Q\left(A\cup B\right)=Q(A)+\frac{p}{(p+1)^2}\left|b-c(A)\right|^2+\frac{p^2}{(p+1)^2}\left|b-c(A)\right|^2=Q\left(A\right)+\frac{p+p^2}{(p+1)^2}\left|b-c(A)\right|^2=Q\left(A\right)+\left|b-c(A)\right|^2\frac{p}{p+1}$

1.4.1. *For* $A=\left\{a_1,a_2,\ldots,a_p,b\right\},A^-=\left\{a_1,\ldots,a_p\right\},B=\left\{b\right\}$. ? $Q\left(A^-\right)$ interested after removing b .
$\underbrace{Q\left(A\right)}_{\text{Known}}=Q\left(A^-\cup B\right)=\underbrace{Q\left(A^-\right)}_{\text{Unknown?}}+\left|b-c(A^-)\right|^2\frac{p}{p+1}$
$\frac{b+pc(A^{-1})}{1+p}=c(A)\Rightarrow c(A^-)=\frac{1}{p}\left[\left(1+p\right)c(A)-b\right]$
$\Rightarrow b-c(A^-)=\frac{1+p}{p}\left[b-c(A)\right]$

Lemma. $A=\left\{a_1,\ldots,a_p\right\},B=\left\{b\right\},A^+=\left\{a_1,\ldots,a_p,b\right\}$, *then* $Q\left(A^+\right)=Q\left(A\right)+\frac{p}{p+1}\left|c(A)-b\right|^2$

Lemma. $B=\left\{b_1,\ldots,b_{q-1},b_q\right\},B^-=\left\{b_1,\ldots,b_{q-1}\right\},Q\left(B^-\right)=Q\left(B\right)-\frac{q}{q-1}\left|b_q-c(B)\right|^2$

Claim. $A=\left\{a_1,\ldots,a_p\right\},B=\left\{b_1,\ldots,b_q\right\},A^+=\left\{a_1,\ldots,a_p,b_q\right\},B^-=\left\{b_1,\ldots,b_{q-1}\right\}$.
$\left[Q\left(A\right)+Q(B)\right]-\left[Q\left(A^+\right)+Q\left(B^-\right)\right]=\left[Q(A)-Q\left(A^+\right)\right]+\left[Q(B)-Q\left(B^-\right)\right]=-\frac{p}{p+1}\left|c(A)-b_q\right|^2+\frac{q}{q-1}\left|b_q-c(B)\right|^2$

1.5. **Comparision between batch K-means and Incremental K-means.**

Case 1. Batch K-means, $\Pi=\left\{\pi_1,\ldots,\pi_k\right\}$

Step1: compute $c(\pi_i),i=1,\ldots,k$
Step2: $if\ a\in\pi_i,\ m_i=\left|\pi_i\right|;\ if\ a\in\pi_j,m=\left|\pi_j\right|$
look at if $\Delta_{Batch}=\left|c_i-a\right|^2-\left|c_j-a\right|^2>0$, then a should belong to π_j

Case 1. Incremental K-means,

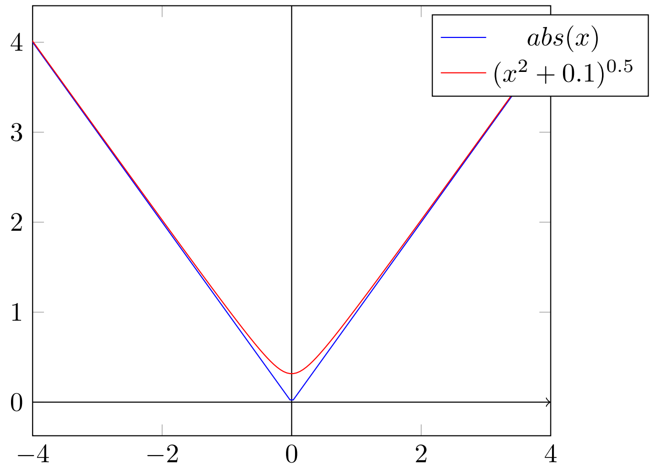
Step2: $if\ a\in\pi_i,\ m_i=\left|\pi_i\right|;\ if\ a\in\pi_j,m=\left|\pi_j\right|$
if $\Delta_{Incremental}=\frac{m_i}{m_i-1}\left|a-c\left(\pi_i\right)\right|^2-\frac{m_j}{m_j+1}\left|a-c\left(\pi_j\right)\right|^2>0$, then we should move a from π_i to π_j .
then $\Delta_{Batch}-\Delta_{Incremental}=-\left\{\frac{m_i}{m_i-1}\left|a-c\left(\pi_i\right)\right|^2-\frac{m_j}{m_j+1}\left|a-c\left(\pi_j\right)\right|^2\right\}+\left\{\left|c_i-a\right|^2-\left|c_j-a\right|^2\right\}$
$=\frac{1}{m_i-1}\left|a-c\left(\pi_i\right)\right|^2-\frac{1}{m_j+1}\left|c_j-a\right|^2\geq 0$
which explains why Batch K-means can miss some important move.

1.6. **2-steps K-means.** Step 1: Run K-means until it stops
Step 2: one iteration of "Incremental K-means"
If (Change is detected) goto Step 1
else Stop

Example. Init partitions as $\Pi_1=\left\{0\right\}\cup\left\{1,3\right\}$, batch K-means will stop
Incremental K-means, $\pi_1=\left\{\Phi\right\},\pi_2=\left\{0,1,3\right\},Q=\left(\frac{4}{3}\right)^2+\left(\frac{1}{3}\right)^2+\left(\frac{5}{3}\right)^2$
$\pi_1=\left\{0,1\right\},\pi_2=\left\{3\right\},Q=\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^2=\frac{1}{2}$
$\pi_1=\left\{0,3\right\},\pi_2=\left\{1\right\},Q=\left(\frac{3}{2}\right)^2+\left(\frac{3}{2}\right)^2=4.5$

1.7. **Quality function.** Given $\left\{a_1,\ldots,a_m\right\}$ build K clusters, $x_j\in R^n,x=\begin{bmatrix}x_1\\x_2\\\vdots\\x_k\end{bmatrix}\in R^{nk}=R^N$

$D_i\left(x_1,\ldots,x_k\right)=\min_{1\leq l\leq k}\left|x_l-a_i\right|^2$
function $D_i:R^N\mapsto R$
$f(x)=\sum_{i=1}^m D_i\left(x_1,\ldots,x_k\right)$, trying to do $\min_{x\in R^N}f(x)$



1.8. **for** $f(x)=|x|$ **, not differentiable.**

One way to make it differentiable since there is only one zero point not differentiable, use a family of smooth function $f_x(x)=\left(x^2+s\right)^{1/2}, s>0$ and s is small

1.9. **for** $(x_1,\ldots,x_m)$ **, find** $f(x)=\max x_i$**.**

Problem. use $f_s(x)=s\log\left(\sum_{i=1}^n\exp\left(\frac{x_i}{s}\right)\right)$,

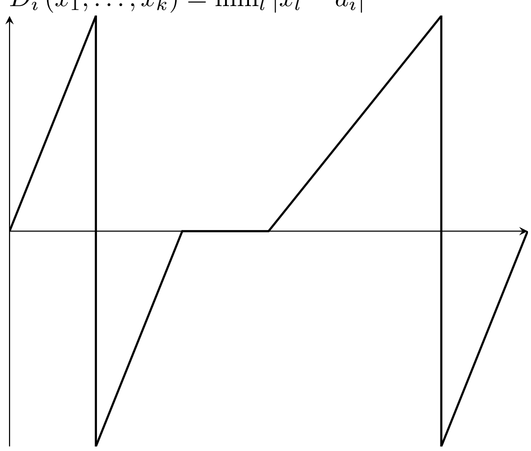
Proposition. $\lim_{s\to0^+}f_s(x)=f(x)$

Proof: for x_1,x_2*,*
if $x_1>x_2$*,*

$$\begin{aligned}s\log\left(\exp\left(\frac{x_1}{s}\right)+\exp\left(\frac{x_2}{s}\right)\right)=&s\log\left[\exp\left(\frac{x_1}{s}\right)\left(1+\exp\left(\frac{x_2-x_1}{s}\right)\right)\right]\\=&s\log\left(\exp\left(\frac{x_1}{s}\right)\right)+s\log\left[1+\exp\left(\frac{x_2-x_1}{s}\right)\right]\\\overset{s\to0^+}{=}&x_1+0\end{aligned}$$

thus $\lim_{s\to0^+}s\log\left[1+\exp\left(\frac{x_2-x_1}{s}\right)\right]=0$

1.10. **Sep 17.** Given $A=\{a_1,\ldots,a_n\},\Pi=\{\pi_1,\ldots,\pi_k\}$
$Q(\Pi)=\sum_{\pi\in\Pi}\sum_{a\in\pi}|a-c(\pi)|^2$
$D_i(x_1,\ldots,x_k)=\min_l|x_l-a_i|^2$



Convex function:
$f:\mathbb{R}^n\mapsto\mathbb{R},\forall t\,and\,0\le t\le1,\ x,y\in R^n,\ f[tx+(1-t)y]\le tf(x)+(1-t)f(y)$

Lemma 2. $f(y)\ge f(x)+\nabla f(x)\cdot(y-x)$

Proof. *for* t close to 0, $f[x+t(y-x)]=f[(1-t)x+ty]\le(1-t)f(x)+tf(y)$

$$\begin{aligned}
\Leftrightarrow f\left[x+t\left(y-x\right)\right] &\leq (1-t)f(x)+tf(y)\\
f\left[x+t\left(y-x\right)\right]-f(x) &\leq t\left[f(y)-f(x)\right]\\
\frac{f\left[x+t\left(y-x\right)\right]-f(x)}{t(y-x)}(y-x) &\leq f(y)-f(x)\\
(y-x)\lim_{t\to 0}\frac{f\left[x+t\left(y-x\right)\right]-f(x)}{t(y-x)} &\leq f(y)-f(x)\\
(y-x)\cdot\nabla f(x) &\leq f(y)-f(x)
\end{aligned}$$

□

2. 20151022 Entropic mean, $\Phi/f - divergence$, presented by Maria

Ben-Tal *et al.* (1989)

2.1. Then entropic mean.

Definition 3. Let $\phi : \mathbb{R}_+ \mapsto \mathbb{R}$ be a strictly convex differentiable function with $(0,1] \subset dom\phi$, and $\phi(1) = 0, \phi'(1) = 0$. Then the class of such functions are denoted by Φ. The "distance" between 2 pdf was defined as $I_\phi(p,q) \coloneqq \sum_{j=1}^n q_j \phi\left(\frac{p_j}{q_j}\right); p,q \in D_n = \left\{x \in \mathbb{R}^n : \sum_{j=1}^n x_j = 1, x > 0\right\}$. Adopting this concept, we define the distance from x to a_i by $d_\phi(x, a_i) \coloneqq a_i \phi\left(\frac{x}{a_i}\right), 1 \le i \le n$. The optimization problem is now:

$$\min\left\{\sum_{i=1}^n w_i a_i \phi\left(\frac{x}{a_i}\right) : x \in \mathbb{R}_+\right\} \tag{2.1}$$

,

and the resulting optimal solution denoted $\bar{x}_\phi(a) = \bar{x}_\phi(a_1, \ldots, a_n)$ will be called the entropic mean of $(a_1, \ldots, a_n)$.

Lemma 4. *Let $\phi \in \Phi$. Then*

(1) for any $\beta_2 > \beta_1 \ge \alpha > 0$ or $0 < \beta_2 < \beta_1 \le \alpha$, $d_\phi(\beta_2, \alpha) > d_\phi(\beta_1, \alpha)$
(2) for any $\alpha_2 \ge \alpha_1 > \beta > 0$ or $0 < \alpha_2 < \alpha_1 < \beta$, $d_\phi(\beta, \alpha_2) > d_\phi(\beta, \alpha_1)$

Proof. a). Since ϕ is strictly convex, $d_\phi(\cdot, \alpha)$ is strictly convex for any $\alpha > 0$, which can be showed by:
$\frac{\partial d_\phi(\cdot,\alpha)}{\partial \alpha} = \phi\left(\frac{\cdot}{\alpha}\right) + \alpha\phi'\left(\frac{\cdot}{\alpha}\right)\left(-\frac{\cdot}{\alpha^2}\right) = \phi\left(\frac{\cdot}{\alpha}\right) - \frac{\cdot}{\alpha}\phi'\left(\frac{\cdot}{\alpha}\right)$
$\frac{\partial^2 d_\phi(\cdot,\alpha)}{\partial \alpha^2} = \phi'\left(\frac{\cdot}{\alpha}\right)\left(-\frac{\cdot}{\alpha^2}\right) - \left[-\frac{\cdot}{\alpha^2}\phi'\left(\frac{\cdot}{\alpha}\right) + \frac{\cdot}{\alpha}\phi''\left(\frac{\cdot}{\alpha}\right)\left(-\frac{\cdot}{\alpha^2}\right)\right] = \frac{\cdot}{\alpha^3}\phi''\left(\frac{\cdot}{\alpha}\right) > 0$, since $\alpha > 0$.
Thus by gradient inequality for $d_\phi(\cdot, \alpha)$,
$d_\phi(\beta_2, \alpha) = \alpha\phi\left(\frac{\beta_2}{\alpha}\right) > \alpha\phi\left(\frac{\beta_1}{\alpha}\right) + (\beta_2 - \beta_1)\phi'\left(\frac{\beta_1}{\alpha}\right)$
Since $\phi'(1) = 0, \begin{cases}\phi'(x) > 0, & if\ x > 1\\ \phi'(x) < 0, & if\ x < 1\end{cases}$, then $(\beta_2 - \beta_1)\phi'\left(\frac{\beta_1}{\alpha}\right) > 0$, since $\begin{cases}\text{if } \beta_2 > \beta_1 \ge \alpha > 0,\\ \text{if } 0 < \beta_2 < \beta_1 \le \alpha,\end{cases}$
b). it is straight-forward to find out $d_\phi(\beta, \cdot)$ is srictly convex for the 1st argument, then from a), it is similiarly proved. □

Corollary 5. *Let $\phi \in \Phi$ and $\alpha, \beta > 0$. Then $d_\phi(\beta, \alpha) \ge 0$ with equality IFF $\alpha = \beta$.*

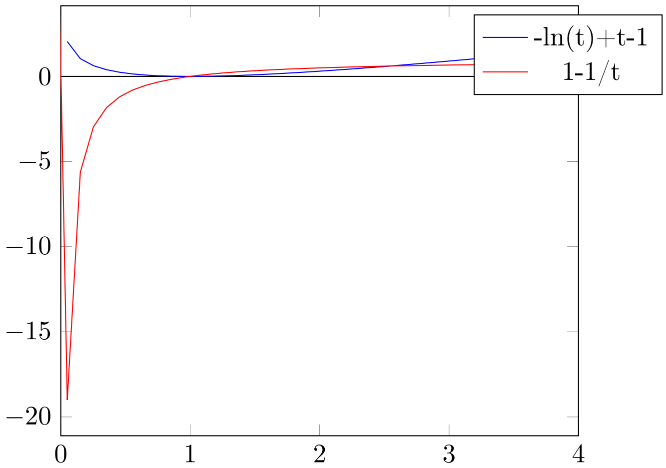
Theorem 6. *Let $\phi \in \Phi$. Then*

(1) There exists a unique continuous function $\bar{x}_\phi$ which solves 2.1 such that
$$\min_{1\le i\le n}\{a_i\} \le \bar{x}_\phi \le \max_{1\le i\le n}\{a_i\}$$
, for all $a_i > 0$. In particular $\bar{x}(\alpha, \ldots, \alpha) = \alpha$.
(2) The mean $\bar{x}_\phi$ is strict, i.e.,
$$\min_{1\le i\le n}\{a_i\} < \max_{1\le i\le n}\{a_i\} \Rightarrow \min_{1\le i\le n}\{a_i\} < \bar{x}_\phi < \max_{1\le i\le n}\{a_i\}$$
.
(3) $\bar{x}_\phi$is homogeneous (scale invariant), i.e.,
$$\bar{x}_\phi(\lambda a) = \lambda \bar{x}_\phi(a)$$
, for all $\lambda > 0, a_i > 0$.
(4) If $w_i = w$ for all i , then $\bar{x}_\phi$is symmetric; i.e., $\bar{x}(a_1, \ldots, a_n)$ is invariant to permutations of the $a_i's > 0$.
(5) $\bar{x}_\phi$ is isotone; i.e., for all i and fixed $\{a_j\}_{j=1}^n > 0, j \ne i, \bar{x}(a_1, \ldots, a_{j-1}, \cdot, a_{j+1}, \ldots, a_n)$ is an increasing function.

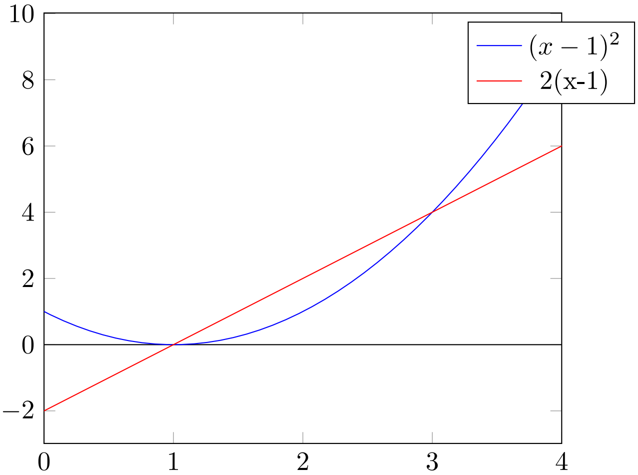
2.2. Examples.

Example 7 (Classical means)**.** Solving the derivative of 2.1: $\sum_{i=1}^n w_i\phi'\left(\frac{x}{a_i}\right) = 0$

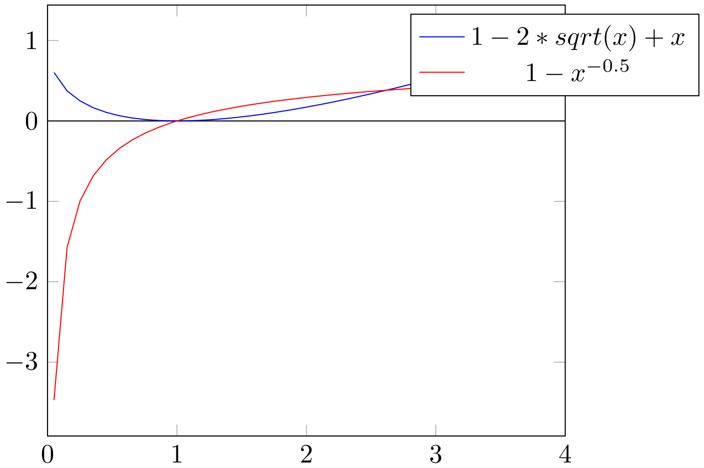
(1) Arithmetic mean. $\phi(t) = -\lg t + t - 1$, $\phi'(t) = -\frac{1}{t} + 1$,then $0 = \sum_{i=1}^n w_i\phi'\left(\frac{x}{a_i}\right) = \sum_{i=1}^n w_i\left(1 - \frac{a_i}{x}\right) \Rightarrow \bar{x} = \sum_{i=1}^n w_i a_i \coloneqq A(a)$.



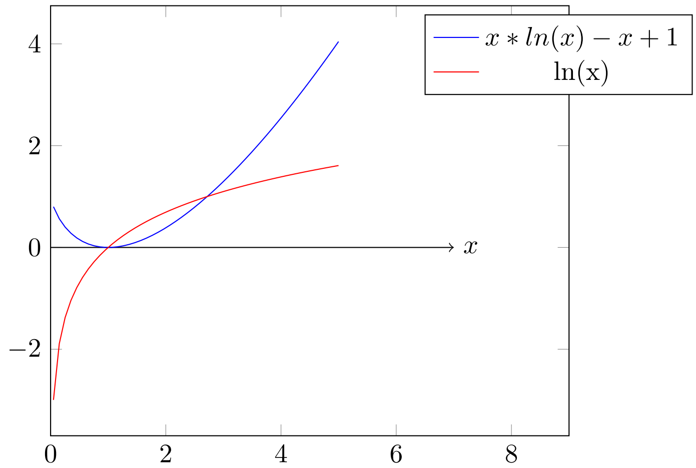
2) Harmonic mean. $\phi(t) = (t-1)^2$, $\phi'(t) = 2(t-1)$, then $0 = \sum_{i=1}^{n} w_i \phi'\left(\frac{x}{a_i}\right) = \sum_{i=1}^{n} w_i \left(\frac{x}{a_i} - 1\right) \Rightarrow \bar{x} = \left(\sum_{i=1}^{n} w_i \frac{1}{a_i}\right)^{-1} := H(a)$.



3) Root mean square. $\phi(t) = 1 - 2\sqrt{t} + t$, $\phi'(t) = 1 - t^{-1/2}$, then $0 = \sum_{i=1}^{n} w_i \phi'\left(\frac{x}{a_i}\right) = \sum_{i=1}^{n} w_i \left(1 - \sqrt{\frac{a_i}{x}}\right) \Rightarrow \bar{x} = \left(\sum_{i=1}^{n} w_i \sqrt{a_i}\right)^2 := R(a)$.

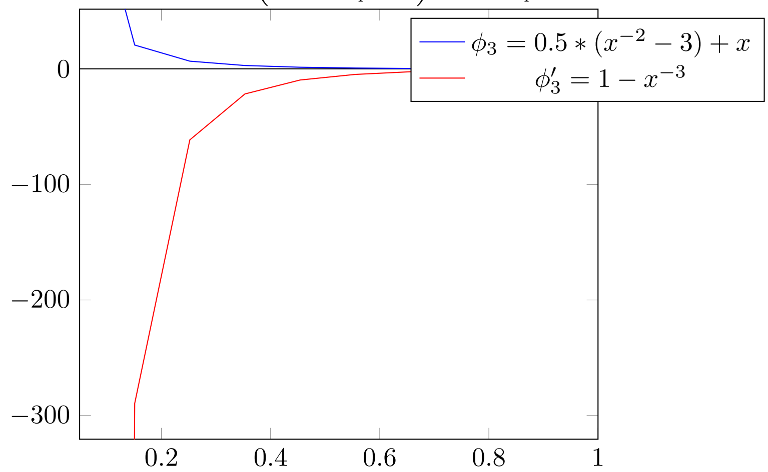


4) Geometric mean. $\phi(t) = t \log t - t + 1$,$\phi'(t) = \lg t$, then $0 = \sum_{i=1}^{n} w_i \phi'\left(\frac{x}{a_i}\right) = \sum_{i=1}^{n} w_i (\log x - \log a_i) \Rightarrow \bar{x} = \exp\left(\sum_{i=1}^{n} w_i \log a_i\right) = \prod_{i=1}^{n} a_i^{w_i} := G(A)$.



5) Mean of order p. $\phi_p(t) = \frac{t^{1-p}-p}{p-1} + t$, $p \neq 1, p > 0$. $\phi_p'(t) = 1 + \frac{(1-p)t^{-p}}{p-1} = 1 - t^{-p}$, then $0 = \sum_{i=1}^n w_i \phi'\left(\frac{x}{a_i}\right) = \sum_{i=1}^n w_i \left(1 - \frac{a_i^p}{x^p}\right) \Rightarrow x^p = \sum w_i a_i^p \Rightarrow \bar{x} = \left(\sum w_i a_i^p\right)^{1/p}$. To extend $\bar{x_p}$for negative order, one may choose $\tilde{\phi}_q(t) = \frac{t^q - qt}{q-1} + 1, q > 0, q \neq 1 \Rightarrow \bar{x}_q = \left(\sum w_i a_i^{1-q}\right)^{1/(1-q)}$, since $\tilde{\phi}_q'(t) = \frac{qt^{q-1}-q}{q-1} = \frac{q}{q-1}\left(t^{q-1}-1\right)$, then $0 = \sum_{i=1}^n w_i \tilde{\phi}'\left(\frac{x}{a_i}\right) = \sum_{i=1}^n w_i \left(\frac{x^{q-1}}{a_i^{q-1}} - 1\right) \Rightarrow \bar{x}_q = \left(\sum w_i a_i^{1-q}\right)^{1/(1-q)}$.

Note, $\tilde{\phi}_q(t) = t\phi_p\left(\frac{1}{t}\right) = t\left(t^{-1} + \frac{t^{p-1}-p}{p-1}\right) = 1 + \frac{t^p - pt}{p-1}$, hence $\tilde{\phi}_q$ is also strictly convex $\forall t > 0$.



And for $p = q = 0.5$ yielding the Root mean square R,
$\bar{x} = \left(\sum w_i a_i^p\right)^{1/p} = \bar{x} = \left(\sum w_i a_i^{0.5}\right)^2$
$q = 2$ gives harmonic mean H, $\bar{x}_q = \left(\sum w_i a_i^{-1}\right)^{-1}$.
Application of $L'Hospital's$ rule shows:

$$\begin{aligned}
\lim_{p\to 1} \phi_p(t) &= \lim_{p\to 1}\left(\frac{t^{1-p}-p}{p-1} + t\right) \\
&= t + \lim_{p\to 1}\left(-t^{1-p}\log t - 1\right) \\
&= t - \log t - 1
\end{aligned}$$

which corresponds to Arithmetic mean's ϕ.
and

$$\begin{aligned}
\lim_{q\to 1} \phi_q(t) &= \lim_{q\to 1}\left(\frac{t^q - qt}{q-1} + 1\right) \\
&= 1 + \lim_{q\to 1}\left(t^q \log t - t\right) \\
&= 1 + t\log t - t
\end{aligned}$$

which corresponds to Geometric mean's ϕ.

9) Composition of means.

Let $\phi(t) = \frac{1}{3}\left(-2\log t + t^2 - 1\right), \Rightarrow 0 = \sum w_i \left(\frac{x}{a_i} - \frac{a_i}{x}\right) \Rightarrow \bar{x}_\phi(a) = \left(\frac{\sum w_i a_i}{\sum w_i / a_i}\right)^{1/2} = \sqrt{A(a)H(a)}$, i.e., the geometric mean of A and H.

2.3. **Comparision of means.**

Theorem 8. *Let $\phi, \psi \in \Phi$ and denote by $\bar{x}_\phi, \bar{x}_\psi$ respectively the corresponding entropic means. If there exists a constant $K \neq 0$ such that*

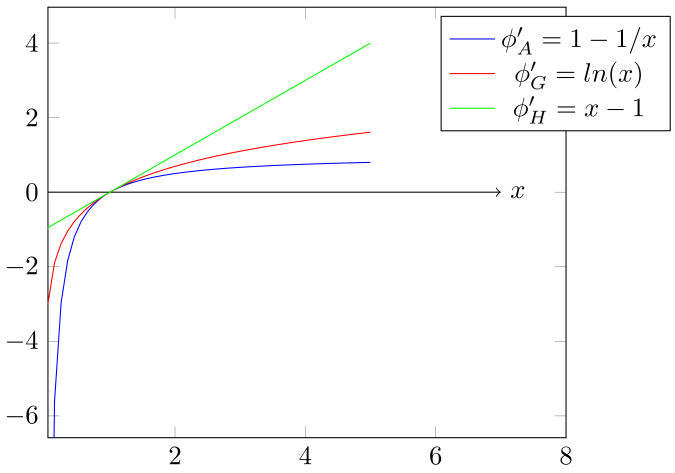
$$K\phi'(t) \leq \psi'(t), \forall t \in \mathbb{R}_+ \setminus \{1\}$$

then,

$$\bar{x}_\phi(a) \geq \bar{x}_{psi}(a).$$

Proof by contradiction.

Example 9. Show the classical inequalities $A(a) \geq G(a) \geq H(a)$ are an easy consequence of the theorem above.



Theorem 10. *Let $\phi_1, \phi_2 \in \Phi$ and $\phi_\lambda(t) := \lambda\phi_1(t) + (1-\lambda)\phi_2(t)$, then for all $0 \leq \lambda \leq 1$,*

$$\min\{\bar{x}_{\phi_1}(a), \bar{x}_{\phi_2}(a)\} \leq \bar{x}_{\phi_\lambda}(a) \leq \max\{\bar{x}_{\phi_1}(a), \bar{x}_{\phi_2}(a)\}$$

Proof. 1st note $\forall\lambda \in [0,1], \phi_\lambda \in \Phi$. Now $\bar{x}_{\phi_\lambda}$ is obtained from

$$\sum_{i=1}^n w_i \left\{\lambda\phi_1'\left(\frac{\bar{x}_{\phi_\lambda}}{a_i}\right) + (1-\lambda)\phi_2'\left(\frac{\bar{x}_{\phi_\lambda}}{a_i}\right)\right\} = 0$$

Assume $\bar{x}_{\phi_\lambda} < \min(\bar{x}_{\phi_1}, \bar{x}_{\phi_2})$, then since ϕ_1', ϕ_2' are strictly increasing we have

$$\lambda\sum_{i=1}^n w_i\phi_1'\left(\frac{\bar{x}_{\phi_1}}{a_i}\right) + (1-\lambda)\sum_{i=1}^n w_i\phi_2'\left(\frac{\bar{x}_{\phi_2}}{a_i}\right) > 0. \tag{2.2}$$

but from the optimality conditions for $\bar{x}_{\phi_1}, \bar{x}_{\phi_2}$, the left hand of 2.2 should be equal to 0, hence the contradiction.

Similarly for $\bar{x}_{\phi_\lambda} > \max(\bar{x}_{\phi_1}, \bar{x}_{\phi_2})$. □

Example 11. let $\phi_1(t) = -\log t + t - 1$ and $\phi_2(t) = (t-1)^2$. Then $\bar{x}_{\phi_1} = A(a)$ and $\bar{x}_{\phi_2} = H(a)$. Consider $\lambda = \frac{2}{3}, \phi_\lambda(t) = \lambda\phi_1(t) + (1-\lambda)\phi_2(t)$ which is the function to derive the geometric mean of harmonic mean and arithematic mean.

Then $H(a) \leq \sqrt{A(a)H(a)} \leq A(a)$ since $H(a) \leq A(a)$.

2.4. **Entropic mean for random variables.** Let A be a non-negative RV with distribution F and support $suppA := [\alpha, \beta], 0 \leq \alpha \leq \beta \leq +\infty$. A natural generalization of problem 2.1 is

$$\min\left\{E\left[A\phi\left(\frac{x}{A}\right)\right] := \int_\alpha^\beta t\phi\left(\frac{x}{t}\right)dF(t) : x \in \mathbb{R}_+\right\},$$

$$Pr\{A = a_i\} := w_i, \text{for the discrete case above}$$

Theorem 12. *then for any positive random variable A:*

(1) There exists a unique $\bar{x}_\phi$ which solves 2.1 such that $\bar{x}_\phi \in suppA$.
(2) If a is a degenerate RV, i.e., A=C where C is a positive finite constant, then $\bar{x}_\phi = C$.
(3) For all $\lambda > 0, \bar{x}_\phi(\lambda A) = \lambda\bar{x}_\phi(A)$.

The optimality condition equation :

$$\int_{\alpha}^{\beta}\phi'\left(\frac{x}{t}\right)dF(t)=0$$

Example 13 (θth Quantile.)**.** Let

$$\phi(\xi)=\begin{cases}(1-\theta)(\xi-1) & ,\xi>1\\ \theta(1-\xi) & ,0<\xi\le 1\end{cases}\ (0<\theta<1)$$

Remark 14. The differentiability assumption on ϕ can be relaxed, since ϕ is convex, its left and right derivative $\phi'_{-}(x)$ and $\phi'_{+}(x)$ exist and finite and increasing. Moreover, the subdifferential of ϕ is $\partial\phi(x)=\left[\phi'_{-}(x),\ \phi'_{+}(x)\right]$. then $0\in\partial\phi(x)$ and it is guaranteed for all strictly convex function $\phi(x)>0,\forall x\in\mathbb{R}_{+}\setminus\{1\}$.

Clearly $\phi(\xi)$ is not differentiable at $\xi=1$, but $0\in\partial\phi(1)$, thus not matter.
The objective function of 2.1 is

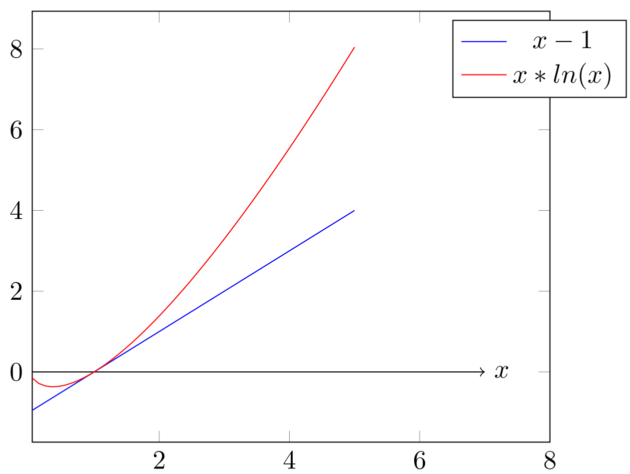
$$\begin{aligned}
h(x):=&E\left[t\phi\left(\frac{x}{A}\right)\right]=\int_0^{\infty}t\phi\left(\frac{x}{t}\right)dF(t)\\
=&(1-\theta)\int_0^x(x-t)\,dF(t)+\theta\int_x^{\infty}(t-x)\,dF(t)\\
=&(1-\theta)\left[x\int_0^x dF(t)-\int_0^x tdF(t)\right]+\theta\left[\int_x^{\infty}tdF(t)\right]-\theta x\left[1-\int_0^x dF(t)\right]\\
=&\underbrace{(1-\theta)\,xF(x)}-(1-\theta)\int_0^x tdF(t)+\theta\left[E(A)-\int_0^x tdF(t)\right]-\theta x+\underbrace{\theta xF(x)}\\
=&xF(x)-\int_0^x tdF(t)+\underbrace{\theta\int_0^x tdF(t)}+\theta E(A)-\underbrace{\theta\int_0^x tdF(t)}-\theta x\\
=&xF(x)-\int_0^x tdF(t)+\theta E(A)-\theta x
\end{aligned}$$

Then $0=h'(\bar{x}_{\phi})=F(x)+xf(x)-xf(x)-\theta\Rightarrow F(\bar{x}_{\phi})=\theta$, i.e., $\bar{x}_{\phi}$ is the θth quantile of the continuous RV A. In particular for $\theta=\frac{1}{2},\bar{x}_{\phi}$ is the median.

2.5. **Extremal principle for the HLP Generalized mean.**

Claim 15. Means not processing the properties listed in 6 cannot be derived from the solution of problem 2.1 with the entropy type distance $d_{\phi}(x,a_i)=a_i\phi(x/a_i)$.

Inequality: $\forall p > 0$, $p \ln p \geq p - 1$



Misc (Xiaowei Song)
Hölder's Inequality:
$\left|u^T v\right| \leq \left(\sum |u_i|^p\right)^{\frac{1}{p}} \left(\sum |v_i|^q\right)^{\frac{1}{q}}$, where $\frac{1}{p} + \frac{1}{q} = 1$
For

$$\Psi(z) = \log\left(\sum_{i=1}^{k} e^{-z_i}\right), z \in R^k$$

1. Ψ is convex (5 points)
Proof:
let $z_\lambda = \frac{1}{p}x + \frac{1}{q}y$,where $x, y \in R^k, p, q \in R^+$ *and* $\frac{1}{p} + \frac{1}{q} = 1$, $\lambda = \frac{1}{p}, 1-\lambda = \frac{1}{q}$, $0 < \lambda < 1$

$$\begin{aligned}
\Psi\left(\frac{1}{p}x + \frac{1}{q}y\right) = \Psi(z_\lambda) &= \log\left(\sum_{i=1}^{k} e^{-z_i}\right) \\
&= \log\left(\sum_{i=1}^{k} \exp\left[-\left(\frac{1}{p}x_i + \frac{1}{q}y_i\right)\right]\right) \\
&= \log\left(\sum_{i=1}^{k} \exp\left[-\left(\frac{1}{p}x_i\right)\right] \exp\left[-\left(\frac{1}{q}y\right)\right]\right) \\
&= \log\left(\sum_{i=1}^{k} \left[\exp\left(-x_i\right)\right]^{\frac{1}{p}} \left[\exp\left(-y_i\right)\right]^{\frac{1}{q}}\right) \\
&= \log\left(u^T v\right)
\end{aligned}$$

where

$$u = \begin{bmatrix} [\exp(-x_1)]^{\frac{1}{p}} \\ [\exp(-x_2)]^{\frac{1}{p}} \\ \vdots \\ [\exp(-x_k)]^{\frac{1}{p}} \end{bmatrix}, v = \begin{bmatrix} [\exp(-y_1)]^{\frac{1}{q}} \\ [\exp(-y_2)]^{\frac{1}{q}} \\ \vdots \\ [\exp(-y_k)]^{\frac{1}{q}} \end{bmatrix}$$

thus from Hölder's Inequality, we can have

$$\begin{aligned}
u^T v = \left|u^T v\right| &\leq \left(\sum |u_i|^p\right)^{\frac{1}{p}} \left(\sum |v_i|^q\right)^{\frac{1}{q}} \\
&= \left(\sum_{i=1}^{k} \left|\left[\exp\left(-x_i\right)\right]^{\frac{1}{p}}\right|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{k} \left|\left[\exp\left(-x_i\right)\right]^{\frac{1}{p}}\right|^p\right)^{\frac{1}{q}} \\
&= \left(\sum_{i=1}^{k} \exp\left(-x_i\right)\right)^{\frac{1}{p}} \left(\sum_{i=1}^{k} \exp\left(-y_i\right)\right)^{\frac{1}{p}}
\end{aligned}$$

(where $u^T v > 0$ since $u_i > 0, v_i > 0, i = 1, \ldots, k$, thus $u^T v = \left|u^T v\right|$)
thus

$$\begin{aligned}
\Psi\left(\frac{1}{p}x + \frac{1}{q}y\right) = \Psi(z_\lambda) &= \log\left(u^T v\right) \\
&\leq \log\left(\sum_{i=1}^{k} \exp\left(-x_i\right)\right)^{\frac{1}{p}} \left(\sum_{i=1}^{k} \exp\left(-y_i\right)\right)^{\frac{1}{p}} \\
&= \frac{1}{p}\log\left(\sum_{i=1}^{k} \exp\left(-x_i\right)\right) + \frac{1}{q}\log\left(\sum_{i=1}^{k} \exp\left(-y_i\right)\right) \\
&= \frac{1}{p}\Psi(x) + \frac{1}{q}\Psi(y)
\end{aligned}$$

which proved convexity of functionΨ
2.

$$\Psi(y) - \Psi(z) \leq \sum_{i=1}^{k} (z_i - y_i) \frac{\exp(-y_i)}{\sum_{i=1}^{k} \exp(-y_i)}$$

References

Ben-Tal, Aharon, Charnes, Abraham, & Teboulle, Marc. 1989. Entropic means. *Journal of Mathematical Analysis and Applications*, **139**(2), 537–551. Cited by 0040. 2